Journal of the Korean Statistical Society (2000), **29***: 2, pp 137–153*

Jeffrey's Noninformative Prior in Bayesian Conjoint Analysis †

Man-Suk Oh[1] and Yura Kim[2]

Abstract

Conjoint analysis is a widely-used statistical technique for measuring relative importance that individuals place on the product's attributes. Despite its practical importance, the complexity of conjoint model makes it difficult to anlyze. In this paper, we consider a Bayesian approach using Jeffrey's noninformative prior. We derive Jeffrey's prior and give a sufficient condition under which the posterior derived from the Jeffrey's prior is proper.

Key Words : Conjoint model, Proper posterior, Marketing research, Improper prior.

1. Introduction

Over the past three decades, conjoint analysis has evolved as a primary set of techniques employed by both academics and practitioners of marketing research for measuring consumer tradeoffs among multi-attributed products and services. It is a useful statistical technique for measuring the relative importance that individuals place on attributes of a product or service. In a typical conjoint study, various levels of several key attributes are selected, and product or service profiles are constructed using a level of each attribute. A survey is conducted and respondents are asked to rank the product or service profiles according to their preference. The observed rankings are then used to determine the relative impact of each feature on the individual's overall preference.

An objective of conjoint analysis is to assign a weight, called a partworth, to each level of each feature so that the ranking of the profiles based on the summation of the corresponding partworths reproduces the participant's original overall preference ranking.

---

†This work was supported by BK21, Korea.

[1]Statistics Dept., Ewha Womans University, Seoul 120-750, Korea.

[2]IT Service Team/Information Analysis Part, LG Capital, 11F1 Nara Investment Banking Bldg., 1328-3 Seocho-Dong, Seoul 137-070, Korea.

In estimating the partworths, however, traditional estimation methods such as least squares require each subject to respond to more profiles than product attributes, resulting in lengthy questionnaires for complex and multiattributed product or service. Long questionnaires pose both practical and theoretical problems. Response rate tends to decrease with increasing question numbers, and more importantly, academic evidence indicates that long questionnaires may induce response biases.

Thus, it is desirable to develop experimental design and estimation method that estimates the partworths with shorter questionnaires. Lenk et. al. (1996) introduced a hierarchical conjoint model and showed that Bayesian analysis of the model does not require full rank individual-level design matrices and hence one may use shorter questionnaires.

The hierarchical conjoint model given by Lenk et. al. (1996) describes the variation in a subject's responses and the variation in the subject's partworths over the population as follows:

$$y_i = X_i\beta_i + \epsilon_i,\ \ i = 1, ..., n, \tag{1.1}$$

$$\beta_i = \Theta z_i + \delta_i, i = 1, ..., n. \tag{1.2}$$

In (1.1), y_i is a $m_i \times 1$ vector of observations, X_i is a $m_i \times p$ known design matrix, β_i is a $p \times 1$ vector of regression coefficients for the i-th experimental subject, and ϵ_i is a $m_i \times 1$ vector of errors and $\epsilon_i \sim N(0, \sigma_i^2 I_{m_i})$ independently. In (1.2), Θ is $p \times q$ matrix of regression coefficients, z_i is $q \times 1$ vector of known covariates, and δ_i is $p \times 1$ vector of errors and $\delta_i \sim N_p(0, \Lambda)$ independently. The errors $\{\epsilon_i\}$ and $\{\delta_i\}$ are assumed to be mutually independent. A simpler version of the above model can be given by assuming that z_i's are all equal to 1, Θ is the mean vector for the individual-level coefficients β_i, and the individual variance σ_i^2 are all equal (see Yang and Chen, 1995).

In Bayesian approach to analysis of the conjoint model, noninformative priors are often desired especially when there is no prior information on the parameters of the model or when one wants to compare Bayesian inference with classical inference. Among many possible noninformative priors, Jeffrey's prior is widely used since it is invariant under reparameterization. For the conjoint model, however, Jeffrey's prior is not easy to obtain. Moreover, since the Jeffrey's prior is improper and there are often many parameters in the model compared to observations, one needs to check if the posterior obtained from the Jeffrey's prior is proper. Due to the complexity of the model, analytic evaluation of the posterior is very difficult. However, if one finds some conditions on the data structures under which the posterior is proper then these conditions may be used to design experiment, for instance to determine how many subjects should be taken and how many full rank design matrices are required, which guarantees a proper posterior.

In this paper, we derive the Jeffrey's noninformative prior for the conjoint model by using some matrix results. Then, under the equal individual-level response variance condition, we prove that the corresponding posterior is proper when the number of full rank design matrix is greater than or equal to twice the number of regression coefficient parameters plus 1. This would give a guideline about the minimum number of full rank design matrices given the number of attributes of a product or service.

The rest of this paper is organized as follows. In Section 2, some matrix results are presented and the Jeffrey's prior are derived by using the results. In Section 3, we consider some bounds on Jeffrey's prior density and derive a sufficient condition for the posterior density to be proper, under the assumption that all the individual-level response variances are equal.

2. Jeffrey's Noninformative Prior

2.1. Notations and Some Matrix Results

We will use the following notations throughout this paper. A^t, $|A|$, and $tr(A)$ denote the transpose, determinant, and trace of a square matrix A, respectively. Denote $vec(\ )$ to be the matrix operator which arranges the columns of a matrix into one long column, and $vecp(\ )$ to be the matrix operator which arranges the columns of lower left corner of a symmetric matrix into one long column. The Kronecker product of two matrices, A and B, is denoted by $A \otimes B$. $A \geq 0$ means that A is semi-definite positive. G denotes a $(p(p+1)/2) \times p^2$ constant matrix $(\partial vecV)/(\partial vecpV)$, where V is a $p \times p$ symmetric matrix. We use c to denote a constant independent of parameters, while its value varies from place to place.

We will use the following matrix results in this paper frequently. Results 2.1 and 2.2 can be found in Magnus and Neudecker (1988). Result 2.4 follows from Result 2.1. Results 2.5 and 2.6 are given by Wiens (1985). Yang and Chen (1995) also summarized these results.

Result 2.1 For any $p \times p$ matrices A, B, C, and D,
$(vec(D^t))^t(C^t \otimes A)(vec(B)) = tr(ABCD)$.

Result 2.2 $vec(ABC) = (C^t \otimes A)vec(B)$.

Result 2.3 Suppose $\begin{pmatrix} A & B^t \\ B & C \end{pmatrix} \geq 0$ and $A > 0$, then $\begin{pmatrix} A & B^t \\ B & C \end{pmatrix} \leq |A| \cdot |C|$.

