# Multivariate Regression Analysis of Panel Data with Binary Outcomes applied to Unemployment Data

**Claudia Czado**

**Department of Mathematics and Statistics, York University, 4700 Keele Street, North York, Ontario, M3J 1P3, Canada [1]**

## Summary

In panel studies binary outcome measures together with time stationary and time varying explanatory variables are collected over time on the same individual. Therefore, a regression analysis for this type of data must allow for the correlation among the outcomes of an individual. The multivariate probit model of Ashford and Sowden (1970) was the first regression model for multivariate binary responses. However, a likelihood analysis of the multivariate probit model with general correlation structure for higher dimensions is intractable due to the maximization over high dimensional integrals thus severely restricting ist applicability so far. Czado (1996) developed a Markov Chain Monte Carlo (MCMC) algorithm to overcome this difficulty. In this paper we present an application of this algorithm to unemployment data from the Panel Study of Income Dynamics involving 11 waves of the panel study. In addition we adapt Bayesian model checking techniques based on the posterior predictive distribution (see for example Gelman et al. (1996)) for the multivariate probit model. These help to identify mean and correlation specification which fit the data well.

**Keywords:** Binary time series, panel data, multivariate probit model, Bayesian analysis, Markov Chain Monte Carlo methods, Bayesian model checking, posterior predictive distribution

[1] C. Czado was supported by research grant OGP0089858 of the Natural Sciences and Engineering Research Council of Canada. email: cczado@ma.tum.de

# 1 Introduction

Short time series of binary outcomes in the presence of covariate information are often observed in panel studies such as the well known Panel Study of Income Dynamics (PSID) conducted at the Survey Research Center, Institute of Social Research, University of Michigian, U.S.A. The primary interest in such studies is often the assessment of covariate effects. For this purpose, population averaged or marginal approaches are preferred over transitional or cluster specific approaches (for a review see Pendergast et al. (1996) and Ashby et al. (1992)) The reason for this preference is the ability to interpret the covariate effects unconditionally. For this reason we will not consider lagged response values as covariates in the mean specification in this paper. This excludes models proposed by Heckman and Borjas (1980).

Since the binary outcome is collected over time on the same individual, the binary outcomes are correlated. One naive approach is to ignore this correlation and conduct univariate standard analyses such as the probit or logistic regression. It has been noted by Liang and Zeger (1986, Theorem 1) that even though parameter estimates from univariate analyses ignoring the correlation remain consistent but they are inefficient when the correlation is large (see also Spiess et al. (1996)). This loss in efficiency might lead to overestimating the strength of covariate effects. Liang and Zeger (1986) proposed the use of generalized estimating equations (GEE) instead, which have been extended and used extensively (for example Lipsitz et al. (1991), Liang et al. (1992), Carey et al. (1993), Fitzmaurice and Lipsitz (1995) and Lipsitz et al. (1995), Spiess and Hamerle (1996)). Loss of efficiency also occurs in GEE models with estimated association parameters when time varying covariates are present (Fitzmaurice et al. (1993)). GEE models with estimated association parameters are called GEE2.

More recently, a preference for likelihood based methods for marginal models over the non likelihood based GEE method has been expressed (see for example the comments to Liang et al. (1992)). The earliest likelihood based model for correlated binary regression has been developed long before GEE's by Ashford and Sowden (1970). See for example Amemiya (1986) for a discussion of this approach in econometrics. In the special case of exchangeable correlation or equicorrelation an approximate maximum likelihood analysis has been tractable even in high dimensions (see Ochi and Prentice (1984)), but an exact maximum likelihood analysis of the multivariate probit model has been intractable for dimensions higher than three (see Anderson and Pemberton (1985)). In contrast to GEE models the multivariate probit model is likelihood based.

More recently, two different likelihood based models have been proposed. Both use odds ratios as measures of association between discrete variables. The one model developed by Molenberghs and Lesaffre (1994) is based on marginal odds ratios using a multivariate extension to the bivariate Plackett distribution (Plackett (1965)) for the construction of the joint likelihood. The other model put forward by Fitzmaurice and Laird (1993) for binary time series is formulated in terms of conditional odds ratios assuming a quadratic exponential model for the joint likelihood (Cox (1972), Zhao and Prentice (1990)). The extension of this approach to the ordinal response has been considered by Heagerty and Zeger (1996) and Heumann (1996). It should be noted that these likelihood based models are not easily formulated and while more general require the specification of higher order association parameters in contrast to the multivariate probit model.

