Compatibility of conditionally specified models

Hua Yun Chen
Division of epidemiology & Biostatistics
School of Public Health
University of Illinois at Chicago
1603 West Taylor Street, Chicago, IL 60612

Abstract

A conditionally specified joint model is convenient to use in fields such as spatial data modeling, Gibbs sampling, and missing data imputation. One potential problem with such an approach is that the conditionally specified models may be incompatible, which can lead to serious problems in applications. We propose an odds ratio representation of a joint density to study the issue and derive conditions under which conditionally specified distributions are compatible and yield a joint distribution. Our conditions are the simplest to verify compared with those proposed in the literature. The proposal also explicitly construct joint densities that are fully compatible with the conditionally specified densities when the conditional densities are compatible, and partially compatible with the conditional densities when they are incompatible. The construction result is then applied to checking the compatibility of the conditionally specified models. Ways to modify the conditionally specified models based on the construction of the joint models are also discussed when the conditionally specified models are incompatible.

KEY WORDS: Density decomposition; Density representation; Odds ratio function; Semiparametric models.
1 Introduction

A joint model of data is often desired in many statistical applications, especially in Bayesian inference. However, when the dimension of the data is high, specifying a joint model that captures many features of the data can be much harder than specifying conditional models that capture the different features separately. See for example Besag (1974, 1994) for the spatial data modeling, Hobert and Casella (1998) for Gibbs sampling, and Van Buuren, Boshuizen, and Knook (1999), Van Buuren (2007) and Raghunathan, et al. (2001) for the missing data imputation. One issue with the conditional specification approach is that a joint model that is compatible with all the specified conditional models may not exist. The incompatibility of the conditionally specified models may lead to serious issues on the statistical inference and interpretation in the spatial data analysis, on the convergence of the Gibbs sampling, and on the validity of the imputed values.

The study of the compatibility of conditionally specified models is usually done through the study of the compatibility of the conditionally specified densities. The first issue in the study is to check the compatibility of the densities. This issue has been extensively studied by Besag (1974), Arnold and Press (1989), Hobert and Casella (1998), and more recently Wang and Ip (2008) among others. Besag (1974) concentrated on the form of joint densities that satisfy specific conditional dependence requirements. Arnold and Press (1989) and Arnold and Gokhale (1994) mainly dealt with the case with two variables. Arnold, Castillo, and Sarabia (2001) and discussants gave an excellent introduction to the problem and related issues. Gorieroux and Montfort (1979) discussed the uniqueness of the determined joint distribution. Hobert and Casella (1998) studied the issue in Gibbs sampling and the impact of the incompatibility on the convergence of the Gibbs sampler. Most of the results obtained on checking compatibility are related to the density ratios formulated in Besag (1974), which can be cumbersome when the dimension of the data is moderate to high. Wang and Ip (2008) simplified the conditions for checking compatibility based on the dependence functions rather than density ratio. The second issue in the study is on the modification of the incompatible conditionally specified models so that the modified conditionally specified models do determine a proper joint model for the data. This issue was less discussed in the literature.

In this article, we propose a new way to study the compatibility of conditionally specified densities through the odds ratio representation of a joint density. The odds ratio function was used
jointly with the marginal densities to study bivariate distributions by Osius (2005). Arnold and Gokhale (1994) also implicitly used it through the log-linear model representation of the distribution of a $r \times c$ contingency table. It is unclear how to generalize their approaches to handle high dimensional distributions. Joe (1997, chapter 9) gave a closest result to what we obtained in this paper on the compatibility check. However, like results obtained by others, his result is also based on the density ratio and does not yield a concise expression for the joint density. Our representation differs from those in the literature in that we use odds ratio functions along with conditional densities at a fixed condition to represent the density. The representation is very easy to be generalized to high dimensional problems. We obtain necessary and sufficient conditions for the compatibility in terms of odds ratio functions. The conditions we obtained resemble those of Hobert and Casella (1998) and Wang and Ip (2008), but are simpler and much more transparent. We show that it is sufficient to verify only a few selected permutations on the indices of the odds ratio functions in the compatibility check. Compared with the results of Hobert and Casella (1988) and Wang and Ip (2008), the number of equations to check in our result is much smaller. Furthermore, we propose a simple and natural way to construct the joint density when the conditionally specified densities are compatible. When the conditionally specified densities are incompatible, we propose ways to modify the conditionally specified densities such that the modified conditionally specified densities determine a proper joint density. This will be done through constructing a joint density that is partially compatible with the originally given conditional densities. This means that some but not all of the originally conditionally specified densities may be compatible with the constructed joint density. The odds ratio framework we use can be viewed intuitively in the following way. We first decomposes each of the conditionally specified densities into smaller parts, i.e., the odds ratio function and the conditional density at a fixed condition. Those parts are then reassembled into a functioning machine, the joint density. In the process of assembling, some parts may have to be modified to fit into the functioning machine. If all the parts were well manufactured (compatible), no modification to the parts is needed in the assembly and different ways of assembling yield the same machine. However, if modifications to the parts are necessary in the process of assembling, the functioning machines manufactured may vary depending on the order of the parts assembled and how the parts are modified. One of the advantages of the proposed framework is that the result
can be easily generalized from the conditionally specified densities to the conditionally specified models, *i.e.*, families of conditionally specified densities.