Result 2.4 If for $p\times p$ matrices A and B, $0 \leq A \leq B$, then $A\otimes A \leq A\otimes B \leq B\otimes B$. If $p \times p$ matrices $A_i \geq 0$, $i = 1, ..., n$ then

$$\sum_{i=1}^{n} A_i \otimes A_i \leq \left(\sum_{i=1}^{n} A_i\right) \otimes \left(\sum_{i=1}^{n} A_i\right).$$

Result 2.5 If A is a $p \times p$ matrix, then $|G(A \otimes A)G^t| = |GG^t|\,|A|^p + 1$.

Result 2.6 For a $p \times p$ symmetric matrix V, $vec(V) = G^t vecp(V)$.

2.2. Jeffrey's Noninformative Prior

In the hierarchical conjoint model, β_i's are random coefficients connecting (1.1) and (1.2), and Θ, Λ, $\{\sigma_i^2\}$ are population parameters. Therefore, the likelihood function depends only on $(\Theta, \Lambda, \{\sigma_i^2\})$. From

$$\Theta z_i = \begin{pmatrix} \theta_{11} & \cdots & \theta_{1q} \\ \vdots & \vdots & \vdots \\ \theta_{p1} & \cdots & \theta_{pq} \end{pmatrix} \begin{pmatrix} z_{i1} \\ \vdots \\ z_{iq} \end{pmatrix} = (z_i^t \otimes I)vec(\Theta) = w_i \boldsymbol{\theta},$$

where $w_i = z_i^t \otimes I_p$ and $\boldsymbol{\theta} = vec(\Theta)$, we can write the equation (1.2) as $\beta_i = w_i\boldsymbol{\theta} + \delta_i$, for $i = 1, 2, \cdots, n$. We will use both Θz_i and $w_i\boldsymbol{\theta}$ interchangeably depending on convenience.

Since $y_i = X_i\beta_i + \epsilon_i$, $\beta_i \sim MVN(\Theta z_i, \Lambda)$, $\epsilon_i \sim N(0, \sigma_i^2 I_{m_i})$, and β_i and ϵ_i are independent,

$$y_i \sim MVN(X_i\Theta z_i, X_i\Lambda X_i^t + \sigma_i^2 I_{m_i}). \tag{2.1}$$

Thus, the likelihood function is given as

$$\begin{aligned} L(\Theta, \Lambda, \{\sigma_i^2\}|data) &= \prod_{i=1}^{n} \left|X_i\Lambda X_i^t + \sigma_i^2 I_{m_i}\right|^{-1/2} \\ &\times \; exp[-\frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\Theta z_i)^t(X_i\Lambda X_i^t + \sigma_i^2 I_{m_i})^{-1}(y_i - X_i\Theta z_i)] \end{aligned} \tag{2.2}$$

$$= \prod_{i=1}^{n} \left|X_i \Lambda X_i^t + \sigma_i^2 I\right|^{-1/2}$$
$$\times \quad exp[-\frac{1}{2}\sum_{i=1}^{n}(y_i - X_i w_i \boldsymbol{\theta})^t (X_i \Lambda X_i^t + \sigma_i^2 I)^{-}1(y_i - X_i w_i \boldsymbol{\theta})], \tag{2.3}$$

where *data* are $\{(y_i, X_i),\ i = 1, ..., n\}$.

The following result was given by Tracy and Jinadasa (1988).

Lemma 2.1 *If a* $p \times 1$ *random vector* W *follows a* $MVN(\mu, V)$, *then the Fisher information matrix* $I(\mu, V)$ *for* μ *and* V *is given by*

$$I(\mu, V) = \begin{pmatrix} V^{-1} & 0 \\ 0 & \frac{1}{2}G(V^{-}1 \otimes V^{-}1)G^t \end{pmatrix}.$$

Using Lemma 2.1, the Fisher information matrix for Θ, Λ, and $\{\sigma_i^2\}$ in the hierarchical conjoint model can be obtained as in the following theorem.

Theorem 2.1 *For the hierarchical conjoint model with the likelihood function given in (2.2), the Fisher information matrix for* $(\Theta, \Lambda, \{\sigma_i^2\})$ *is*

$$I(\Theta, \Lambda, \{\sigma_i^2\})$$
$$= \begin{pmatrix} \sum_{i=1}^{n} I_i(\Theta) & 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{2}\sum_{i=1}^{n} I_i(\Lambda) & \frac{1}{2}I_1(\Lambda, \sigma_1^2) & \frac{1}{2}I_2(\Lambda, \sigma_2^2) & \cdots & \frac{1}{2}I_n(\Lambda, \sigma_n^2) \\ 0 & \frac{1}{2}I_1^t(\Lambda, \sigma_1^2) & \frac{1}{2}I_1(\sigma_1^2) & 0 & \cdots & 0 \\ 0 & \frac{1}{2}I_2^t(\Lambda, \sigma_2^2) & 0 & \frac{1}{2}I_2(\sigma_2^2) & & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \frac{1}{2}I_n^t(\Lambda, \sigma_n^2) & 0 & 0 & \cdots & \frac{1}{2}I_n(\sigma_n^2) \end{pmatrix},$$

where

$$I_i(\Theta) = A_i, \quad I_i(\Lambda) = G(B_i \otimes B_i)G^t, \quad I_i(\Lambda, \{\sigma_i^2\}) = G\ vec(C_i),$$
$$I_i(\{\sigma_i^2\}) = tr((X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-2}), \quad A_i \equiv w_i^t X_i^t (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-1} X_i w_i,$$
$$B_i \equiv X_i^t (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-1} X_i, \quad C_i \equiv X_i^t (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-2} X_i.$$