Markov Chain Monte Carlo (MCMC) methods have been used very successfully for the analysis of many previously intractable problems (for example see Besag et al. (1995) and the many references cited therein) and have become a standard tool for statistical model analysts (see the recent books of Gelman et al. (1995), Gelfand and Smith (1995) and Gilks et al. (1996)). In this paper we present the results of a multivariate probit analysis using MCMC methods for a dataset studying the unemployment dynamics of a group of individuals followed for 11 years, thus demonstrating the tractability of the multivariate probit analysis with general correlation structure in high dimensions.

For the analysis of panel data with discrete response a more restricted alternative to the multivariate probit model are panel probit models allowing for a subject specific random effect (Hsiao (1986), Baltagi (1996)). These models are not marginal models and are only comparable to multivariate probit models with exchangeable correlation. Further, these panel probit models with random effect remain tractable in high dimensions since maximization over high dimensional integrals can be reduced to maximization over one dimensional integrals, where Gaussian quadrature can be applied in this situation (see Butler and Moffit (1982)). Another nonlikelihood based alternative for this situation which avoids numerical integration is to use a minimum-distance estimator (see Chamberlain (1984)).

The paper is organized as follows. In Section 2 details on the unemployment data and the results of an initial explanatory analysis are presented.

In Section 3, the multivariate probit model and its analysis based on MCMC methods are introduced. The results and their interpretation for the unemployment data are given in Section 4. Model checking for the multivariate probit model based on the posterior predictive distribution is developed and discussed in Section 5.

## 2 Description of the Unemployment Data

This section provides details and some results of an explanatory data analysis of the unemployment data analyzed later. The data collected is part of the Panel Study of Income Dynamics (PSID) conducted at the Survey Research Center, Institute of Social Research, University of Michigian, U.S.A. (http://www.umich.edu/˜psid/). This panel study emphasizes the dynamic aspects of economic and demographic behavior.

We are interested in investigating the unemployment dynamics of individuals who were initially unemployed but remained available to the labor market during the whole study period. We excluded individuals who retired or became permanently disabled during the study period as well as individuals who became house keepers and students at any time period during the course of study. In 1981, 837 individuals interviewed for the PSID reported that they were looking for work. Of these 837 individuals, 166 remained available to the labor market until 1992. In addition to the 11 measurements of the yearly employment status (1 = working, 0 = looking for work or temporarily laid off), the gender (1 = male, 0 = female), the age in 1981 in years and the number of actual grades of schooling completed in 1981 were reported. This last measurement will be taken as an indicator of the level of education of an individual. Primary interest is the modeling of the time dependency of the unemployment dynamics for this group of individuals while adjusting for gender, education and age. In this paper the probability of being unemployed is investigated while the duration of unemployment could also be studied (see for example Niesing et al. (1994)). We will now present the results of an explanatory data analysis to help us formulate reasonable marginal models and models for the association present among the responses. To assess the effect of time for the marginal model, Figure 2.1 plots the proportion of unemployed individuals for each year from 1982 to 1992. It clearly shows a nonlinear time trend. We are interested in formulating a model which allows for probit margins. The probit scale, i.e. $\Phi^{-1}(p)$ where $\Phi$ is the standard normal cdf, is therefore the appropriate scale to assess the effects of the

