Comparing the Slack-Variable Mixture Model With Other Alternatives

Lulu KANG
Department of Applied Mathematics
Illinois Institute of Technology
Chicago, IL 60616
(ikang2@iit.edu)

Javier Cruz SALGADO
Department of Manufacturing and Industrial Engineering
Centro de Innovaciones Aplicadas en Tecnologías Competitivas (CIATEC)
León, Gto, México
(jsalgado.picyt@ciatec.mx)

William A. BRENNEMAN
Statistics Department
The Procter & Gamble Company
Cincinnati, OH 45040
(brenneman.wa@pg.com)

There have been many linear regression models proposed to analyze mixture experiments including the Scheffé model, the slack-variable model, and the Kronecker model. The use of the slack-variable model is somewhat controversial within the mixture experiment research community. However, in situations that the slack-variable ingredient is used to fill in the formulation and the remaining ingredients have constraints such that they can be chosen independently of one another, the slack-variable model is extremely popular by practitioners mainly due to the ease of interpretation. In this article, we advocate that for some mixture experiments the slack-variable model has appealing properties including numerical stability and better prediction accuracy when model-term selection is performed. We also explain how the effects of the slack-variable model components should be interpreted and how easy it is for practitioners to understand the components effects. We also investigate how to choose the slack-variable component, what transformation should be used to reduce collinearity, and under what circumstances the slack-variable model should be preferred. Both simulation and practical examples are provided to support the conclusions.

KEY WORDS: Condition number; Mixture experiments; Scheffé model; Slack-variable model; Variance inflation factor (VIF); Variable transformation.

1. INTRODUCTION

Mixture experiments are commonly used in many fields such as chemical, pharmaceutical, and food industries. In such experiments, the product is developed by mixing different ingredients or components. Different from other types of experiments, the design variables in mixture experiments are the proportions of different components. In this article, we focus on the mixture experiments with proportions as the only input variables. Other types of mixture experiments involving the total amount of the mixture or certain process variables (Piepel and Cornell 1985; Kowalski, Cornell, and Vining 2002; Goldfarb et al. 2004) are out of the scope of this article.

For a mixture experiment of $q$ components, their proportions $x_1, x_2, \ldots, x_q$ must satisfy

\[ x_1 + x_2 + \cdots + x_q = 1. \quad (1) \]

This constraint differentiates mixture experiments from other types of experiments. Not only does the experimental design region become constrained, but the resulting model from a mixture experiment also has to satisfy the constraint. The most frequently used models include the Scheffé model (S-model) and the slack-variable model (SV-model). The Kronecker model (K-model) is less commonly used but it is recommended in Prescott et al. (2002) because its Fisher information matrix has a smaller condition number than that of the S-model in some cases. The quadratic order S-, SV-, and K-models have the form

**S-model**

\[ E(Y) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q + \beta_{12} x_1 x_2 + \cdots + \beta_{q-1,q} x_{q-1} x_q, \quad (2) \]

**SV-model**

\[ E(Y) = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_{q-1} x_{q-1} + \sum_{i=1}^{q-1} \gamma_i x_i^2 + \sum_{i < j} \gamma_{ij} x_i x_j, \quad (3) \]

**K-model**

\[ E(Y) = \alpha_{11} x_1^2 + \alpha_{22} x_2^2 + \cdots + \alpha_{qq} x_q^2 + \alpha_{12} x_1 x_2 + \cdots + \alpha_{q-1,q} x_{q-1} x_q. \quad (4) \]

There are several challenges in modeling mixture experiments. First, the regular polynomial regression models, such as the second-order polynomial $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_q x_q + \beta_{12} x_1 x_2 + \cdots + \beta_{q-1,q} x_{q-1} x_q + \beta_{11} x_1^2 + \cdots + \beta_{qq} x_q^2$, contain the intercept term and thus cannot...
be directly used for mixture experiments. The linear constraint in Equation (1) must be imposed on the relevant terms in the model. Otherwise, the regular polynomial model matrix $F$ generated based on the constrained mixture design will not be of full rank and thus the information matrix $F'F$ will become singular. Different representations of the constraint will lead to different models. To convert the regular second-order polynomial model into the S-model (2), the intercept $\beta_0$ is multiplied by $x_1 + x_2 + \cdots + x_q$, and the quadratic terms are replaced by $x_i^2 = x_i(1 - \sum_{j \neq i} x_j)$ yielding (2) after simplification. The SV-model can be obtained from the S-model by replacing one component proportion, for example $x_q$, with $1 - \sum_{i=1}^q x_i$. The variable $x_q$ is called a slack variable. The K-model (4) can be generated from the S-model by replacing each linear term with $x_i = x_i(x_1 + x_2 + \cdots + x_q)$, so it only contains the second-order terms.

Second, interpreting mixture models is quite difficult. For the S- and K-models, since all the proportions $x_i$’s are involved in the models, increasing/decreasing one unit of $x_i$ means that the sum of the other $q - 1$ components has to be decreased/increased by one unit, but it is uncertain how each of $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_q$ should be changed. So the first- and second-order model terms should not be interpreted as the main effects and interactions as they would be in nonmixture models. There are some useful techniques for interpretation of the S-model, such as the Cox and Piepel effect directions (see Smith 2005, pp. 257–268). But compared with these interpretation methods, the SV-model does offer a more straightforward way to understand the components effects for practitioners, which are to be discussed in detail in Section 2.

Third, also due to the constraint in Equation (1), model-term selection has to be done carefully. For the S-model, none of the first-order terms should be removed from the model; otherwise, the constraint Equation (1) is violated. For the same reason, none of the second-order terms in the K-model should be removed as they are all involved in the constraint via $x_i = x_i(x_1 + x_2 + \cdots + x_q)$ for $i = 1, \ldots, q$.

Finally, the information matrix of mixture models can easily become ill-conditioned. Denote $F_{n \times k}$ as the model matrix of $n$ runs and $k$ terms (including the intercept if the model contains it). The information matrix is then equal to $F'F$, with $(F'F)^{-1}$ being the covariance matrix of the estimated coefficients (without the variance of the noise). It is well known that for the linear regression model if the information matrix is ill-conditioned, the least-square estimation becomes numerically unstable, that is a small perturbation in the response $y$ will cause large change in the least-square estimates. But the unknown true relationship between the response and component proportions should not change much for a small perturbation. In the context of this paper, numerical stability refers to the numerical stability of the least-square estimation, which essentially depends on the condition of the information matrix. One obvious cause of the ill-conditioning problem is that for some mixture models, such as the S- and K-models, some columns of $F$ are dependent on each other because of Equation (1). Other reasons are related to the experimental design. In some experiments, certain component proportions have to be varied in a narrow range or even close to zero. One such example is shown in Section 2 where the minimal range of components is $[0, 0.5\%]$ resulting in the condition numbers of the mixture models to be quite large. All of these factors can lead to collinearity of the model matrix. In some very complicated mixture experiments, such as the one in Piepel, Cooley, and Jones (2005), more linear constraints are involved and the collinearity problem can become more severe. Sometimes the design-related issues could be overcome by using a linear transformation of the design variables, such as the $L$- and $U$-, and modified-pseudo-components transformation. But there are no universal rules to deal with the ill-conditioning and collinearity problems as discussed in many papers including Montgomery and Voth (1994), Prescott et al. (2002), and Cornell and Gorman (2003). In our context, the three concepts of numerical stability, condition of the information matrix, and collinearity essentially refer to the same concept and sometimes we write them interchangeably. We use the common criteria of the condition number of the information matrix and the variance inflation factor (VIF) values to measure the numerical stability.