The remainder of the article is organized in the following way. In section 2, we propose to represent a joint density in terms of the odds ratio functions and conditional densities at a fixed condition. We show that the decomposition has the property of variation independence. In section 3, we apply the odds ratio representation of a joint density to study the issue of compatibility in conditionally specified densities and models. A sufficient and necessary condition for the compatibility of the conditionally specified densities is obtained. Results are then applied to the conditionally specified models. Construction of joint models that are partially compatible with the conditionally specified models are demonstrated by a typical example in section 4. The article concludes with a brief discussion on the extension of the proposed approach to the compatibility problem with more complicated ways of model specification.

### 2 Decomposition of a joint density in the odds ratio framework

Let $Y$ be the random vector whose distribution is of interest. Suppose that $Y$ is divided into $t$ groups as $Y_j, j = 1, \cdots, t$, where $Y_j$ has dimension $d_j, j = 1, \cdots, t$. As in Besag (1974), we make the following positivity assumption on the joint distribution of $Y_j, j = 1, \cdots, t$. Namely, if the marginal density of $Y_j, p(y_j) > 0$, for all $j$, then the joint density of $(Y_1, \cdots, Y_t), p(y_1, \cdots, y_t) > 0$.

Consider first the case with $t = 2$. Let $dy_j$ denote the reference measure the density of $Y_j$ was obtained. For a given joint density $p(y_1, y_2)$, as in Chen (2003; 2004; 2007), define the odds ratio function

$$
\eta(y_1, y_1^0; y_2, y_2^0) = \frac{p(y_1, y_2)p(y_1^0, y_2^0)}{p(y_1^0, y_2)p(y_1, y_2^0)},
$$

where $(y_1^0, y_2^0)$ is a point in the sample space. In the following, we suppress $(y_1^0, y_2^0)$ from the odds ratio expression and use $\eta(y_1; y_2)$ to denote the odds ratio function. It is easy to see from the definition that

$$
p(y_1, y_2) = \eta(y_1; y_2)g_2(y_2|y_1)g_1(y_1|y_2) \frac{\int p(y_1, y_2^0)dy_1}{p(y_1^0, y_2)} \frac{\int p(y_1^0, y_2)dy_2}{p(y_1^0, y_2^0)},
$$

where $g_1$ and $g_2$ are respectively conditional densities of $y_1$ given $y_2$, and of $y_2$ given $y_1$. Since
\[ \int \int p(y_1, y_2) dy_1 dy_2 = 1, \text{ it follows that} \]

\[ p(y_1, y_2) = \frac{\eta(y_1; y_2)g_2(y_2|y_1^0)g_1(y_1|y_2^0)}{\int \int \eta(y_1; y_2)g_2(y_2|y_1^0)g_1(y_1|y_2^0) dy_1 dy_2}. \]

Note that the odds ratio function is the same for the two conditional densities, that is,

\[ \eta(y_1; y_2) = \frac{g_1(y_1|y_2)g_2(y_2|y_1^0)}{g_1(y_1^0|y_2)g_2(y_2|y_1)} = \frac{g_2(y_2|y_1)g_2(y_2^0|y_1)}{g_2(y_2^0|y_1)g_2(y_2|y_1^0)}, \]