Proof : Lemma 2.1 yields

$$I_i(\Theta) = \left[\frac{\partial (X_i w_i \boldsymbol{\theta})}{\partial\, \boldsymbol{\theta}}\right] (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-}1 \left[\frac{\partial (X_i w_i \boldsymbol{\theta})}{\partial\, \boldsymbol{\theta}}\right]^t$$
$$= w_i^t X_i^t (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-1} X_i w_i \tag{2.4}$$

and

$$I_i(\Lambda,\{\sigma_i^2\}) = \left[\frac{\partial\ vec(X_i\Lambda X_i^t+\sigma_i^2 I)}{\partial\ vec(\Lambda,\{\sigma_i^2\})}\right]\cdot(\Lambda_i^{*-1}\otimes\Lambda_i^{*-1})^{-1}\cdot\left[\frac{\partial\ vec(X_i\Lambda X_i^t+\sigma_i^2 I)}{\partial\ vec(\Lambda,\{\sigma_i^2\})}\right]^t,$$

where $\Lambda_i^* = X_i\Lambda X_i^t + \sigma_i^2 I_{m_i}$. If we let λ_{lm} be the (l,m)-th element of Λ, then

$$\frac{\partial\ vec(X_i\Lambda X_i^t+\sigma_i^2 I_{m_i})}{\partial\ \lambda_{lm}} = vec\left(X_i\frac{\partial\Lambda}{\partial\lambda_{lm}}X_i^t\right) = vec\left(X_i\Delta_{lm}X_i^t\right),$$

where $\Delta_{lm} = e_le_m^t + e_me_l^t$ if $l<m$ and $\Delta_{mm} = e_me_m^t$. Here e_l denotes the column vector with 1 in the l-th row and 0 elsewhere. Then (2.4) and Result 2.1 imply

$$\begin{aligned} I_i(\lambda_{lm},\lambda_{hk}) &= \left[vec\left(X_i\Delta_{lm}X_i^t\right)\right]^t(\Lambda_i^*-1\otimes\Lambda_i^*-1)\left[vec\left(X_i\Delta_{hk}X_i^t\right)\right] \\ &= tr\left[\Lambda_i^{*-1}\cdot X_i\Delta_{hk}X_i^t\cdot\Lambda_i^{*-1}\cdot X_i\Delta_{lm}X_i^t\right] \\ &= tr\left[B_i\cdot\Delta_{hk}\cdot B_i\cdot\Delta_{lm}\right] \\ &= \left[vec\left(\Delta_{lm}\right)\right]^t\cdot(B_i\otimes B_i)\cdot\left[vec\left(\Delta_{lm}\right)\right], \end{aligned}$$

and hence

$$I_i(\Lambda) = G(B_i\otimes B_i)G^t. \tag{2.5}$$

Now, from $\partial/\partial\sigma_i^2\{vec(X_i\Lambda X_i^t+\sigma_i^2 I_{m_i})\} = vec(I_{m_i})$ and Result 2.1,

$$\begin{aligned} I_i(\lambda_{lm},\{\sigma_i^2\}) &= \left[vec\left(X_i\Delta_{lm}X_i^t\right)\right]^t\cdot\left(\Lambda_i^{*-1}\otimes\Lambda_i^{*-1}\right)\cdot\left[vec\left(I_{m_i}\right)\right] \\ &= tr\left[\Lambda_i^{*-1}\cdot I_{m_i}\cdot\Lambda_i^{*-1}\cdot X_i\Delta_{lm}X_i^t\right] \\ &= \left[vec\left(\Delta_{lm}\right)\right]^t\cdot\left(D_i^t\otimes D_i^t\right)\cdot\left[vec\left(I_{m_i}\right)\right], \end{aligned} \tag{2.6}$$

where $D_i = (X_i\Lambda X_i^t+\sigma_i^2 I_{m_i})^{-1}X_i$. Thus, Result 2.2 and (2.6) gives

$$I_i(\Lambda,\{\sigma_i^2\}) = \left[\frac{\partial vec(\Lambda)}{\partial vecp(\Lambda)}\right]\cdot(D_i^t\otimes D_i^t)\cdot vec(I_{m_i}) = G\cdot vec(C_i).$$

Using Result 2.1,

$$I_i(\{\sigma_i^2\}) = \left[vec\left(I_{m_i}\right)\right]^t\cdot\left(\Lambda_i^{*-1}\otimes\Lambda_i^{*-1}\right)\cdot\left[vec\left(I_{m_i}\right)\right] = tr((X_i\Lambda X_i^t+\sigma_i^2 I_i)^{-2}). \tag{2.7}$$

The result follows from (2.4)-(2.7). □

We now can obtain the Jeffrey's prior for the hierarchical conjoint model.

Theorem 2.2 *The Jeffrey's prior of the hierarchical conjoint model is given as*

$$\pi(\Theta, \Lambda, \{\sigma_i^2\}) = \left|\sum_{i=1}^{n} I_i(\Theta)\right|^{1/2} \cdot \left|\begin{array}{ccccc} \sum_{i=1}^{n} I_i(\Lambda) & I_1(\Lambda, \sigma_1^2) & I_2(\Lambda, \sigma_2^2) & \cdots & I_n(\Lambda, \sigma_n^2) \\ I_1^t(\Lambda, \sigma_1^2) & I_1(\sigma_1^2) & 0 & \cdots & 0 \\ I_2^t(\Lambda, \sigma_2^2) & 0 & I_2(\sigma_2^2) & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_n^t(\Lambda, \sigma_n^2) & 0 & 0 & \cdots & I_n(\sigma_n^2) \end{array}\right|^{1/2}$$

Proof : The proof is straightforward from the Fisher information matrix given in Theorem 2.1 and hence is omitted.

3. A Sufficient Condition for the Posterior to be Proper

Throughout this section, we will denote the largest eigenvalue of Λ by λ_{max} and the smallest value of $\{\sigma_i^2\}$ by σ_*^2.

3.1. Bounds on the Jeffrey's Prior

The results in this section will be used to find a sufficient condition under which the posterior density is proper when we use the Jeffrey's prior.

Proposition 3.1 *Let* $A_i = w_i^t X_i^t (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-1} X_i w_i$ *and* $B_i = X_i^t (X_i \Lambda X_i^t + \sigma_i^2 I_{m_i})^{-1} X_i$ *then*

$$\left|\sum_{i=1}^{n} A_i\right| \leq \frac{c}{\sigma_*^{2p}}, \quad (3.1)$$

$$\left|\sum_{i=1}^{n} B_i\right| \leq \frac{c}{\sigma_*^{2p}}, \quad (3.2)$$

$$\left|\sum_{i=1}^{n} A_i\right| \leq \frac{c}{\lambda_{max}\sigma_*^{2(p-1)}}, \quad (3.3)$$

$$\left|\sum_{i=1}^{n} B_i\right| \leq \frac{c}{\lambda_{max}\sigma_*^{2(p-1)}}. \quad (3.4)$$

Proof : Inequality (3.1) and (3.2) are trivial. For (3.3), since $\Lambda \geq 0$, Λ can be decomposed as

$$\Lambda = Q \begin{pmatrix} \lambda_{max} & 0 \\ 0 & \Lambda^* \end{pmatrix} Q^t,$$

where Λ^* is a $(p-1) \times (p-1)$ diagonal matrix of eigenvalues other than λ_{max} and Q is a $p \times p$ orthonormal matrix.