Different mixture models have different pros and cons. The S-model enjoys a great popularity because it incorporates the constraint Equation (1) in the simplest way but can cause some computational problems. Since it is very rare for the regression model not to include the intercept, some statistical software such as R does not recognize the S-model as a mixture model. For models without an intercept, R simply assumes that the mean of the response is 0 and the response is equal to 0 under the null hypothesis. Consequently, the other statistics and inferences calculated based on this assumption, such as $R^2$, adjusted $R^2$, $F$-ratio, and ANOVA, are incorrect. If users neglect this detail, they can be misled by the software output. The correct calculation of these statistics is very basic and can be found in Cornell (2002). In addition, it may not be possible to interpret the effects of the S-model and model-term selection of the effects is also restricted. The K-model sometimes has a less ill-conditioned information matrix than the S-model, but it shares all the other disadvantages of the S-model. In this article when comparing different models, the common criteria $R^2$ and $R_{adj}^2$, model size, and the estimated standard error $\hat{\sigma}$ are used to evaluate the model fit, the simplicity, and the precision of the models, respectively. Prediction accuracy is another important aspect in our comparison and it is measured by the prediction error, defined by $E(Y(X) - \hat{Y}(X))^2$ (Hastie et al. 2009) and can be estimated by new test data, which are unfortunately not available in all the examples here. Thus we use cross-validation, the simplest and most widely used method to estimate prediction error (Hastie et al. 2009).

In this article, we discuss how the SV-model can overcome these problems and under what circumstances it performs better than the others. In Section 2, we use one practical example to illustrate why the SV-model is so attractive to practitioners. In Section 3, we study which component proportion should be used as the slack variable to minimize the collinearity in the SV-model. We compare the SV-model with the S- and K-models in Section 4 based mainly on the prediction accuracy and numerical stability. Examples and a simulation study are also shown in Section 4. Section 5 provides the conclusion. Some more extended simulation studies are contained in the supplement.
2. SV-MODEL AND FILLER COMPONENT

2.1 Filler and Its Effects

A typical example of what is often referred to as a filler is water within shampoo, detergent, and many other liquid products. For some solid products, flour acts as filler within products such as baked goods, lactose, microcrystalline cellulose, and cornstarch. Sugar is the common filler in drug tablets, caplets, and pills. Fillers are so common that they exist in most mixture products. There are two rules of thumb for telling if a component should be considered as a filler. First, fillers are usually not an active part of the formulation, meaning that they do not actively affect certain properties of the mixture product. Second, fillers often make up a large percentage of the entire mixture or have the largest range of possible proportion values. There are exceptions to these rules. However, the main way to determine a filler is to use expert knowledge to understand how the formulators think about their mixture components. If a component is used to “fill in” the mixture after the other components have been determined, then this “fill in” component should be treated as a filler.

The effects of the fillers over the entire mixture are different from the other components. Take sweet tea as an example. We can increase the sweetness of the tea by adding more sugar or by reducing the amount of the water. In both ways, the percentage of sugar is increased and the percentage of water is decreased. But when people explain why the tea has become sweeter, most would say that it is because the percentage of sugar is increased. One would say the sugar is the ingredient that affects the sweetness of the tea. But this does not mean that water has no effect. It only means that the water affects the sweetness negatively as a solvent.

For an experiment with a filler component \( x_q \), if \( x_q \) is used to “fill in” the formula so that Equation (1) is satisfied, then both the design and analysis can be performed through classic nonmixture techniques. For the analysis, it is intuitive to let the filler proportion be the slack variable. As the slack variable, the filler proportion does not appear in the SV-model and we simply interpret all the other component effects with respect to the filler which greatly helps practitioners understand how each component affects the overall mixture property. For the experimental design, if the nonfiller components only have to meet lower and upper bounds without any other constraints, and the upper bounds are such that \( \sum_{i=1}^{7} U_i \leq 1 \), we can then use classical experimental design methods such as fractional factorial designs, orthogonal arrays, and optimal designs for the nonfiller components and use the filler to fill in the mixture.

2.2 A Real Example

Following is a real mixture experiment for a hair product. There are eight active components and water, which is considered as the filler. The upper and lower bounds of each component proportion (in percentage) and the range of the response \( y \) are shown in Table 1. We have used a 61-run I-optimal design (no replicated points, generated by JMP Pro 10) on \( x_1 \sim x_8 \) for the complete quadratic SV-model Equation (5) with water as the slack variable

\[
E(Y) = \beta_0 + \sum_{i=1}^{8} \beta_i x_i + \sum_{i=1}^{8} \beta_i x_i^2 + \sum_{i \neq j}^{8} \beta_{ij} x_i x_j. \tag{5}
\]

Then, four replicates at the centroid of the constrained region are added. So, the total run size is 65. First, we fit the model in Equation (5) and then use stepwise regression to obtain a reduced model. BIC is used to select the model terms. The parameter estimates and model fit summaries are in Tables 2 and 3. We estimate the noise variance \( \hat{\sigma}^2 \) as the sample variance of the four replications and calculate the lack-of-fit test as \( \text{SSE} - \frac{3\text{df}_{\text{SSE}}}{\hat{\sigma}^2} \) as in Table 3. The ANOVA F-test statistic of the reduced model with respect to the complete model in Equation (5) is 1.2179 with a \( p \)-value of 0.33. The analysis shows that the reduced model contains less than half of the parameters in the complete model. The design and analysis for this experiment are certainly routine as they follow a classical design and analysis approach. Any standard statistical software is sufficient for the task. Model-term selection is a necessary step since it produces a much more parsimonious model with smaller estimated variance and improved precision. If S- or K-model were used, the design and analysis would not be so convenient.

More importantly, the interpretation of the SV-model is clear and simple: when the proportion of a nonfiller component is varied, the proportions of other active components remain the same and only the filler amount is changed to make up the difference. Therefore, it is very easy to understand the role of an individual component. For the hair product example, Figure 1 shows the prediction profiler and confidence intervals of \( E(Y) \) for all the factors based on the reduced model. From Figure 1, we can see that components \( x_1, x_4, x_7, \) and \( x_5 \) affect \( E(Y) \) linearly. More specifically, \( E(Y) \) has strong positive trend with \( x_1 \) and \( x_4 \), strong negative trend with \( x_7 \), and less obvious negative trend with \( x_5 \). The components \( x_2, x_3, x_6, \) and \( x_8 \) strongly affect \( E(Y) \) quadratically. The readers should be reminded that all the interpretations of the prediction profilers in Figure 1 are with respect to the water, meaning that each profile is plotted by varying the corresponding component proportion while the other nonwater components are fixed at the midpoints and only the amount of water is changed accordingly. If the other nonwater components’ proportions are fixed at other values, there are some slight changes to the plot. This can be observed from an interactive prediction profiler plot.

The individual linear, quadratic, and interaction model terms can be interpreted in the same way as other classical regression models. The only difference is that the interpretations are with respect to the filler. The significant linear effect \( \beta_i \)}
indicates significant main effect of the corresponding component. Similarly, a significant quadratic effect indicates that increasing the corresponding component while decreasing the filler the same amount, \( E(Y) \) would change in a nonlinear fashion. The interactions can also be interpreted in the usual way.