By similar arguments (Chen, 2003, 2007), we can represent \( g_2(y_2|y_1) \) and \( g_1(y_1|y_2) \) respectively as

\[ g_1(y_1|y_2) = \frac{\eta(y_1; y_2)g_1(y_1|y_2^0)}{\int \eta(y_1; y_2)g_1(y_1|y_2^0) dy_1}, \]

and

\[ g_2(y_2|y_1) = \frac{\eta(y_1; y_2)g_2(y_2|y_1^0)}{\int \eta(y_1; y_2)g_2(y_2|y_1^0) dy_2}. \]

One of the advantages of these representations is that they clarified the common components shared by different densities. Namely, the odds ratio function and two conditional density at fixed conditions. The three components in the expressions uniquely identify the densities and they are variation independent if no restriction is imposed on the joint density. Now, if we are given an odds ratio function and two densities respectively for \( Y_1 \) and \( Y_2 \) that are subject to weak integrable conditions, we can easily construct a joint density using the three components. Before we proceed, we give sufficient conditions for a given function of \( (y_1, y_2) \) to be an odds ratio function, \textit{i.e.}, the ratio of the odds of a density function. Let \( \eta(y_1; y_2) \) be a given function of \( (y_1, y_2) \). If it is an odds ratio function, there should exist \( (y_1^0, y_2^0) \) in the domain of \( \eta \) such that \( \eta(y_1; y_2^0) = 1 = \eta(y_1^0; y_2) \) for all \( y_1 \) and \( y_2 \) such that \( (y_1^0, y_2) \) and \( (y_1, y_2^0) \) are in the domain of \( \eta \). The following lemma gives easily verifiable sufficient conditions for a given function to be an odds ratio function.

\textbf{Lemma 1.} For a given function \( \eta(y_1; y_2) \geq 0 \), suppose that \( \mathcal{Y}_1 = \{y_1 | \eta(y_1; y_2) > 0\} \) is the same for any given \( y_2 \) and there exists \( (y_1^0, y_2^0) \) such that \( \eta(y_1^0; y_2) = 1 = \eta(y_1^0; y_2^0) \) for all \( y_1 \) and \( y_2 \). Denote the common domain by \( \mathcal{Y} \). Assume that there is a density \( f_0(y_1) \) on \( \mathcal{Y} \) with respect to a known measure denoted by \( dy_1 \) (\textit{e.g.}, count or Lebesgue measure) such that

\[ 0 < \int \eta(y_1; y_2)f_0(y_1)dy_1 < +\infty, \]

for all \( y_2 \). Then \( \eta(y_1; y_2) \) is an odds ratio function.
Proof: Define

\[ f(y_1|y_2) = \frac{\eta(y_1; y_2)f_0(y_1)dy_1}{\int \eta(y_1; y_2)f_0(y_1)dy_1} \]

It is easy to check that \( f(y_1|y_2) \) thus defined is a density function on \( Y \) and \( \eta \) is the odds ratio function of this density.

Lemma 2. Given an odds ratio function \( \eta(y_1; y_2) > 0 \) and two density functions \( f_1 \) and \( f_2 \), if \( \int \int \eta(y_1; y_2)f_1(y_1)f_2(y_2)dy_1dy_2 < +\infty \), then

\[ p(y_1, y_2) = \frac{\eta(y_1; y_2)f_1(y_1)f_2(y_1)}{\int \int \eta(y_1; y_2)f_1(y_1)f_2(y_2)dy_1dy_2}, \]

is a joint density function with conditional densities for \( Y_1 \) given \( Y_2 \) and for \( Y_1 \) given \( Y_2 \) being respectively

\[ g_1(y_1|y_2) = \frac{\eta(y_1; y_2)f_1(y_1)}{\int \eta(y_1; y_2)f_1(y_1)dy_1}, \]

and

\[ g_2(y_2|y_1) = \frac{\eta(y_1; y_2)f_2(y_2)}{\int \eta(y_1; y_2)f_2(y_2)dy_2}. \]

Proof: Straightforward.