Denoting $X_iQ = (h_1^{(i)}, \cdots, h_p^{(i)})$, where $h_j^{(i)}$ is a $m_i \times 1$ vector, it follows that

$$\begin{aligned} X_i \Lambda X_i^t + \sigma_i^2 I_{m_i} &\geq (h_1^{(i)}, \cdots, h_p^{(i)}) Diag(\lambda_{max}, 0, .., 0)(h_1^{(i)}, ..., h_p^{(i)})^t + \sigma_i^2 I_{m_i} \\ &= \lambda_{max} h_1^{(i)} h_1^{(i)t} + \sigma_i^2 I_{m_i}. \end{aligned}$$

Therefore,

$$\begin{aligned} & Q^t X_i^t (X_i \Lambda' X_i^t + \sigma_i^2 I_{m_i})^{-1} X_i Q \\ \leq \; & (h_1^{(i)}, ..., h_p^{(i)})^t \cdot \frac{1}{\sigma_i^2} \left[\frac{I_{m_i} - h_1^{(i)} h_1^{(i)t}}{h_1^{(i)t} h_1^{(i)} + \sigma_i^2 / \lambda_{max}} \right] (h_1^{(i)}, \cdots, h_p^{(i)}). \end{aligned} \tag{3.5}$$

Denote the right hand side of (3.5) as $\begin{pmatrix} h_{11}^{(i)} & * \\ * & H_{p-1,p-1}^{(i)} \end{pmatrix}$. Then

$$h_{11}^{(i)} = h_1^{(i)t} \cdot \frac{1}{\sigma_i^2} \left[I_{m_i} - \frac{h_1^{(i)} h_1^{(i)t}}{h_1^{(i)t} h_1^{(i)} + \sigma_i^2 / \lambda_{max}} \right] \cdot h_1^{(i)} = \frac{1}{\lambda_{max}} \cdot \frac{h_1^{(i)t} h_1^{(i)}}{h_1^{(i)t} h_1^{(i)} + \sigma_i^2 / \lambda_{max}} \leq \frac{1}{\lambda_{max}},$$

$$H_{p-1,p-1}^{(i)} \leq (h_2^{(i)}, ..., h_p^{(i)})^t \cdot \frac{I_{m_i}}{\sigma_i^2} \cdot (h_2^{(i)}, \cdots, h_p^{(i)}).$$

From these and Result 2.3,

$$\begin{aligned} \left| \sum_{i=1}^n A_i \right| &\leq \left| \sum_{i=1}^n h_{11}^{(i)} \right| \cdot \left| \sum_{i=1}^n H_{p-1,p-1}^{(i)} \right| \leq \frac{n}{\lambda_{max}} \left| \sum_{i=1}^n (h_2^{(i)}, ..., h_p^{(i)})^t \cdot \frac{I_{m_i}}{\sigma_i^2} \cdot (h_2^{(i)}, \cdots, h_p^{(i)}) \right| \\ &\leq \frac{n}{\lambda_{max} \sigma_*^{2(p-1)}} \cdot \left| \sum_{i=1}^n (h_2^{(i)}, ..., h_p^{(i)})^t (h_2^{(i)}, \cdots, h_p^{(i)}) \right| \\ &\leq \frac{c}{\lambda_{max} \sigma_*^{2(p-1)}}. \end{aligned} \tag{3.6}$$

Note that the last step of (3.6) follows from the fact that the elements of $X_iQ = (h_1^{(i)}, \cdots, h_p^{(i)})$ are uniformly bounded. The proof for (3.4) is similar to that of (3.3).

Proposition 3.2 *An upper bound on the Jeffrey's prior density is given as*

$$\pi(\Theta, \Lambda, \{\sigma_i^2\}) \leq \frac{c}{\lambda_{max}^{p/2} \, \sigma_*^{p^2+p+2n}}. \tag{3.7}$$

Proof : We first give

$$\pi(\Theta, \Lambda, \{\sigma_i^2\}) \leq \left| \sum_{i=1}^n I_i(\Theta) \right|^{1/2} \cdot \left| \sum_{i=1}^n I_i(\Lambda) \right|^{1/2} \cdot \prod_{i=1}^n |I_i(\sigma_i^2)|^{1/2} \tag{3.8}$$

which is obtained from Theorem 2.2 and Results 2.3. Now we consider upper bounds on each term of the right-hand side of (3.8).

From (2.4) and Proposition 3.1,

$$\left|\sum_{i=1}^{n} I_i(\Theta)\right|^{1/2} = \left|\sum_{i=1}^{n} A_i\right|^{1/2} \leq \frac{c}{\sigma_*^p}. \tag{3.9}$$

From (2.5) and Results 2.4 and 2.5 ,

$$\begin{aligned}\left|\sum_{i=1}^{n} I_i(\Lambda)\right|^{1/2} &= \left|G\left[\sum_{i=1}^{n}(B_i \otimes B_i)\right]G^t\right|^{1/2} \leq \left|G\left[\left(\sum_{i=1}^{n} B_i \otimes \sum_{i=1}^{n} B_i\right)\right]G^t\right|^{1/2} \\ &= \left|GG^t\right|^{1/2}\left|\sum_{i=1}^{n} B_i\right|^{(p+1)/2} = \left|GG^t\right|^{1/2}|\sum_{i=1}^{n} B_i|^{1/2}|\sum_{i=1}^{n} B_i|^{p/2}.\end{aligned} \tag{3.10}$$

Applying the bounds (3.2) and (3.4) yields

$$|\sum_{i=1}^{n} I_i(\Lambda)|^{1/2} \leq c \cdot \left(\frac{1}{\sigma_*^p}\right)\left(\frac{1}{\lambda_{max}\sigma_*^{2(p-1)}}\right)^{p/2}. \tag{3.11}$$

And from (2.7) and (3.3),

$$\begin{aligned}\left|I_i(\sigma_i^2)\right|^{1/2} &= \left|tr((X_i\Lambda X_i^t + \sigma_i^2 I_{m_i})^{-2})\right|^{1/2} = \left[tr\{\frac{1}{\sigma_i^4}(X_i\frac{\Lambda}{\sigma_i^2}X_i^t + I_{m_i})^{-2}\}\right]^{1/2} \\ &\leq \left[tr\{\frac{1}{\sigma_*^4}(X_i\frac{\Lambda}{\sigma_i^2}X_i^t + I_{m_i})^{-2}\}\right]^{1/2} \leq \frac{c}{\sigma_*^2}.\end{aligned} \tag{3.12}$$

The inequality (3.7) follows from these bounds.