Although it is not so in this example, sometimes a nonfiller component would disappear from the SV-model after the model-term selection. In this case, we should not consider that this nonfiller component has no influence over the mixture property measured by the response. Instead, it inactively affects the mixture property as a solvent, similar to the filler, whose related effects are not contained in the model either. To further simplify the structure of the mixture, we can even combine this nonfiller component and the filler into a new filler. Such structure is known as the mixture-of-mixtures experiment (Kang, Joseph, and Brenneman 2011).

There is another hidden advantage of the SV-model. In the case described at the end of Section 2.1, if an SV-model is used then an orthogonal or optimal design can be applied on the nonfiller components resulting in a well-conditioned information matrix. In the hair product experiment, an I-optimal design (without the four replications) for the SV-model is used. The condition number (using the original scale) of the SV-model, \( \lambda \), has the smallest \( \lambda \) compared to the other models. 

Prescott et al. (2002) showed that the quadratic K-model has better conditioned information matrix. However, having the smallest \( \lambda \) cannot guarantee the smallest condition number, and the conjecture is not true as counterproved by the hair product example.

### 2.3 Literature Review and Comparisons With S-Model

The SV-model shows up frequently in the research literature. Recent papers include Cornell (2000) and Khuri (2005). The latest paper by Piepel and Landmesser (2009) provided a complete review of the SV-model. The authors summarized four situations where SV-models are often used in practice. However, the common view in the literature is that the SV-model “can be misleading in making conclusions and in developing models for response variables,” thus other mixture models such as the S-model, K-model, or component–slope linear mixture model (Piepel 2006, 2007) should be used for mixture experiments. However, as discussed earlier, the SV-model provides a simpler interpretation than the S- and K-model when the filler component is used as the slack variable, which is an advantage worth addressing. Certainly, we need to be careful not to misinterpret some model effects as being causal, so domain knowledge is very important in understanding the components properties from the SV-model.

In the S-model, the effects are confounded due to the constraint in Equation (1). Therefore, there is no way to tell if a “significant” effect is actually statistically significant or it is due to its confounding effects. For example, assume the effect \( x_i \) is “significant” for the linear S-model \( E(Y) = \sum_{i=1}^{q} \beta_i x_i \). If we change \( x_i \) to \( x_i + \Delta \), the sum of the other proportions \( \sum_{j \neq i}^{q} x_j \) has to be changed to \( \sum_{j \neq i}^{q} x_j - \Delta \). How should \( \Delta \) be divided into the rest of the \( q - 1 \) proportions? There are some

| Parameter | Complete | Reduced | Parameter | Complete | Reduced | Parameter | Complete | Reduced |
|-----------|----------|---------|-----------|----------|---------|-----------|----------|---------|
| \( \beta_0 \) | 1.1544* | 1.1471* | \( \beta_{77} \) | -0.0542 | \( \beta_{34} \) | 0.0103 |
| \( \beta_1 \) | -0.0009 | -0.0012* | \( \beta_{88} \) | -0.3826* | \( \beta_{35} \) | -0.0574 |
| \( \beta_2 \) | -0.1302* | -0.1030* | \( \beta_{12} \) | 0.0002 | \( \beta_{36} \) | 0.0086 |
| \( \beta_3 \) | -0.3239* | -0.3056* | \( \beta_{13} \) | 0.0012* | \( \beta_{37} \) | 0.0057 |
| \( \beta_4 \) | -0.1712* | -0.1547* | \( \beta_{14} \) | -0.0006 | \( \beta_{38} \) | -0.0197 |
| \( \beta_5 \) | -0.1575 | -0.1075* | \( \beta_{15} \) | 0.0001 | \( \beta_{45} \) | 0.0863 | 0.0960* |
| \( \beta_6 \) | 0.2480* | 0.2527* | \( \beta_{16} \) | 0.0002 | \( \beta_{46} \) | 0.0389* | 0.0362* |
| \( \beta_7 \) | 0.0550 | 0.1862* | \( \beta_{17} \) | 0.0000 | \( \beta_{47} \) | 0.1349* | 0.1191* |
| \( \beta_8 \) | 0.1763 | 0.1484* | \( \beta_{18} \) | -0.0004 | \( \beta_{48} \) | 0.1260* | 0.1129* |
| \( \beta_{11} \) | 0.0000 | | \( \beta_{23} \) | 0.0290 | \( \beta_{56} \) | 0.0408 |
| \( \beta_{22} \) | 0.1240* | 0.1123* | \( \beta_{24} \) | -0.0992* | \( \beta_{57} \) | 0.1252* | 0.1066* |
| \( \beta_{33} \) | 0.2393* | 0.2223* | \( \beta_{25} \) | 0.0820 | \( \beta_{58} \) | 0.0138 |
| \( \beta_{44} \) | 0.0178 | | \( \beta_{26} \) | -0.0150 | \( \beta_{57} \) | -0.0210 |
| \( \beta_{55} \) | 0.0016 | | \( \beta_{27} \) | -0.0050 | \( \beta_{58} \) | -0.0094 |
| \( \beta_{66} \) | -0.1540* | -0.1576* | \( \beta_{28} \) | 0.1067 | 0.0973* | \( \beta_{58} \) | -0.1446* | -0.1504* |

**Table 2.** Parameter estimation for the complete and reduced SV-model for hair product example

| Model | \( R^2 \) | \( R^2_{\text{adj}} \) | \( \hat{\sigma}^2 = \text{SSE/df} \) | Lack-of-fit (p-value) | \( F \)-ratio (p-value) |
|-------|---------|-----------------|-----------------|---------------------|----------------------|
| Complete | 0.9290 | 0.7727 | 0.0273/20 = 0.0014 | 2.0956 (0.2977) | 5.9436 (<0.0001) |
| Reduced | 0.8950 | 0.8437 | 0.0403/43 = 0.0009 | 1.3525 (0.4652) | 17.4489 (<0.0001) |

**Table 3.** Summaries of the complete and reduced model fits for hair product example
techniques for interpretation of the S-model, such as the Cox and Piepel effect directions (see Smith 2005, pp. 257–268), however in practice these interpretations are very difficult for practitioners. For an ordinary linear model, the mean response \( E(Y) \) should be changed to \( E(Y) + \beta_i \Delta \). But for the S-model, it is not clear how \( E(Y) \) should be changed as we need to account for how \( x_j \) for \( j \neq i \) should be changed individually. Therefore, the classic interpretation of the significant effects cannot be directly used for S-model or any other models that have equality constraints between effects. Great care is needed in interpreting the quadratic mixture effects as well.

To conclude Section 2, we recommend using the SV-model for mixture experiments when a filler component is identified and there are no lower and upper bounds (other than \([0, 1]\)) on its proportion. Meanwhile, the upper bounds \( U_i \) of the nonfiller components sum to at most 1 \( (\sum_{i=1}^{q-1} U_i \leq 1) \) and the lower bounds \( L_i \) of the nonfiller components sum must be larger than 0 \( (\sum_{i=1}^{q-1} L_i > 0) \), to avoid the case that the mixture is made purely from the filler. Under these two conditions, we can use a nonconstrained design on the nonfiller components’ proportions, which is a much easier task than the constrained mixture experimental design. It is then a natural choice to use the SV-model with the filler proportion as the slack variable and interpret effects with respect to the filler proportions. This method is called the slack variable approach in the literature.