One consequence of the lemmas is that, given two conditional densities \( g_1(y_1|y_2) \) and \( g_1(y_2|y_1) \), there exists a joint density \( p(y_1, y_2) \) with \( g_1 \) and \( g_2 \) as the conditional densities if and only if the two odds ratio functions computed from the two conditional densities are equal and the integrable condition on the odds ratio function is satisfied. The consistency check in the case of \( t = 2 \) is relatively easy to be performed and various authors had derived somewhat equivalent results. However, note that the generalizable of the approach to carry out compatibility check and to construct a joint density in high dimensional problems is much more important in practice (Besag, 1974, 1996; Besag and Kooperberg, 1995). The odds ratio representation we proposed here is very easy to be extended to deal with high dimensional problems while other approaches either may encounter difficulties in the generalization or can be cumbersome in the generalized form.

For the general case with \( t > 2 \), a joint density on \((y_1, \ldots, y_t)\) can be first represented as

\[
p(y_1, \ldots, y_t) = \frac{\eta_t(y_1; \{y_{t-1}, \ldots, y_1\})g_t(y_1|y_{t-1}^0, \ldots, y_1^0)g_{t-1}(y_{t-1}, \ldots, y_1|y_t^0)}{\int \cdots \int \eta_t(y_1; \{y_{t-1}, \ldots, y_1\})g_t(y_1|y_{t-1}^0, \ldots, y_1^0)g_{t-1}(y_{t-1}, \ldots, y_1|y_t^0)dy_tdy_{t-1} \cdots dy_1}. \tag{1}
\]
We can then apply a similar representation to $g_{t-1}(y_{t-1}, \ldots, y_1 | y_t^0)$ and repeat this step. Eventually, we obtain

$$p(y_1, \ldots, y_t) = \frac{\prod_{j=2}^{t-1} \eta_j(y_j; \{y_{j-1}, \ldots, y_1\} | y_t^0, \ldots, y_j^0) \prod_{j=1}^{t} f_j(y_j)}{\int \cdots \int \prod_{j=2}^{t-1} \eta_j(y_j; \{y_{j-1}, \ldots, y_1\} | y_t^0, \ldots, y_j^0) \prod_{j=1}^{t} f_j(y_j) dy_j} \quad (2)$$

where $f_j(y_j) = g_j(y_j | y_t^0, \ldots, y_{j+1}^0, y_{(j-1)}^0, \ldots, y_1^0)$, for $j = 1, \ldots, t$, and

$$\eta_j(y_j; \{y_{j-1}, \ldots, y_1\} | y_t^0, \ldots, y_j^0) = \frac{g_j(y_j | y_t^0, \ldots, y_{j-1}^0, y_1) g_j(y_t^0 | y_t^0, \ldots, y_{j-1}^0, y_1) g_j(y_{j+1}^0 | y_t^0, \ldots, y_{j-1}^0, y_1) g_j(y_{j+2}^0 | y_t^0, \ldots, y_{j-1}^0, y_1) \cdots g_j(y_1^0 | y_t^0, \ldots, y_{j-1}^0, y_1)}{g_j(y_j^0 | y_t^0, \ldots, y_{j-1}^0, y_1) g_j(y_t^0 | y_t^0, \ldots, y_{j-1}^0, y_1) g_j(y_{j+1}^0 | y_t^0, \ldots, y_{j-1}^0, y_1) g_j(y_{j+2}^0 | y_t^0, \ldots, y_{j-1}^0, y_1) \cdots g_j(y_1^0 | y_t^0, \ldots, y_{j-1}^0, y_1)} = \eta(y_j; \{y_t^0, \ldots, y_{j+1}^0, y_1\}).$$

The joint density can also be equivalently represented as

$$p(y_1, \ldots, y_t) = \frac{\prod_{j=2}^{t-1} \eta_j(y_j; \{y_t^0, \ldots, y_{j+1}^0, y_{j-1}^0, \ldots, y_1^0\}) \prod_{j=1}^{t} f_j(y_j)}{\int \cdots \int \prod_{j=2}^{t-1} \eta_j(y_j; \{y_t^0, \ldots, y_{j+1}^0, y_{j-1}^0, \ldots, y_1^0\}) \prod_{j=1}^{t} f_j(y_j) dy_j} \quad (3)$$