3.2. A Sufficient Condition for the Posterior to be Proper

In this section, we assume common variance $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$. By using the upper bound on Jeffrey's prior given in Proposition 3.2, we will derive a sufficient condition for the posterior to be proper when the Jeffrey's prior is used.

3.2.1. The Posterior Density Function

Denote $\pi(\Theta, \Lambda, \sigma^2)$ to be the Jeffrey's prior. From derivations similar to those given in Section 2, we obtain

$$\pi(\Theta, \Lambda, \sigma^2) \propto \left|\sum_{i=1}^{n} I_i(\Theta)\right|^{1/2} \cdot \begin{pmatrix} \sum_{i=1}^{n} I_i(\Lambda) & I_1(\Lambda, \sigma^2) \\ I_1^t(\Lambda, \sigma^2) & I_1(\sigma^2) \end{pmatrix}^{1/2}.$$

An upper bound on $\pi(\Theta,\Lambda,\sigma^2)$ can be obtained by substituting σ_*^2 by σ^2 in (3.7) of Proposition 3.2, resulting in

$$\pi(\Theta,\Lambda,\sigma^2) \leq \frac{c}{\lambda_{max}^{p/2}\ \sigma^{p^2+p+2n}}. \tag{3.13}$$

Also, we obtain the likelihood

$$\begin{aligned} L(\Theta,\Lambda,\sigma^2|data) &= \prod_{i=1}^{n} \left|X_i\Lambda X_i^t + \sigma^2 I_{m_i}\right|^{-\frac{1}{2}} \\ &\quad \times exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\Theta z_i)^t(X_i\Lambda X_i^t + \sigma^2 I_{m_i})^{-1}(y_i - X_i\Theta z_i)\}. \end{aligned}$$

Let $\pi(\Theta,\Lambda,\sigma^2|data)$ be the joint posterior density of $\Theta,\Lambda,$ and σ^2, derived from the Jeffrey's prior $\pi(\Theta,\Lambda,\sigma^2)$. Then

$$\begin{aligned} \pi(\Theta,\Lambda,\sigma^2|data) &\propto \pi(\Theta,\Lambda,\sigma^2)\cdot\prod_{i=1}^{n} \left|X_i\Lambda X_i^t + \sigma^2 I_i\right|^{-\frac{1}{2}} \\ &\times \ exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - X_i\Theta z_i)^t(X_i\Lambda X_i^t + \sigma^2 I_{m_i})^{-1}(y_i - X_i\Theta z_i)\}. \end{aligned}$$

To show that the posterior is proper, one needs to show that

$$\int \pi(\Theta,\Lambda,\sigma^2|data)d\Theta d\Lambda d\sigma^2 < \infty.$$

Due to the complexity of the posterior density, however, analytic integration of the posterior is almost impossible and hence it is not easy to see whether the joint posterior density is proper or not. Especially when we deal with unbalanced observations, it becomes extremely hard. Thus, it seems necessary to pose some conditions on the data structures in order to obtain a proper joint posterior density. For example, we need to know how many subjects should be taken and how many full rank design matrices X_i are needed to obtain a proper posterior.

3.2.2. Auxiliary Variables

The complexity of the Jeffrey's prior can be eased by using an upper bound

$$\pi(\Theta,\Lambda,\sigma^2) \leq \frac{c}{|\Lambda|^{1/2}\ \sigma^{p^2+p+2n}},$$

which is obtained from (3.13) and the fact that $\lambda_{max}^{p/2} \geq |\Lambda|^{1/2}$.

However, the likelihood is still highly complicated especially in terms of Λ and σ_i^2 . To handle this, we use $\beta_1, \beta_2, ..., \beta_n$, given in the hierarchical conjoint model, as auxiliary variables and consider the likelihood of $(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2)$.

Note that $\pi(\Theta, \Lambda, \sigma^2|data) = \int \pi(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data) d\beta_1, .., \beta_n$ and

$$\pi(\Theta, \beta_1, ..., \beta_n, \Lambda, \sigma^2|data) \le \frac{1}{\sigma^{p^2+p+2n+\sum_{i=1}^n m_i}} |\Lambda|^{-(n+1)/2}$$
$$\times exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - X_i\beta_i)^t (y_i - X_i\beta_i) - \frac{1}{2} \sum_{i=1}^n (\beta_i - \Theta z_i)^t \Lambda^{-1} (\beta_i - \Theta z_i\}. \tag{3.14}$$

If we denote the right hand side of (3.14) as $\pi^*(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$ then it is obvious that if $\pi^*(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$ is proper then $\pi(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$ is also proper. Even though $\pi^*(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$ has more variables, it is possible to integrate out Λ and σ^2. Thus, from now on we will focus on $\pi^*(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$ instead of $\pi(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$.

Integrating out Λ and σ^2 from $\pi^*(\Theta, \beta_1, .., \beta_n, \Lambda, \sigma^2|data)$ gives

$$\pi^*(\Theta, \beta_1, ..., \beta_n|data)$$
$$\propto \int \frac{1}{\sigma^{p^2+p+2n+\sum_{i=1}^n m_i}} \cdot exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - X_i\beta_i)^t (y_i - X_i\beta_i)\} d\sigma^2$$
$$\times \int |\Lambda|^{-(n+1)/2} \cdot exp\{-\frac{1}{2} \sum_{i=1}^n (\beta_i - \Theta z_i)^t \Lambda^{-1} (\beta_i - \Theta z_i)\} d\Lambda$$
$$\propto \left| \sum_{i=1}^n (\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t \right|^{-\frac{n-p}{2}} \{\sum_{i=1}^n (y_i - X_i\beta_i)^t (y_i - X_i\beta_i)\}^{-\frac{p^2+p+2n-2+\sum_{i=1}^n m_i}{2}}. \tag{3.15}$$

Note that in the above derivation, we need $n \ge 2p$ to integrate out Λ.

In the next section, we will prove that the right hand side of (3.15) is proper if the number of full-rank X_i 's is greater than $2p + 1$.

3.2.3. A Sufficient Condition for the Posterior to be Proper

Suppose that there are n_2 of X_i 's that are of rank less than p and these X_i 's are the last n_2 of X_i 's, i. e.,

$$r(X_1) = r(X_2) = \cdots = r(X_{n_1}) = p,$$

$$r(X_{n_1+1}) < p, \cdots, r(X_n) < p,$$

where $n_1 = n - n_2$ and $r(X_i)$ is the rank of X_i . Then the following proposition will give an upper bound of $\pi^*(\Theta, \beta_1, ..., \beta_n|data)$ given in (3.15). Here, we assume $n_2 \geq 1$; otherwise we do not need the following proposition.