Table 4. Correlation criterion values for Examples 1–4

| Correlation criterion | \( r_1^2 \) | \( r_2^2 \) | \( r_3^2 \) | \( r_4^2 \) |
|-----------------------|------------|------------|------------|------------|
| Example 1             | 0.3334     | 0.3334     | 0.3332     | 0.9998     |
| Example 2             | 0.9844     | 0.94213    | 0.0447     |             |
| Example 3             | 0.4930     | 0.4165     | 0.0129     | 0.8648     |
| Example 4             | 0.2739     | 0.7419     | 0.6210     |             |

3. CHOICE OF SLACK VARIABLE

If the filler ingredient is known, we can just use its component as slack variable in the SV-model. But there are cases when it is not clear which component should be the filler. The SV-model can still be used in order to alleviate the collinearity problem or to achieve more accurate predictions. The choice of the slack variable is crucial to the performance of the SV-model as shown in Cornell (2000).

If a filler component is not involved in the experiment, then it does not matter which component is used as the slack variable in the sense of interpretation. Therefore, we should choose the slack variable for the sake of numerical stability and prediction accuracy. Here, we only focus on the numerical stability of the model, since it is well known that model-term selection is more accurate for models with less collinearity.

Table 5. maxVIFs, MVIFs, and condition numbers for Examples 1–4 with original scales

| Model | \( SV_{x_1} \)     | \( SV_{x_2} \)     | \( SV_{x_3} \)     | \( SV_{x_4} \)     |
|-------|-------------------|-------------------|-------------------|-------------------|
| Example 1  | maxVIF | 425341.74 | 425341.75 | 434077.28 | 66.29 |
|         | MVIF   | 156755.7   | 156755.7 | 158035.2 | 35.74 |
|         | CN     | 222626     | 222626 | 223704 | 30037 |
| Example 2  | maxVIF | 27.83     | 124.85    | 64295.03 |             |
|         | MVIF   | 19.24      | 68.71     | 69758.25 |             |
|         | CN     | 1161       | 1493      | 3831   |             |
| Example 3  | maxVIF | 9152.53    | 38857.22  | 43353298 | 2701.84 |
|         | MVIF   | 11610      | 10439.87  | 133371201 | 853.97 |
|         | CN     | 72937      | 72901     | 391490 | 66420 |
| Example 4  | maxVIF | 376.16     | 63.35     | 106.36 |             |
|         | MVIF   | 202.40     | 42.62     | 68.34  |             |
|         | CN     | 270        | 152       | 184   |             |
To quantify the collinearity, two criteria are commonly used in the literature. The first one is the condition number of the information matrix. Let $\lambda_{\max}$ and $\lambda_{\min}$ be the maximum and minimum eigenvalues of the positive definite matrix $F'F$. The 2-norm conditional number (CN) is $\text{cond}(F'F) = \sqrt{\lambda_{\max}/\lambda_{\min}}$. The other criterion is the variance inflation factor (VIF). The VIF associated with the regression coefficient $\beta_j$ for the $j$th column in $F$ is given by $\text{VIF}(\beta_j) = (1 - R_j^2)^{-1}$, where $R_j^2$ is the coefficient of determinant for regressing $F(\cdot, j)$ against all the other columns $F(\cdot, -j)$. Here, the notation $A(\cdot, k)$ denotes the $k$th column of $A$ and $A(\cdot, -k)$ denotes $A$ without the $k$th column. The larger the VIF is, the worse the collinearity is for the effect $\beta_j$. To evaluate the overall collinearity level of a model, we propose the maximum VIF (maxVIF) and the mean variance inflation factor (MVIF),

$$\text{maxVIF} = \max_{j=1,\ldots,k} \text{VIF}(\beta_j) \quad \text{and} \quad \text{MVIF} = \frac{1}{k} \sum_{j=1}^{k} \text{VIF}(\beta_j),$$

where $k$ is the number of effects in the model excluding the intercept. For the same number of components and the fixed order of the SV-model, using any component proportion as the slack variable results in SV-models with the same number of effects. This allows us to compare the maxVIF and MVIF for each SV-model. Both maxVIF and MVIF are often used as indicators of the severity of collinearity. It can be shown that the expected value of the sum of the squared errors of the estimated coefficients $\hat{\beta}_j$’s (the standardized version) is proportional to MVIF, and the largest VIF is proportional to the largest squared error of the estimated coefficient (Kutner et al. 2004).

To choose the best SV-model using different components as slack variables, one can directly construct each possible SV-model and choose the best one according to the aforementioned criteria. It is not a difficult task if there are not many components in the mixture. But this is not an ideal method if there are a large number of components. To avoid the extra computation, we propose a more direct criterion that is only based on the design of the mixture experiment.

### 3.1 Correlation Criterion

Denote $X$ as the design matrix for the mixture experiment of total $q$ components and $n$ experimental settings, and thus $x_{ij}$ is the $j$th component’s proportion of the $i$th run. Define $r_{ij}$ as the correlation between the $i$th and $j$th columns of $X$, and calculate it by

$$r_{ij} = \frac{(X(i, i) - \bar{X}_i)(X(j, j) - \bar{X}_j)}{\sqrt{(X(i, i) - \bar{X}_i)\left(\sum_{q} X(i, q) - \bar{X}_i\right)(X(j, j) - \bar{X}_j)\left(\sum_{q} X(j, q) - \bar{X}_j\right)}},$$

where $X(i, j)$ is the $i$th column of $X$, $\bar{X}_i$ is the sample mean of $X(i, \cdot)$, and $I$ is a column of 1’s. To evaluate whether the $i$th column is collinear with any other columns, we can use its average squared correlation with all the other $q - 1$ columns. Denote it as $r_i^2$, and it can be calculated by

$$r_i^2 = \frac{\sum_{j \neq i} r_{ij}^2}{q - 1} = \frac{(X(i, j) - \bar{X}_i)\text{EX}(X(\cdot, j))\text{diag}(\text{EX}(X(\cdot, j))^{-1}X(\cdot, j)\text{EX}(X(\cdot, j)))}{(q - 1)\text{EX}(X(\cdot, j))\text{EX}(X(\cdot, j))},$$

where $\text{E} = 1 - 1/n1'1'$ and $I$ is the $n \times n$ identity matrix. We choose the component that has the largest $r_i^2$ as the slack variable.

The intuition behind this criterion is very straightforward. If the $i$th component has the largest $r_i^2$, it indicates that $X(i, \cdot)$ is mostly correlated with some or all the other columns in $X$ and thus it is likely to be heavily correlated with the quadratic order columns too. If we include this component in the model, we should expect the worst collinearity especially between the $i$th component’s effects (linear or second-order) and all the other effects. This intuition can be explained using a simple 3-component mixture and the linear SV-model.