### 3 Compatibility of conditionally specified densities and models

Consider first the compatibility of conditionally specified densities. More specifically, given a set of conditional densities $g_j(y_j | y_{-j})$, $j = 1, \ldots, t$ where $y_{-j} = (y_l, l \neq j)$, we want to know if there exists a joint distribution such that all the given conditional densities are its corresponding conditional densities. Let

$$\eta_j(y_j; y_{-j}) = \frac{g_j(y_j | y_{-j}) g_j(y_t^0 | y_{-j})}{g_j(y_t^0 | y_{-j}) g_j(y_t^0 | y_{-j})}.$$

Then $g_j(y_j | y_{-j})$ is determined by $\eta_j(y_j; y_{-j})$ and $g_j(y_t^0 | y_{-j})$ as

$$g_j(y_j | y_{-j}) = \frac{\eta_j(y_j; y_{-j}) g_j(y_t^0 | y_{-j})}{\int \eta_j(y_j; y_{-j}) g_j(y_t^0 | y_{-j}) dy_j}.$$

Theorem 1. For a given set of conditional densities $g_j(y_j | y_{-j})$, there exists a joint density $p(x_1, \ldots, x_t)$ with conditional densities being $g_j$, $j = 1, \ldots, t$ if and only if

$$\int \cdots \int \prod_{j=2}^{t} \eta_j(y_j; \{y_t^0, \ldots, y_{j+1}^0, y_{j-1}^0, \ldots, y_1^0\}) \prod_{j=1}^{t} \{g_j(y_j | y_{-j})dy_j\} < +\infty, \quad (4)$$

and

$$\prod_{l=2}^{t} \eta_l(y_l; \{y_t^0, \ldots, y_{l+1}^0, y_{l-1}^0, \ldots, y_1^0\}) = \prod_{l=2}^{t} \eta_{\sigma(l)}(y_{\sigma(l)}; \{y_{\sigma(l)}^0, \ldots, y_{\sigma(l)+1}^0, y_{\sigma(l)-1}^0, \ldots, y_{\sigma(1)}^0\}) \quad (5)$$

for those permutation $\sigma_j$ such that $\sigma_j(\{t, \ldots, 1\}) = \{j, t, \ldots, j+1, j-1, \ldots, 1\}$, $j = 1, \ldots, t-1$ (or equivalently for all permutation $\sigma$ on $\{1, \ldots, t\}$).
Proof: If $g_j(y_j|y_{-j})$, $j = 1, \cdots, t$ are compatible, then it follows from the odds ratio representation of a joint density that (4) holds and the joint density can be represented as

$$p_\sigma(y_1, \cdots, y_t) = \frac{\prod_{l=2}^t \eta_{\sigma(l)}(y_{\sigma(1)}, \{y_0^0, y_0^1, \cdots, y_0^{l-1}, y_{\sigma(l-1)}, \cdots, y_{\sigma(1)}\}) \prod_{l=1}^t g_l(y_l|y_{0,l})}{\int \cdots \int \prod_{l=2}^t \eta_{\sigma(l)}(y_{\sigma(1)}, \{y_0^0, y_0^1, \cdots, y_0^{l-1}, y_{\sigma(l-1)}, \cdots, y_{\sigma(1)}\}) \prod_{l=1}^t \{g_l(y_l|y_{0,l})dy_l\}},$$

(6)

for any permutation $\sigma$. That is, $p_\sigma$ for all permutation $\sigma$ are equal. By setting $y_j = y_j^0$ for all $j$, it follows that the denominators are all equal for different $\sigma$. Next, by canceling the denominator and $\prod_{l=1}^t g_l(y_l|y_{0,l})$ from the numerator from the equations, we obtain that (5) holds for all the permutations.

To show the reverse, assume that (5) is true for all $\sigma_j$, $j = 1, \cdots, t - 1$. Then $p_{\sigma_j}$, $j = 1, \cdots, t$ are the same because of (5). Since the conditional density for $Y_{\sigma_j(t)} = Y_j$ given the other variables computed from the joint density $p_{\sigma_j}$ is exactly the same as $g_j(y_j|y_{-j})$. Hence $g_j(y_j|y_{-j})$ corresponds to the conditional densities from a single joint distribution. It then follows that the $\{g_j(y_j|y_{-j}), j = 1, \cdots, t\}$ are compatible.