Proposition 3.3 *If $n_1 \geq 2p$, then*

$$\pi^*(\Theta, \beta_1, ..., \beta_{n_1}|data) \leq c\left|\sum_{i=1}^{n_1}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t\right|^{-(n_1-p)/2} \times\{\sum_{i=1}^{n_1}(y_i - X_i\beta_i)^t(y_i - X_i\beta_i)\}^{-(p^2+p+2n-2+\sum_{i=1}^{n} m_i)/2}. \quad (3.16)$$

Proof : From (3.15),

$$\pi^*(\Theta, \beta_1, ..., \beta_{n_1}|data) \leq c\left|\sum_{i=1}^{n}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t\right|^{-(n-p)/2} \times\{\sum_{i=1}^{n_1}(y_i - X_i\beta_i)^t(y_i - X_i\beta_i)\}^{-(p^2+p+2n-2+\sum_{i=1}^{n} m_i)/2}.$$

Now, we first integrate out β_n from the right hand side of the above inequality. By writing $A = \sum_{i=1}^{n-1}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t$ and $\beta_n^* = A^{-1/2}(\beta_n - \Theta z_n)$, we have

$$\int\left|\sum_{i=1}^{n}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t\right|^{-(n-p)/2} d\beta_n$$
$$= |A|^{-(n-p)/2}\int\left|A^{-1/2}(\beta_n - \Theta z_n)(\beta_n - \Theta z_n)^t A^{-1/2} + I\right|^{-(n-p)/2} d\beta_n$$
$$= |A|^{-(n-p-1)/2}\int\left|\beta_n^*\beta_n^{*t} + I\right|^{-(n-p)/2} d\beta_n^* = |A|^{-(n-p-1)/2}\int(1 + \beta_n^{*t}\beta_n^*)^{-(n-p)/2}d\beta_n^*$$
$$\propto \left|\sum_{i=1}^{n-1}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t\right|^{-(n-p-1)/2}.$$

Repeating the above procedure to integrate out $\beta_{n_1+1}, ..., \beta_{n-1}$ gives (3.16). Note that since $\int(1 + \beta^* t\beta^*)^{-(n-k)/2}d\beta^* < \infty$ only if $n - k > p$, we need the condition $n_1 \geq 2p$.

Now, we integrate out Θ from (3.16).

Proposition 3.4 *If $n_1 \geq 2p + 1$, then*

$$\pi^*(\beta_1, ..., \beta_{n_1}|data) \leq c\left|\sum_{i=1}^{n_1}\beta_i\beta_i^t - \hat{\Theta}(\sum_{i=1}^{n_1} z_i z_i^t)\hat{\Theta}^t\right|^{-(n_1-p-1)/2}$$

$$\times\{\sum_{i=1}^{n_1}(y_i - X_i\beta_i)^t(y_i - X_i\beta_i)\}^{-(p^2+p+2n-2+\sum_{i=1}^n m_i)/2}, \tag{3.17}$$

where $\hat{\Theta} = (\sum_{i=1}^{n_1}\beta_i z_i^t)(\sum_{i=1}^{n_1} z_i z_i^t)^{-1}$.

Proof :

$$\begin{aligned}\sum_{i=1}^{n_1}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t &= \Theta(\sum_{i=1}^{n_1} z_i z_i^t)\Theta^t - 2(\sum_{i=1}^{n_1}\beta_i z_i^t)\Theta^t + \sum_{i=1}^{n_1}\beta_i\beta_i^t \\ &= \left[\Theta - (\sum_{i=1}^{n_1}\beta_i z_i^t)(\sum_{i=1}^{n_1} z_i z_i^t)^{-1}\right]\left(\sum_{i=1}^{n_1} z_i z_i^t\right) \\ &\quad \times \left[\Theta - (\sum_{i=1}^{n_1}\beta_i z_i^t)(\sum_{i=1}^{n_1} z_i z_i^t)^{-1}\right]^t \\ &\quad + \sum_{i=1}^{n_1}\beta_i\beta_i^t - (\sum_{i=1}^{n_1}\beta_i z_i^t)(\sum_{i=1}^{n_1} z_i z_i^t)^{-1}(\sum_{i=1}^{n_1}\beta_i z_i^t)^t.\end{aligned}$$

By writing $\hat{\Theta} = (\sum_{i=1}^{n_1}\beta_i z_i^t)(\sum_{i=1}^{n_1} z_i z_i^t)^{-1}$, $A = \sum_{i=1}^{n_1}\beta_i\beta_i^t - \hat{\Theta}(\sum_{i=1}^{n_1} z_i z_i^t)\hat{\Theta}^t$, $\Theta^* = A^{-1/2}(\Theta - \hat{\Theta})(\sum_{i=1}^{n_1} z_i z_i^t)^{1/2}$, we have

$$\begin{aligned}&\int \left|\sum_{i=1}^{n_1}(\beta_i - \Theta z_i)(\beta_i - \Theta z_i)^t\right|^{-(n_1-p)/2} d\Theta \\ &= \int \left|(\Theta - \hat{\Theta})(\sum_{i=1}^{n_1} z_i z_i^t)(\Theta - \hat{\Theta})^t + A\right|^{-(n_1-p)/2} d\Theta \\ &= |A|^{-(n_1-p)/2}\int \left|A^{-1/2}(\Theta - \hat{\Theta})(\sum_{i=1}^{n_1} z_i z_i^t)(\Theta - \hat{\Theta})^t A^{-1/2} + I\right|^{-(n_1-p)/2} d\Theta \\ &= |A|^{-(n_1-p-1)/2}\int \left|\Theta^*\Theta^{*t} + I\right|^{-(n_1-p)/2} d\Theta^* \\ &\propto |A|^{-(n_1-p-1)/2} = \left|\sum_{i=1}^{n_1}\beta_i\beta_i^t - \hat{\Theta}(\sum_{i=1}^{n_1} z_i z_i^t)\hat{\Theta}^t\right|^{-(n_1-p-1)/2} .\square\end{aligned}$$

Now we need to integrate out $\beta_1, ..., \beta_{n_1}$. Under the condition that there exists at least one $i \leq n_1$ such that

$$y_i \neq X_i(X_i^t X_i)^{-1}X_i^t y_i, \tag{3.18}$$

it can be proven that there exist positive constants δ^* and δ such that

$$\sum_{i=1}^{n_1}(y_i - X_i\beta_i)^t(y_i - X_i\beta_i) \geq \delta^* + \delta\sum_{i=1}^{n_1}\beta_i^t\beta_i.$$