### Table 6. maxVIFs, MVIFs, and condition numbers for Examples 1–4 with designs scaled in [−1,1]

| Model         | $SV_{s1}$ | $SV_{s2}$ | $SV_{s3}$ | $SV_{s4}$ |
|---------------|-----------|-----------|-----------|-----------|
| Example 1     | maxVIF    | 45.76     | 45.76     | 45.76     |
|               | MVIF      | 138.39    | 138.39    | 138.39    |
|               | CN        | 2544.66   | 2544.66   | 2544.66   |
| Example 2     | maxVIF    | 1.06      | 1.42      | 2848.66   |
|               | MVIF      | 5.19      | 6.06      | 4414.70   |
|               | CN        | 21.37     | 29.80     | 61093.43  |
| Example 3     | maxVIF    | 80.17     | 737.50    | 5628574.2 |
|               | MVIF      | 336.05    | 1675.75   | 20805871  |
|               | CN        | 6068.54   | 31659.86  | 371990245 |
| Example 4     | maxVIF    | 25.00     | 6.31      | 12.96     |
|               | MVIF      | 77.97     | 18.11     | 40.27     |
|               | CN        | 548.13    | 71.25     | 234.54    |

### Table 7. maxVIFs, MVIFs, and CNs for a different model using a different transformation

| Model         | $[-1,1]$ Scale | L-Pseudocomponent |
|---------------|----------------|-------------------|
|               | maxVIF | MVIF | CN  | maxVIF | MVIF | CN |
| SV-model      | 1.49   | 1.26 | 19.50 | 35.54 | 12.80 | 1.87E+10 |
| S-model       | 55070.61 | 8743.32 | 860907.8 | 358201.8 | 61252.93 | 4.32E+11 |
| K-model       | 1410.01 | 456.34 | 9.59E+8 | 23.12 | 12.04 | 8.02E+9 |
| U-Pseudocomponent | maxVIF | MVIF | CN | maxVIF | MVIF | CN |
| SV-model      | 36.21  | 12.84 | 1.74E+10 | 22.22 | 2.67 | 3.40E+9 |
| S-model       | 341974 | 58942.4 | 3.64E+11 | 127159.3 | 12015.9 | 3.19E+10 |
| K-model       | 23.51  | 12.05 | 6.42E+9 | 552544.9 | 33316.2 | 2.20E+12 |
If we use the 3rd component as slack variable, the SV-model is

\[ E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \] (6)

The MVIF for this model is

\[ \text{MVIF} = \frac{1}{2} (\text{VIF}(\beta_1) + \text{VIF}(\beta_2)) = \frac{1}{2} \left( \frac{1}{1 - R_1^2} + \frac{1}{1 - R_2^2} \right). \]

Here, \( R_1^2 \) is the coefficient of determination for the simple linear regression equation \( x_1 = \alpha_0 + \alpha_1 x_2 \). We know that for simple linear regression \( R_1^2 = r_{12}^2 \). So, the MVIF for Equation (6) is

\[ \text{MVIF}_1 = \frac{1}{2} \left( \frac{1}{1 - r_{12}^2} + \frac{1}{1 - r_{23}^2} \right) = \frac{1}{1 - r_{12}^2}. \]

In the same way, we can derive the MVIF values for SV-models using the 1st and 2nd component as slack variable, respectively

\[ \text{MVIF}_1 = \frac{1}{1 - r_{12}^2}, \quad \text{MVIF}_2 = \frac{1}{1 - r_{23}^2}. \]

Without loss of generality, let \( r_2^2 \) be the largest compared to \( r_1^2 \) and \( r_3^2 \) meaning that \( r_{12}^2 \) is smaller than both \( r_{23}^2 \) and \( r_{31}^2 \). Thus MVIF3 is the smallest among all and Equation (6) has the least collinearity. Apparently, the SV-model represented by Equation (6) also has the smallest maxVIF value which is equal to \( \text{VIF}(\beta_1) = \text{VIF}(\beta_2) = 1/(1 - R_{12}^2) \). This shows that using the component with the largest \( r_2^2 \) as the slack variable can lead to the SV-model with the smallest collinearity, and thus the most numerically stable model. This proof cannot be extended to general mixture experiment settings. Thus, next we use four numerical examples to show that this criterion can effectively pick the best slack variable and lead to the most numerically stable SV-model.

### 3.2 Four Examples

In the following, we describe four examples that are chosen from the literature. In Table 4, we compute the correlation criterion for each component for all examples. It can be seen in Table 5 that choosing the component with the largest correlation criterion to be slack variable is consistent with using criteria maxVIF, MVIF, and CN.

**Example 1.** Artificial Example in Piepel and Landmesser (2009): Piepel and Landmesser (2009) used an artificial example on mixtures of two drugs \((x_1 \text{ and } x_2)\), an enhancer \((x_3)\), and a filler \((x_4)\), where 0.01 \(\leq x_1 \leq 0.03\), 0.01 \(\leq x_2 \leq 0.03\), and 0 \(\leq x_3 \leq 0.02\). An 18-point, face-centered cube design was used which contains 8 factorial points, 6 face centroids, and 4 center point replicates.

**Example 2.** Artificial Example in Prescott et al. (2002): Prescott et al. (2002) described an experiment in which 0.4 \(\leq x_1 \leq 0.945\), 0.05 \(\leq x_2 \leq 0.5\), and 0.005 \(\leq x_3 \leq 0.1\). The 13-Point D-optimal design was generated by software DesignExpert®.

**Example 3.** Pharmaceutical Solution Example in Cornell (2000): Cornell (2000) described a study where the solubility of butoconazole nitrate, an antifungal agent, was studied as a function of the proportions of the cosolvents polyethylene glycol 400 \((x_1)\), glycerin \((x_2)\), polysor polysorbate 60 \((x_3)\), along with water \((x_4)\). Constraints on the component proportions were

\[
0.10 \leq x_1 \leq 0.40, \quad 0.005 \leq x_3 \leq 0.03, \\
0.10 \leq x_2 \leq 0.40, \quad 0.30 \leq x_4 \leq 0.795.
\]

A 10-point D-optimal design was selected for fitting a quadratic model. The design consisted of 6 of the 10 extreme vertices and midpoints of 4 of the edges of the constraints region.

**Example 4.** Vapor Pressure Example in Cornell and Gorman (2003): Cornell and Gorman (2003) showed a study involving three components and seven design points in the reduced region constrained by the inequalities 0.15 \(\leq x_1 \leq 0.5\), 0.2 \(\leq x_2 \leq 0.7\), and 0.15 \(\leq x_3 \leq 0.65\). In this example, the component \(x_2\) is water and would naturally be considered the filler. But the design of the experiment does not use the slack variable approach. Instead, the design is a transformed simplex-centroid design.

All four examples show that the largest criterion \( r_2^2 \) chooses the most suitable slack variable. Unintentionally, all four examples include a filler component and we note that the criterion chooses the filler as slack variable in all examples. Example 1 clearly uses the slack variable approach where a face-centered cube design is used for the nonfiller components and the filler takes the rest of the proportion. Examples 2 and 3 use the software-generated D-optimal designs for the quadratic S-model. Since the S-model and the SV-model are equivalent, the D-optimal designs for S- and SV-model are the same, as shown in Proposition 1 and its corollary in Section 4. Example 4 uses transformed simplex-centroid design that is also a D-optimal design for a quadric model.

To analyze mixture experiments, it is suggested to perform some linear transformation on the components’ proportions to reduce the ill-conditioning problem. The correlation criterion is scalefree so transformation will not change the values of \( r_i^2 \)’s. While the transformation will change the absolute values of maxVIF, MVIF, and CN, it will not change the sorted orders of the criteria of all the SV-models if the same transformation is used (see Table 6).