Compared to the conditions in Theorems 1 and 2 of Hobert and Casella (1998), our Theorem 1 has two improvements. First, $t - 1$ rather than $t! - 1$ equations need to be checked to see if the conditional densities are compatible. The reduction is substantial even for $t = 3$. Second, the conditions in terms of odd ratio functions are simpler because redundancy in terms of the conditional densities at a fixed condition in the conditional density expressions is canceled. Compared with theorem 4 of Wang and Ip (2008), our result is also simpler because we need to check $t - 1$ equations based on odds ratio functions rather than $t(t - 1)/2$ equations based on the conditional densities.

Condition (5) in Theorem 1 is relatively easy to check. For example, when $t = 2$, the condition reduces to $\eta_1(y_1; y_2) = \eta_2(y_2; y_1)$. When $t = 3$, the conditions become

$$\eta_3(y_3; \{y_2, y_1\})\eta_2(y_2; y_1|y_3^0) = \eta_2(y_2; \{y_1, y_3\})\eta_3(y_3; y_1|y_2^0) = \eta_1(y_1; \{y_2, y_3\})\eta_3(y_3; y_2|y_1^0),$$

or equivalently

$$\eta_3(y_3; \{y_2, y_1\})\eta_2(y_2; \{y_3^0, y_1\}) = \eta_2(y_2; \{y_1, y_3\})\eta_3(y_3; \{y_1, y_2^0\}) = \eta_1(y_1; \{y_3, y_2\})\eta_3(y_3; \{y_2, y_1^0\}).$$

Note that, to find incompatibility in the conditional models, we only need to show that one of the equalities does not hold or the integral does not converge.
The joint densities constructed in Theorem 1 is very useful in checking the compatibility of conditionally specified models. The compatibility of conditionally specified models can be defined as that there exists a joint model such that each of the conditionally specified models can be obtained as the corresponding conditional model of the joint model. Given a set of conditionally specified models \( \{g_j(y_j|y_{-j}, \theta_j), \theta_j \in \Theta_j\}, j = 1, \cdots, t \), it is not difficult to construct a joint model \( \{p_\sigma(y_1, \cdots, y_t, \theta), \theta \in \Theta\} \) from Theorem 1, where \( \Theta \) is the range of \( \theta \). We can then recompute the conditional models of the joint model. If all the conditional models based on the joint model are equivalent to the conditionally specified models after possible reparametrization, we can conclude that the conditionally specified models are compatible. Otherwise, they are incompatible. For incompatible conditionally specified models, we can likewise construct different joint models based on the conditionally specified models. Each of the constructed joint models is partially compatible with the conditionally specified models.

4 Modifications to incompatible conditionally specified models based on the joint model construction

When the conditionally specified models are compatible, the constructed joint model is very useful in reparameterizing the conditional models. This is almost always needed in practice. In the case that the conditionally specified models are incompatible, the construction approach may yield several joint models, each of which is partially compatible with the conditionally specified models. Those constructed joint models can also be useful in practice. We use the following example to illustrate the idea. Suppose that \( Y_1, \cdots, Y_t \) are a set of variables modeled by conditional models. For the simplicity of presentation, assume that \( Y_1, \cdots, Y_t \) are all continuous variables. However, similar discussions can be applied to the case with all discrete variables, or a mixture of discrete and continuous variables equally well. Assume the fully conditional models are

\[
Y_j = \beta_{j0} + \sum_{k \neq j} \beta_{jk} Y_k + \epsilon_j,
\]

9
where $\epsilon_j \sim N(0, \sigma_j)$, $j = 1, \cdots, t$. Note that $\epsilon_j$, $j = 1, \cdots, t$ are usually not independent (Besag, 1974). The regression form implies that the conditional densities are

$$f_j(y_j | y_{-j}) = \frac{1}{\sqrt{2\pi\sigma_j}} \exp \left\{ -\frac{1}{2\sigma_j^2} \left( y_j - \beta_{j0} - \sum_{k \neq j} \beta_{jk} y_k \right)^2 \right\}.$$  