From this inequality and Proposition 3.4, it follows that

$$\begin{aligned}\pi^*(\beta_1,...,\beta_{n_1}|data) &\le c\left|\sum_{i=1}^{n_1}\beta_i\beta_i^t - \hat\Theta(\sum_{i=1}^{n_1} z_iz_i^t)\hat\Theta^t\right|^{-(n_1-p-1)/2} \\ &\times\left[\delta^* + \delta\sum_{i=1}^{n_1}\beta_i^t\beta_i\right]^{-(p^2+p+2n-2+\sum_{i=1}^n m_i)/2}. \quad (3.19)\end{aligned}$$

Note that

$$\begin{aligned}\sum_{i=1}^{n_1}\beta_i\beta_i^t - \hat\Theta(\sum_{i=1}^{n_1} z_iz_i^t)\hat\Theta^t &= \sum_{i=1}^{n_1}\beta_i\beta_i^t - (\sum_{i=1}^{n_1}\beta_iz_i^t)(\sum_{i=1}^{n_1}z_iz_i^t)^{-1}(\sum_{i=1}^{n_1}\beta_iz_i^t)^t \\ &= (\beta_1,...,\beta_{n_1})\left[I_{n_1} - z(z^tz)^{-1}z^t\right](\beta_1,...,\beta_{n_1})^t \\ &= (\beta_1,...,\beta_{n_1})QD^{1/2}\begin{pmatrix} I_{n_1-1} & 0 \\ 0 & 0\end{pmatrix}D^{1/2}Q^t(\beta_1,...,\beta_{n_1})^t, \quad (3.20)\end{aligned}$$

where I_{n_1} and I_{n_1-1} are the $n_1\times n_1$ and $(n_1-1)\times(n_1-1)$ identity matrices, respectively, $z=(z_1,...,z_{n_1})^t$ is a $n_1\times q$ matrix, Q is an $n_1\times n_1$ orthogonal matrix and D is then$n_1\times n_1$ diagonal matrix of eigenvalues of $[I_{n_1} - z(z^tz)^-1z^t]$.

Consider the transformation

$$\boldsymbol{\eta}^t \equiv (\eta_1,\cdots,\eta_{n_1}) = (\beta_1,...,\beta_{n_1})QD^{1/2}, \quad (3.21)$$

where $\boldsymbol{\eta}$ is a $p\times n_1$ matrix. Then from (3.20) and (3.21),

$$\sum_{i=1}^{n_1}\beta_i\beta_i^t - \hat\Theta(\sum_{i=1}^{n_1} z_iz_i^t)\hat\Theta^t = \sum_{i=1}^{n_1-1}\eta_i\eta_i^t \text{ and } \sum_{i=1}^{n_1}\beta_i^t\beta_i \ge \frac{1}{d_{max}}\sum_{i=1}^{n_1}\eta_i^t\eta_i, \quad (3.22)$$

where d_{max} is the maximum value of elements of $D = diag(d_1,\cdots,d_{n_1-1},0)$. From (3.19) and (3.21)-(3.22), a bound of the marginal posterior density for $\eta_1,...,\eta_{n_1}$ is

$$\pi^*(\eta_1,..,\eta_{n_1}|data) \le c\cdot\left|\sum_{i=1}^{n_1-1}\eta_i\eta_i^t\right|^{-(n_1-p-1)/2}\times\left[\delta^* + \frac{\delta}{d_{max}}\sum_{i=1}^{n_1}\eta_i^t\eta_i\right]^{-(p^2+p+2n-2+\sum_{i=1}^n m_i)/2} \quad (3.23)$$

Integrating out η_{n_1} from (3.23) yields

$$\pi^*(\eta_1,..,\eta_{n_1-1}|data) \le c\cdot\left|\sum_{i=1}^{n_1-1}\eta_i\eta_i^t\right|^{-(n_1-p-1)/2}\cdot\left[\delta^* + \frac{\delta}{d_{max}}\sum_{i=1}^{n_1-1}\eta_i^t\eta_i\right]^{-(p^2+2n-2+\sum_{i=1}^n m_i)/2} \quad (3.24)$$

Now, the only thing left is to verify that $\pi^*(\eta_1, ..., \eta_{n_1-1}|data)$ is proper. Denote $\boldsymbol{\eta}_1 = (\eta_1, \cdots, \eta_{n_1-1})$ which is a $p \times (n_1 - 1)$ matrix. Then $\boldsymbol{\eta}_1$ can be decomposed as

$$\boldsymbol{\eta}_1 = P \begin{pmatrix} \xi_1 & 0 & \cdots & 0 & 0 \\ 0 & \xi_2 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \xi_p & 0 \end{pmatrix} \Gamma^t \equiv PT\Gamma^t, \tag{3.25}$$

where P and Γ are $p \times p$ and $(n_1 - 1) \times (n_1 - 1)$ orthogonal matrices and $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_p \geq 0$.

Rewrite Γ as $\Gamma = (\Gamma_1, \Gamma_2)$, where Γ_1, Γ_2 are $(n_1 - 1) \times p$ and $(n_1 - 1) \times (n_1 - p - 1)$ matrices, respectively. Note that $\boldsymbol{\eta}_1$ does not depend on Γ_2 since $\boldsymbol{\eta}_1 = P \cdot Diag(\xi_1, \cdots, \xi_p) \cdot \Gamma_1^t$. From here on, we assume that Γ_2 is a function of Γ_1 such that Γ is an orthogonal matrix.