### 4. COMPARISON WITH OTHER MODELS

In this section, we compare the S-, SV-, and K-model that are commonly used in mixture experiments. Previously, we propose that the SV-model is more appealing in terms of interpretation. In this section, we focus on the numerical stability and prediction accuracy of the mixture models. First, we use two examples to show the criteria for comparing the three models.

| Model               | Size | \( R^2 \) | \( R^2_{adj} \) | \( \hat{\sigma} \) | LOOCV | 13-fold CV |
|---------------------|------|-----------|----------------|-----------------|-------|------------|
| SV-model *          | 20   | 0.8928    | 0.8475         | 0.03025         | 0.0221| 0.0225     |
| S-model *           | 34   | 0.9058    | 0.8056         | 0.03416         | 0.0255| 0.02974    |
| K-model             | 45   | 0.9290    | 0.7727         | 0.03693         | 0.04493| 0.04496    |

Table 8. Comparison of the reduced SV-model, reduced S-model, and full K-model.
4.1 Numerical Stability and Prediction Accuracy Examples

As suggested in the literature, we use linear transformation to alleviate the numerical instability. But different transformation methods work the best for different mixture models. For the SV-model, we can scale the designs of the proportions into [–1, 1], which is the typical scale used in classical design and analysis of experiments. For the S- and K-model, L-, U-, and modified-pseudocomponent transformations are recommended according to the literature. Thus, we try all four transformation methods and choose the best one for each mixture model. The model-term selection process is done by backward stepwise regression based on both AIC and BIC. For the two examples, the reduced S- and SV-model are the same by both criteria.

Example 5. Hair Product Example: For the hair product experiment that involves an additive \( x_1 \) and three lubricant blends \( x_2, x_3, \) and \( x_4 \). The component proportions need to satisfy

\[
0.07 \leq x_1 \leq 0.18, \quad 0.37 \leq x_3 \leq 0.70, \\
0 \leq x_2 \leq 0.30, \quad 0 \leq x_4 \leq 0.15.
\]

Later, John (1984) used these data to show how ridge regression can reduce the ill-conditioning problem for the S-model. Based on Table 9, scaling the components’ proportions into [–1, 1] range works the best for all three models. Using this transformation, we compare the reduced SV-, reduced S-, and full K-model. The comparisons of the model size, \( R^2 \), \( R_{adj}^2 \), and leave-one-out and 5-fold cross-validation prediction errors are shown in Table 10. It also confirms that the reduced SV-model is the most parsimonious, the best fit, and the most accurate model.

4.2 Simulation Study

We conduct a simulation study to compare the prediction performance of the SV-, S-, and K-model under different mixture experiment settings. In order to visualize the experimental designs, we only assume a three-component mixture. We use four different settings on the constraints of the component proportions.

Case 1: \( 0 \leq x_1 \leq 0.1, 0 \leq x_2 \leq 0.2, \) and \( 0 \leq x_3 \leq 1 \),
Case 2: \( 0 \leq x_1 \leq 0.4, 0 \leq x_2 \leq 0.5, \) and \( 0 \leq x_3 \leq 1 \),
Case 3: \( 0 \leq x_1 \leq 0.5, 0 \leq x_2 \leq 0.7, \) and \( 0 \leq x_3 \leq 1 \),
Case 4: \( 0 \leq x_i \leq 1 \) for \( i = 1, 2, 3 \).

In Cases 1–3, the component \( x_3 \) can be seen as the filler since it does not have upper or lower bounds other than 0 and 1. But only in Cases 1 and 2, the component proportions \( x_1 \) and \( x_2 \) can

| Table 9. maxVIFs, MVIFs, and CNs for a different model using a different transformation |
|----------------------------------|------------------|--------|
| Model               | maxVIF | MVIF  | CN    |
| [–1, 1] Scale        |        |       |       |
| SV-model            | 66.59  | 25.12 | 2310.30 |
| S-model             | 126.18 | 52.05 | 161.23  |
| K-model             | 85.63  | 52.05 | 1994.48 |
| Modified L-Pseudocomponent |        |       |       |
| SV-model            | 9.69   | 3.06  | 52.95  |
| S-model             | 14.01  | 4.20  | 91.34  |
| K-model             | 10.15  | 3.75  | 37.56  |

| Table 10. Comparison of the reduced SV-, reduced S-model, and full K-model |
|------------------|------|------------------|-----|-------|-----|-----|
| Model            | Size | \( R^2 \)          | \( R_{adj}^2 \) | \( \hat{\sigma} \) | LOOCV | 5-fold CV |
| SV-model         | 8    | 0.8557           | 0.7114 | 51.34  | 50.96 | 52.17 |
| S-model          | 9    | 0.8593           | 0.6717 | 54.76  | 57.76 | 94.06 |
| K-model          | 9    | 0.8596           | 0.5789 | 59.91  | 66.86 | 95.89 |

Example 6. Additive in Lubricants Experiment Example in Snee (1975) and John (1984); Snee (1975) showed a mixture experiment that involves an additive \( x_1 \) and three lubricant blends \( x_2, x_3, \) and \( x_4 \). The component proportions need to satisfy

\[
0.07 \leq x_1 \leq 0.18, \quad 0.37 \leq x_3 \leq 0.70, \\
0 \leq x_2 \leq 0.30, \quad 0 \leq x_4 \leq 0.15.
\]
Experimental Design

R-square

Adjusted-R-square

leave-one-out CV error

6-fold CV error

Figure 2. Simulation study for mixture experiment $0 \leq x_1 \leq 0.1$ and $0 \leq x_2 \leq 0.2$: $R^2$, $R^2_{adj}$, $\hat{\sigma}$, the leave-one-out cross-validation prediction error and 6-fold cross-validation prediction error.

be varied independent of each other since the sum of their upper bounds is less than 1. Although Case 3 still contains the filler, $U_1 + U_2 > 1$ and thus the design on $x_1$ and $x_2$ is not constraint-free. But we can still use the SV-model with $x_3$ as the slack variable. Case 4 is the classic mixture experiment setting with no additional restrictions on the upper and lower bounds. Thus any component can be treated as the filler.

The 12-run D-optimal designs for the four cases are shown in Figures 2–5 and are generated by JMP Pro 10®. The constrained region for the first two settings is a parallelogram and thus the D-optimal design contains the centroid, the four vertices, and the four midpoints of the edges, which is also the central composite design. Three extra design points are replicated at three vertices marked in the plots. The D-optimal design for Case 3 contains five vertices, two midpoints on two edges, and one centroid. Three vertices are replicated twice. For Case 4, the D-optimal design is the simplex design with each design point replicated twice.

We use the quadratic S-model $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \varepsilon$ as the underlying true response function. To simulate the response, we randomly generate the true coefficients from $\beta_i \sim \text{Uniform}(0, 5)$ for $i = 1, 2, 3$ and $\beta_i \sim \text{Uniform}(0, 1)$ for $i = 4, 5, 6$, which is consistent with the hierarchical ordering principle (Wu and Hamada 2011). The noise follows $\varepsilon \sim N(0, 0.5^2)$ which is normal distribution with 0 mean and standard deviation 0.5. We simulate $B = 100$ datasets from these distributions. For each dataset, we first fit the full SV-, S-, and K-model and then use backward elimination with BIC to obtain the reduced the SV-, reduced S-, and full K-model. The densities of the five criteria $R^2$, $R^2_{adj}$, $\hat{\sigma}$, leave-one-out, and 6-fold cross-validation prediction error of the reduced SV-, reduced S-, and full K-model are shown in Figures 2–5.
The first two cases forming the parallelograms are more challenging for modeling as the $R^2$ and $R_{adj}^2$ are relatively small and widely spread, while the two criteria are much more improved for the Cases 3 and 4. For Cases 1 and 2, the density plots of $\hat{\sigma}$ and two cross-validation errors show that the reduced SV-model has slightly better modeling precision and more accurate predictions than the reduced S-model. This advantage is less obvious in Case 3 and even disappears in Case 4. The full K-model has the worst prediction accuracy for all cases.