The odds ratio function and the conditional density at a fixed condition are respectively

$$\eta_j(y_j; y_{-j}) = \exp \left\{ \sum_{k \neq j} \frac{\beta_{jk}}{\sigma_j^2} (y_j - y_{j0})(y_k - y_{k0}) \right\},$$

and $f_j(y_j | y_{-j0})$. A joint model obtained based on Theorem 1 has density proportional to

$$\exp \left\{ \sum_{j=2}^t \sum_{k=1}^{j-1} \frac{\beta_{jk}}{\sigma_j^2} (y_j - y_{j0})(y_k - y_{k0}) \right\} \prod_{j=1}^t \frac{1}{\sqrt{2\pi\sigma_j}} \exp \left\{ -\frac{1}{2\sigma_j^2} \left( y_j - \beta_{j0} - \sum_{k \neq j} \beta_{jk} y_k - \sum_{k=j+1}^t \beta_{kj} \frac{\sigma_j^2}{\sigma_k^2} (y_k - y_{k0}) \right)^2 \right\}.$$

From this joint density, it can be derived that the conditional density for $Y_j$ given $Y_{-j}$ is

$$f^*_j(y_j | y_{-j}) = \frac{1}{\sqrt{2\pi\sigma_j}} \exp \left\{ -\frac{1}{2\sigma_j^2} \left( y_j - \beta_{j0} - \sum_{k=1}^{j-1} \beta_{jk} y_k - \sum_{k=j+1}^t \beta_{jk} y_{k0} - \sum_{k=j+1}^t \beta_{kj} \frac{\sigma_j^2}{\sigma_k^2} (y_k - y_{k0}) \right)^2 \right\}.$$

Without further restrictions on the parameters, $f^*$ and $f$ define the same model with different parametrizations. This implies that the conditionally specified models are compatible. When the parameters satisfy the constraints,

$$\beta_{kj} / \sigma_k^2 = \beta_{jk} / \sigma_j^2$$

for all $k \neq j$, the equivalent models become identical. In practice, if the data were generated from the joint normal models and no restriction is imposed on the parameters in the conditionally specified models, the consistent estimator of the parameters based on the conditionally specified models automatically satisfies the constraints. However, taking the constraints into consideration in estimation will increase the estimation efficiency.

More sophisticatedly specified conditional models may include either interaction and/or second order terms. This seemingly harmless and natural addition to the conditional models, however, can instantly destroy the compatibility of the conditional models. The question now becomes how to minimally and sensibly amend the conditional models to make them compatible. To illustrate the idea, suppose that we only add one high order term to the conditional model of $Y_1$ given $Y_2, \cdots, Y_t$. That is,

$$Y_1 = \beta_{10} + \sum_{k=2}^t \beta_{1k} Y_k + \alpha_{12} Y_2^2 + \epsilon_1.$$
It is easy to see that the odds ratio function obtained from this model is

$$\log \eta_1 \{ Y_1; (Y_2, \cdots, Y_t) \} = \sum_{k=2}^{t} \frac{\beta_{1k}}{\sigma_1} (y_1 - y_{10})(Y_k - Y_{k0}) + \frac{\alpha_{12}}{\sigma_1} (Y_1 - Y_{10})(Y_2 - Y_{20}),$$

and the odds ratio functions from the other conditional models are

$$\log \eta_j \{ Y_j; (Y_i, i \neq j) \} = \sum_{k \neq j} \frac{\beta_{jk}}{\sigma_j} (y_j - y_{j0})(Y_k - Y_{k0}),$$

for \( j = 2, \cdots, t \). All the conditional densities at a fixed condition that are obtained from the odds ratio decomposition of the conditional densities are normal densities.

The conditionally specified models can be seen as incompatible by checking the conditional odds ratios of \( Y_1 \) and \( Y_2 \) given the rest of variables from \( g_1 \) and \( g_2 \) when \( \alpha_{12} \neq 0 \). It is hard to modify the conditionally specified models directly in an attempt to correct the conflicts in those models. By applying Theorem 1 to this problem, we can construct many joint models which are partially compatible with the conditionally specified models. Note that if the second order terms in the model \( Y_1 \) given \( Y_{-1} \) is very important to keep, we can select \( Y_1 \) before \( Y_2 \) in the order of the odds ratio function combination to keep this high order association term there. To resolve the incompatibility, we can redefine \( \log \eta_2 \) as

$$\log \eta_2^*(Y_2; Y_{-2}) = \sum_{k \neq 2} \frac{\beta_{2k}}{\sigma_2} (y_2 - y_{20})(Y_k - Y_{k0}) + \frac{\alpha_{12}}{\sigma_1} (Y_1 - Y_{10})(Y_2 - Y_{20}),$$