Lemma 3.1 *The Jacobian of the orthogonal decomposition given by (3.25) is*

$$\left|\frac{\partial(\boldsymbol{\eta}_1)}{\partial(P, T, \Gamma_1)}\right| \propto (\prod_{i=1}^{p} \xi_i)^{n_1-p-1} \prod_{1 \leq i < j \leq p} \left|\xi_i^2 - \xi_j^2\right|.$$

Proof : Using the exterior product method in Muirhead (1982) and $\boldsymbol{\eta}_1 = PT\Gamma^t$,

$$d\boldsymbol{\eta}_1 = dP \cdot T \cdot \Gamma^t + P \cdot dT \cdot \Gamma^t + P \cdot T \cdot d\Gamma^t.$$

By the fact that $d\Gamma^t\Gamma = -\Gamma^t d\Gamma$,

$$P^t \cdot d\boldsymbol{\eta}_1 \cdot \Gamma = P^t \cdot dP \cdot T - T \cdot \Gamma^t \cdot d\Gamma + dT. \tag{3.26}$$

According to Muirhead (1982, Ch. 2), the exterior product of elements in the matrix on the left-hand side of (3.26) is

$$\left|P^t\right|^p \cdot |\Gamma|^{n_1-1} \cdot d\boldsymbol{\eta}_1 = d\boldsymbol{\eta}_1. \tag{3.27}$$

Write $P = (P_1, P_2, ..., P_p)$ and $\Gamma = (\gamma_1, \gamma_2, ..., \gamma_{n_1-1})$, then

$$P^t \cdot dP \cdot T = (P_1, ..., P_p)^t (dP_1, ..., dP_p) \begin{pmatrix} \xi_1 & 0 & \cdots & 0 & 0 \\ 0 & \xi_2 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \xi_p & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -\xi_2 P_2^t dP_1 & \cdots & -\xi_p P_p^t dP_1 & 0 \\ \xi_1 P_2^t dP_1 & 0 & \cdots & -\xi_p P_p^t dP_2 & 0 \\ \xi_1 P_3^t dP_1 & \xi_2 P_3^t dP_2 & \cdots & -\xi_p P_p^t dP_3 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \xi_1 P_p^t dP_1 & \xi_2 P_p^t dP_2 & \cdots & 0 & 0 \end{pmatrix}. \tag{3.28}$$

Similarly, we have

$$T \cdot \Gamma^t \cdot d\Gamma$$
$$= \begin{pmatrix} 0 & -\xi_1 \gamma_2^t d\gamma_1 & \cdots & -\xi_1 \gamma_p^t d\gamma_1 & -\xi_1 \gamma_p + 1^t d\gamma_1 & \cdots & -\xi_1 \gamma_{n_1-1}^t d\gamma_1 \\ \xi_2 \gamma_2^t d\gamma_1 & 0 & \cdots & -\xi_2 \gamma_p^t d\gamma_2 & -\xi_2 \gamma_p + 1^t d\gamma_2 & \cdots & -\xi_2 \gamma_{n_1-1}^t d\gamma_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \xi_p \gamma_p^t d\gamma_1 & \xi_p \gamma_p^t d\gamma_2 & \cdots & 0 & -\xi_p \gamma_p + 1^t d\gamma_p & \cdots & -\xi_p \gamma_{n_1-1}^t d\gamma_p \end{pmatrix} \tag{3.29}$$

and

$$dT = \begin{pmatrix} \xi_1 & 0 & \cdots & 0 & 0 \\ 0 & \xi_2 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \xi_p & 0 \end{pmatrix}. \tag{3.30}$$

From (3.28)-(3.30), the exterior product of the right hand side of (3.26) is

$$\begin{aligned} &\wedge_{i=1}^p d\xi_i \wedge_{1\le i<j\le p} \left[\left(\xi_i P_j^t dP_i - \xi_j \gamma_j^t d\gamma_i \right) \wedge \left(-\xi_j P_j^t dP_i + \xi_i \gamma_j^t d\gamma_i \right) \right] \\ &\quad \times \wedge_{1\le i\le p, p+1\le j\le n_1-1} \xi_i \gamma_j^t d\gamma_i \\ &= \left(\prod_{i=1}^p \xi_i \right)^{n_1-p-1} \prod_{1\le i<j\le p} \left| \xi_i^2 - \xi_j^2 \right| \quad \wedge_{i=1}^p d\xi_i \\ &\quad \times (P_j^t dP_i \wedge \gamma_j^t d\gamma_i) \cdot \wedge_{1\le i\le p, p+1\le j\le n_1-1} \gamma_j^t d\gamma_i. \end{aligned} \tag{3.31}$$

From (3.27) and (3.31),

$$\left| \frac{\partial(\ \eta_1)}{\partial(P, T, \Gamma_1)} \right| \ \propto\ (\prod_{i=1} p\xi_i)^{n_1-p-1} \prod_{1\le i<j\le p} \left| \xi_i^2 - \xi_j^2 \right|.$$

Now we can give a sufficient condition for the posterior to be proper.

Theorem 3.1 *If $n_1 \ge 2p+1$ and condition (3.18) is satisfied then the posterior, derived from the Jeffrey's prior, for the hierarchical conjoint model is proper.*

Proof : Using the decomposition given in (3.25), Lemma 3.1, and (3.24),

$$
\begin{aligned}
&\pi^*(\eta_1, ..., \eta_{n_1-1}|data)d\eta_1 \cdots d\eta_{n_1-1} \\
&\leq c \cdot \int (\xi_1^2 \cdots \xi_p^2)^{-(n_1-p-1)/2} \cdot \left[\delta^* + \delta \sum_{i=1}^{p} \xi_i^2\right]^{-(p^2+2n-2+\sum_{i=1}^{n} m_i)/2} \\
&\times \left|\frac{\partial(\eta_1)}{\partial(P, T, \Gamma_1)}\right| \cdot \left(\prod_{i=1}^{p} d\xi_i\right) \cdot (P^t dP) \cdot (\Gamma_1^t d\Gamma_1) \\
&\propto \left|\prod_{1\leq i<j\leq p} \xi_i^2 - \xi_j^2\right| \cdot \left[\delta^* + \delta \sum_{i=1}^{p} \xi_i^2\right]^{-(p^2+2n-2+\sum_{i=1}^{n} m_i)/2} \\
&\times \left(\prod_{i=1} p d\xi_i\right) \cdot (P^t dP) \cdot (\Gamma_1^t d\Gamma_1) \quad < \quad \infty,
\end{aligned}
$$

because $(P^t dP)$ and $(\Gamma_1^t d\Gamma_1)$ are finite Haar measure (Muirhead, 1982, Chap2).

References

Lenk, Peter J. , DeSarbo, Wayne S., Green, Paul E., and Young, Martin R. (1996), Hierarchical Bayes Conjoint Analysis : Recovery of Partworth Heterogenetiy from Reduced Experimental Designs, *Marketing Science*, **15**, pp173-191.

Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Tracy, D. S., and Jinadasa, K. G. (1988), Patterned Matrix Derivatives, *Canadian Journal of Statistics*, **16**, 411-418.

Wiens, D. P. (1985), On some pattern-reduction matrices which appear in statistics, *Linear Algebra Application*, **67**, 233-258.

Yang, Ruoyong and Chen, Ming-Hui (1995), Bayesian Analysis for Random Coefficient Regression Models Using Noninformative Priors, *Journal of Multivariate Analysis*, **55**, 283-311.