It should be noted that the simulation results returned in Figures 2–5 are only based on backward elimination with BIC. In the supplement, we use other model-term selection methods, including backward elimination with AIC, Lasso, and LAR procedure (Hastie et al. 2009). The latter two are very popular in data mining field. All of the simulation results show that different model-term selection methods perform similarly. Thus, the conclusions we have summarized can be applied to other similar model building methods. In the supplement, we conduct another simulation study for a 10-component mixture experiment, where the four model-term selection methods are used as well. Besides these four cases, we have also studied other scenarios. For example, we have studied the 4-component mixture case. We have also assumed the Scheffé cubic model as the true response, different distributions for the true coefficients in the model, and different constraint settings. We omit these results from the supplement due to space limit. From all these simulation studies, we can draw the following conclusions for mixture experiment with small or relative larger number of components.

1. Model-term selection should be performed to obtain more predictive models, which produce more accurate predictions. The K-model has the worst prediction accuracy since model-term selection cannot be used for the K-model.
2. If all the components are symmetric with respect to each other in both constraints and design, the SV- and S-model perform similarly, and there is no difference choosing any component proportion as slack variable.

3. If there is a filler component \( x_q \in [0, 1] \), but \( \sum_{i=1}^{q-1} U_i > 1 \), then we cannot use a nonconstrained design for the nonfiller components. The SV-model is only slightly more accurate than the S-model in these situations.

4. If there is a filler component \( x_q \in [0, 1] \), and \( \sum_{i=1}^{q-1} U_i \leq 1 \), then the SV-model would be more precise and more predictive than the corresponding S- and K-models.

### 4.3 Numerical Stability

As shown in Section 4.1, apart from the choice of model forms, there can be other factors affecting the numerical stability of the information matrix. The design of the mixture experiment can be an important factor because it appears that the optimal design based on a given model form certainly favors that model compared to all the other model forms. But the following proposition and corollary show that this intuition does not hold for all optimal designs.

**Proposition 1.** Based on the same design \( X \) for a mixture experiment, the model matrices of the S-model \( F_1 \) and the equivalent SV-model \( F_2 \) satisfy \( F_1 = F_2 P \), where \( P \) is an elementary matrix and \(| \det(P) | = 1 \), and thus \( \det(F_1) = \det(F_2) \).

**Corollary 1.** Based on the same design \( X \) for a mixture experiment, Proposition 1 is true for the model matrices of any equivalent mixture models.

If the D-optimal design criterion is used (as in Section 4.2), it does not matter which model form we assume. As long as
the models have the same order and the same number of terms (counting the intercept), the D-optimal criterion value remains the same for different models. Also, from Corollary 1, we can see that for any two equivalent models their model matrices can be transformed from one to the other by $F_1 = F_2 P$. Since $F_1' F_1 = P' F_2' F_2 P'$, the difference between the condition numbers of the information matrices is only caused by $P$. The elementary matrix $P$ is fixed for any two equivalent model forms. But it is impossible to show that $P$ can always enlarge or reduce the condition number of $F_2' F_2$ because it also depends on $F_2$ which, in turn, depends on the experimental design. For the examples in Section 4.2, the matrix $P$ transforming the quadratic SV-model to the quadratic S-model is the same but it does not always increase or decrease the condition number for the S-model. For the simplex design in Case 4, $P$ improves the condition number, as shown in Table 11 but not so for Cases 1–3. To compute Table 11, the model matrix is based on the original scale of the proportions in $[0, 1]$.

Besides the elementary matrix $P$, different transformations of the component proportions can also change the conditioning of the information matrix. In Table 12 the condition numbers for Cases 1–4 are computed after the best transformation method is applied. We consider only the four transformation methods mentioned in Section 4.1. Comparing Tables 11 and 12, it is obvious that transformation is helpful to improve the condition number. Interestingly, the K-model has the smallest condition number.

Table 11. Condition numbers (without transformation) of the S-, SV, and K-model for Cases 1–4

| Models | Case 1    | Case 2    | Case 3    | Case 4    |
|--------|-----------|-----------|-----------|-----------|
| S-model | 3,370,746 | 7162      | 2446      | 86        |
| SV-model | 644,918   | 3955      | 2101      | 1482      |
| K-model | 386,252   | 1090      | 463       | 29        |
under the original scale for Case 1–3. But, after transformation, the SV-model has the smallest condition number for Cases 1–3. Also, for the SV-model, the \([-1, 1]\) scaling consistently works the best for all four simulation cases, and for Examples 5 and 6. However, Case 4 is special as a simplex design is used and the design is symmetric to all three components. In Case 4, both the S- and K-models have smaller CN values than the SV-model.

Due to these reasons, it is not possible to conclude that one model form always has the smallest condition number for all cases, which is why we have used examples and simulations to illustrate cases when the SV-model has better numerical stability than the other two models. We summarize the following results from the simulation (and which cannot be generalized further).

1. \([-1, 1]\) scaling can effectively improve the conditioning of the SV-model.
2. If all the components are symmetric of each other in both constraints and design, the SV-model has the worst conditioning compared with S- and K-model.
3. If there is a filler component \(x_\ell \in [0, 1]\), the SV-model after \([-1, 1]\) scaling is numerically more stable than the S- and K-model with their best transformation methods applied, respectively, for any optimal design criterion.

5. CONCLUSION

In this article, we carefully study the properties of the SV-model and compare the SV-model with the S-model and K-model on three aspects: interpretation, numerical stability, and prediction accuracy. We recommend practitioners to use the slack variable approach when there is a filler component whose proportion is in [0, 1] and the sum of the upper bounds of the nonfiller proportions is no larger than 1 due to the following reasons. (1) The experimental design on the nonfiller components is not constrained and thus much simpler than the constrained mixture experimental design. (2) The interpretation of the SV-model is much easier and clearer for the experimenter to understand. (3) Model-term selection on the SV-model is not restricted and the reduced SV-model can lead to better prediction accuracy than the S- and K-models.

When using the SV-model one can choose the slack variable through expert knowledge (often formulatores will have a filler ingredient already identified) or by using the correlation criterion proposed in Section 3. A reasonable transformation should be used on the design to improve the numerical stability. By simply scaling the proportions into \([-1, 1]\) for the SV-model, we have a more effective transformation than the other transformation methods commonly used for the S-model and the K-model.