In addition, we need to modify \( g_2(y_2|y_{-20}) \) to make it different from the normal density so that the integral in the denominator of the odds ratio representation of the joint density is finite. This can be done easily by modify \( g_2(y_2|y_{-20}) \) to

$$g^*(y_2|y_{-20}) \propto \exp(-y_2^4/\sigma^2).$$

As a result, \( g(y_2|y_{-2}) \) is modified to

$$g^*(y_2|y_{-2}) = \frac{\eta_2^*(Y_2; Y_{-2})g^*(y_2|y_{-20})}{\int \eta_2^*(Y_2; Y_{-2})g^*(y_2|y_{-20})dy_2}.$$
when $\frac{\beta_{kj}}{\sigma_k^2} = \frac{\beta_{jk}}{\sigma_j^2}$ for all $k \neq j$. Note also that different orders of the odds ratio combination in the foregoing expression do not alter the joint model. Rather they determine different parameterizations of the joint model. The joint model can also be expressed as

$$p^*(x_1, \cdots, x_t) = \frac{\prod_{l=1}^{t-1} \eta_l(y_l, \{y_t, \cdots, y_{(l+1)}, y_{(l-1)}, \cdots, y_{10}\}) \prod_{l \neq 2} g_l(y_l | y_{-l}^0) g_2^*(y_2 | y_{-20})}{\int \cdots \int \prod_{l=1}^{t-1} \eta_l(y_l, \{y_t, \cdots, y_{(l+1)}, y_{(l-1)}, \cdots, y_{10}\}) \prod_{l \neq 2} g_l(y_l | y_{-l}^0) g_2^*(y_2 | y_{-20}) dy_l}.$$ 

However, the order of the odds ratio combination in this expression cannot be changed arbitrarily without changing the joint model.

It can be seen that $p^*$ is a joint model that is compatible with all of the conditionally specified models except $g_2(y_2 | y_{-2})$. Note that we have enforced the change of $g_2(y_2 | y_{-20})$ here. It is also possible to allow observed data to determine what distribution to use. It can also be seen from the above arguments that when the conditionally specified models are close to compatible, the modification of the originally conditionally specified models can be very limited in making the conditional models compatible. As a result, most features of the originally conditionally specified models are retained. On the other hand, if the conditionally specified models are in a high degree of conflicts, modifications to the conditional models are likely imperative to make the result of the statistical analysis interpretable even if some features in the conditionally specified models are eliminated in the modifications. However, Theorem 1 gives us some flexibility in determining which features to keep. Theorem 1 also suggests other modifications. For example, if the order of $Y_1$ and $Y_2$ in the odds ratio combination of the last displayed expression is switched, the constructed joint model would be the joint normal model with the second order term removed. This construction may be unfavorable given the importance of the second order term in the conditional model.

The foregoing example provides us with a general strategy to modify conditionally specified models to achieve compatibility and to obtain the joint models. In summary, when the conditionally specified models are incompatible, we first modify the odds ratio functions of the conditionally specified models. In doing so, we can choose the important features in the models and try to maximally keep them in the joint model. To make the integral in the denominator of the potential joint density finite, we may need to modify the conditional density at a fixed condition also. From a practical point of view, we may modify the conditional densities at a fixed condition to be unspecified. Such a modification can generate many interesting models to fit the data. The interpretation of those models can also be attractive in practice.
5 Discussion

We proposed to study the compatibility issue in conditionally specified models in the framework of the odds ratio representation of a joint density. It allows us to obtain the simplest conditions to check in verifying the compatibility of a given set of conditionally specified models. The framework also suggests ways to construct joint models that are partially or fully compatible with the conditional models. Modified conditionally specified models can be obtained based on the constructed joint models. Note that, in the definition and derivation, each $y_j, j = 1, \ldots, t$, can be a vector. The results obtained in this paper apply to blocks of variables in the conditional specification equally well. We have mainly concentrated on the fully conditionally specified models. We note that the framework proposed in this paper can also be applied to studying the compatibility issue in more complicated model specifications, such as the partially conditional specification and the overlapped block conditional specification. The framework can be applied to variables which are discrete, or continuous, or a mixture of both.