### APPENDIX: PROOF OF PROPOSITION 1 AND COROLLARY 1

Without loss of generality, we assume \(x_q\) to be the slack variable. For the linear S- and SV-model, \(F_1 = X\) and \(F_2 = [X(1,1), \ldots, X(q-1,1)]\). Replacing the component \(x_q = 1 - \sum_{i=1}^{q-1} x_i\), then \(F_1 = [X(1,1), \ldots, X(q-1,1), 1 - \sum_{i=1}^{q-1} X(i,1)] = F_1P\), where \(P\) is an elementary matrix

\[
P = \begin{bmatrix}
1 & 0 & 0 & \ldots & -1 \\
0 & 1 & 0 & \ldots & -1 \\
0 & 0 & 1 & \ldots & -1 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{bmatrix}.
\]

Obviously \(\det(P) = 1\), and thus \(\det(F_1F_1) = \det(P^T F_1^T F_1 P) = \det(P) \det(F_1^T F_1) \det(P) = \det(F_1^T F_1)\). For quadratic S- and SV-models, replacing the component variable \(x_q = 1 - \sum_{i=1}^{q-1} x_i\), the vector of the terms in the S-model becomes

\[
\tilde{f}_1(x) = \left[ x_1, \ldots, x_{q-1}, 1 - \sum_{i=1}^{q-1} x_i, x_1, \ldots, x_{q-1}, x_q \right]^T
\]

\[
= \begin{bmatrix}
x_1, \ldots, x_{q-1}, 1 - \sum_{i=1}^{q-1} x_i, x_1, \ldots, x_{q-1}
\end{bmatrix}^T
\]

\[
\times \begin{bmatrix}
1 - \sum_{i=1}^{q-1} x_i
\end{bmatrix}
\]

\[
= [1, x_1, \ldots, x_{q-1}, x_1, \ldots, x_{q-1}, x_q, x_1^2, \ldots, x_q^2, \ldots, x_1^{q-1}, x_q^{q-1}]^T P
\]

\[
= f_2(x) P
\]

where \(f_2(x)\) is the vector of the terms in the equivalent SV-model, and \(P\) is the elementary matrix that performs elementary column operations. (1) Exchange the positions of certain elements in \(f_2(x)\). (2) Multiply certain elements in \(f_2(x)\) by \(-1\). (3) Add certain elements in \(f_2(x)\) by \(\sum k_i e_i\), with \(k_i\) can only be 1 or \(-1\), and \(e_i\) are some elements of \(f_2(x)\). Then, \(f_1(x) = f_2(x) P\). The model matrices are \(F_1 = [f_1(x_1), f_1(x_2), \ldots, f_1(x_q)]\) for \(l = 1, 2\), so \(F_1 = F_2 P\). What’s more, \(P\) is always an upper diagonal matrix with diagonal entries equal to 1 or \(-1\). So, \(\det(P) = 1\) or \(-1\) and \(\det(F_1F_1) = \det(F_1^T F_1)\). For higher order or other forms of the mixture models, the proofs are similar. \(\square\)

### SUPPLEMENTARY MATERIALS

In the supplement we provide more simulation study results to support the conclusions drawn in the article. We compare different variable selection methods for the simulation Cases 1–4. We also show another simulation case where there are ten components in the mixture. These results are consistent with the conclusions stated in Sections 4 and 5.

### ACKNOWLEDGMENTS

This research was supported by the Procter & Gamble Statistics Innovation Grant from the Procter & Gamble Company. We thank the Editor, Associate Editor, and two referees for the valuable comments and suggestions. We also thank Robert Wells, retired Procter & Gamble Hair Care Principal Scientist, for motivating this work by being the first practitioner (many others followed) to question the need for the S-model based on many years of successfully using the SV-model when a large filler component is available.

[Received January 2014. Accepted October 2014.]

## Table 12. Condition numbers of the S-, SV, and K-model for Cases 1–4 after transformation

| Models  | Case 1 | Case 2 | Case 3 | Case 4 |
|---------|--------|--------|--------|--------|
| S-model | 695    | 131    | 154    | 30     |
| SV-model| 15     | 15     | 22     | 210    |
| K-model | 2441   | 523    | 463    | 29     |
REFERENCES

Cornell, J. A. (2000), “Fitting a Slack-Variable Model to Mixture Data: Some Questions Raised,” *Journal of Quality Technology*, 32, 133–147. [258,259,261]

——— (2002), *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data* (3rd ed.), New York: Wiley. [256]

Cornell, J. A., and Gorman, J. W. (2003), “Two New Mixture Models: Living with Collinearity But Removing Its Influence,” *Journal of Quality Technology*, 35, 78–88. [256,261]

Goldfarb, H. B., Anderson-Cook, C. M., Borror, C. M., and Montgomery, D. C. (2004), “Fraction of Design Space Plots for Assessing Mixture and Mixture-Process Designs,” *Journal of Quality Technology*, 36, 169–179. [255]

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The Elements of Statistical Learning* (2nd ed.), New York: Springer. [256,264]

John, R. C. S. T. (1984), “Experiments with Mixtures, Ill-Conditioning, and Ridge Regression,” *Journal of Quality Technology*, 16, 81–96. [262]

Kang, L., Joseph, R. V., and Brenneman, W. (2011), “Design and Modeling Strategies for Mixture-of-Mixtures Experiments,” *Technometrics*, 53, 125–136. [258]

Khuri, A. I. (2005), “Slack-Variable Models Versus Scheffé’s Mixture Models,” *Journal of Applied Statistics*, 32, 887–908. [258]

Kowalski, S. M., Cornell, J. A., and Vining, G. G. (2002), “Split-Plot Designs and Estimation Methods for Mixture Experiments With Process Variables,” *Technometrics*, 44, 72–79. [255]

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). *Applied Linear Statistical Models* (5th ed.), Irwin: McGraw-Hill. [260]

Montgomery, D. C., and Voth, S. E. (1994), “Multicollinearity and Leverage in Mixture Experiments,” *Journal of Quality Technology*, 26, 96–108. [256]

Piepel, G. F. (2006), “A Note Comparing Component-Slope, Scheffé and Cox Parameterization of the Linear Mixture Experiment Model,” *Journal of Applied Statistics*, 33, 397–403. [258]

——— (2007), “A Component Slope Linear Model for Mixture Experiments,” *Quality Technology and Quantitative Management*, 4, 331–343. [258]

Piepel, G. F., Cooley, S. K., and Jones, B. (2005), “Construction of a 21-Component Layered Mixture Experiment Design Using a New Mixture Coordinate-Exchange Algorithm,” *Quality Engineering*, 17, 579–594. [256]

Piepel, G. F., and Cornell, J. A. (1985), “Models for Mixture Experiments When the Response Depends on the Total Amount,” *Technometrics*, 27, 219–227. [255]

Piepel, G. F., and Landmesser, S. M. (2009), “Mixture Experiment Alternatives to Slack Variable Approach,” *Quality Engineering*, 21, 262–276. [258,261]

Prescott, P., Dean, A. M., Draper, N. R., and Lewis, S. M. (2002), “Mixture Experiments: Ill-Conditioning and Quadratic Model Specification,” *Technometrics*, 44, 260–268. [255,256,258,261]

Smith, W. F. (2005), *Experimental Design for Formulation*, ASA-SIAM Series on Statistics and Applied Probability, Philadelphia, PA: ASA-SIAM. [256,259]

Snee, R. D. (1975), “Experimental Designs for Quadratic Models in Constrained Mixture Spaces,” *Technometrics*, 17, 149–159. [262]

Wu, C. F. J., and Hamada, M. S. (2011). *Experiments: Planning, Analysis, and Optimization* (Vol. 552), New York: Wiley. [263]