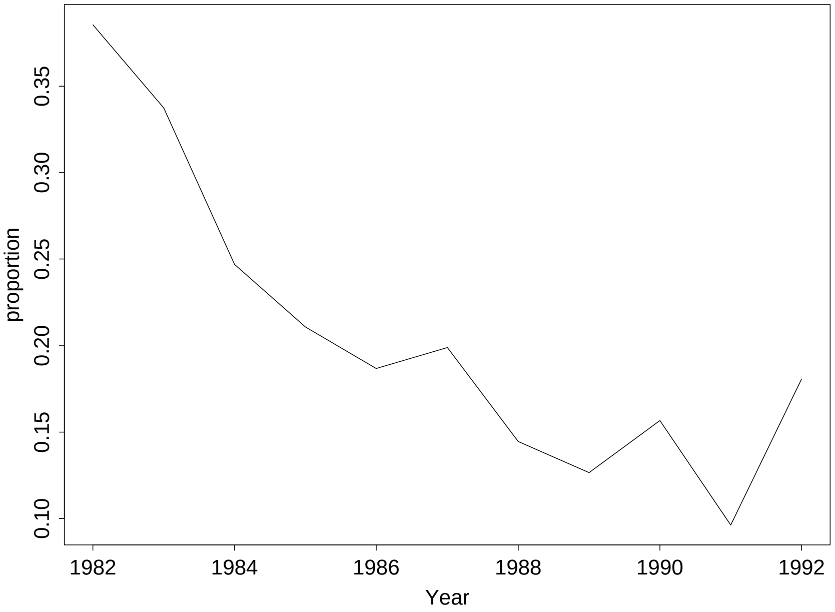

Figure 2.1: Proportion of Unemployment over Time.

covariates. We use 3 age groups (19-25 years, 26-35 years and older than 36 years) to estimate the probit of being unemployed. Similarly we used three education categories (less than 12 years of schooling, 12 years of schooling and more than 12 years of schooling). The estimated probits of being unemployed for each year of each of the covariates are plotted in Figure 2.2. The effect of gender seems to be smaller than the effects of age and education since the lines for age and education are further apart than the lines for gender. Since the lines are somewhat parallel for all panels, interaction between the time stationary covariates gender, age in 1981 and education level in 1981, and the time effect seem to be not present in this data set. We used standard univariate probit analysis ignoring the dependence among responses from the same subject to screen for interaction effects among time stationary covariates. This is reasonable since in general the presence of correlation will reduce the significance of effects. The results of this approach show the presence of an interaction effect between age and education, while other interactions are insignificant.

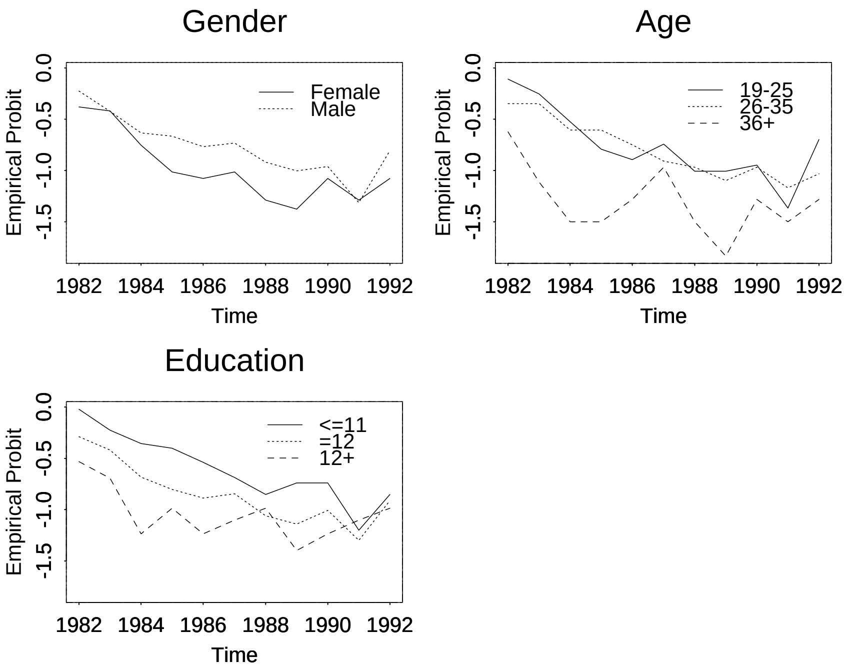

Figure 2.2: Empirical Probits over Time classified by Gender, Age and Education, respectively.

| | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 |
|---|---|---|---|---|---|---|---|---|---|---|
| 83 | 6.82 | | | | | | | | | |
| 84 | 4.59 | 10.38 | | | | | | | | |
| 85 | 3.07 | 4.83 | 6.98 | | | | | | | |
| 86 | 3.76 | 3.53 | 3.93 | 14.55 | | | | | | |
| 87 | 2.27 | 3.00 | 4.83 | 3.93 | 7.40 | | | | | |
| 88 | 2.11 | 4.11 | 2.56 | 5.17 | 3.27 | 9.06 | | | | |
| 89 | 3.80 | 3.85 | 6.79 | 4.36 | 3.26 | 3.78 | 6.50 | | | |
| 90 | 2.10 | 4.89 | 4.88 | 2.88 | 4.40 | 4.86 | 5.62 | 12.48 | | |
| 91 | .70 | 2.12 | 1.97 | 3.39 | 1.01 | 1.39 | 3.13 | 3.81 | 3.90 | |
| 92 | 1.51 | 3.87 | 2.46 | 2.22 | 1.11 | 3.65 | 2.13 | 5.58 | 4.67 | 7.9 |

Table 2.1 : Estimated Odds of being Unemployed between Pairs of Years

We turn now to the effects of correlation among the responses. The estimated ratio between the odds of being unemployed in year $i$ and the odds of being unemployed in year $j$ for $i, j = 1982, \cdots, 1992$ are presented in Table 2.1. This shows that there is some tendency that the associations between the responses decrease as the difference in years increases.

# 3 Multivariate Probit Model for Short Binary Time Series

In this section, the multivariate probit model for binary time series will be formulated and its analysis using MCMC methods are presented. We will assume that the binary time series is completely observed. Approaches on how to handle the case where responses are missing are discussed in Czado (1996).

## 3.1 Model Formulation

To formulate a Bayesian approach, we need to specify the joint distribution of the binary response vector. For this, let $Y_i = (y_{i1}, \cdots, y_{iT})^t$ the binary response vector with binary response, $y_{it} = 1$ or 0, observed at time t and marginal probabilities $\pi_{it} = P(y_{it} = 1)$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$. We assume, that the response vectors $Y_i$ are independently observed. For each response component $y_{it}$, we have covariate information collected in the vector $(x_{it1}, \cdots, x_{itp})$ available. Some of these covariates might be time stationary. For example, if the jth covariate is time stationary, we have $x_{i1j} = \cdots = x_{iTj}$. We consider now marginal models of the following form

$$\pi_{it} = \Phi(\eta_{it}) \text{ where } \eta_{it}(\boldsymbol{\beta}) = \beta_{0t} + \beta_{1t}x_{it1} + \cdots + \beta_{pt}x_{itp} \qquad (1)$$

and $\Phi(\cdot)$ denotes the standard normal distribution function. This formulation is the most general, since it allows for both time varying regression parameters $\beta_{jt}$ as well as time varying covariates. Time stationary regression parameters can be achieved by requiring $\beta_{j1} = \cdots = \beta_{jT} = \beta_j$. For the unemployment data, the models considered will include time varying covariates but only time stationary regression parameters are used.

To give the complete specification of the joint distribution, we introduce independent latent random vectors $Z_i = (Z_{i1}, \cdots, Z_{iT})$ which are jointly normally distributed with mean vector

$$\mu(\beta) = -\eta_i(\beta) = (-\eta_{i1}(\boldsymbol{\beta}), \cdots, -\eta_{iT}(\boldsymbol{\beta}))^t \qquad (2)$$

and covariance matrix $\Sigma_i$ with unit diagonal entries. The dependence between the binary outcomes $y_{it}$ is modeled indirectly through the dependence structure among the latent variables $Z_{it}$. For this, we assume that

$$y_{it} = 1 \Longleftrightarrow Z_{it} < 0.$$

It is easy to see that this equivalence is consistent with the marginal specification given in (1). Note that the latent variable $Z_{it}$ can be interpreted as an unobservable threshold for the response $y_{it}$.

Joint probabilities can now be determined by the joint distribution of $Z_i$. For example

$$\begin{aligned} P(y_{i1} = 1, \cdots, y_{iT} = 1) &= P(Z_{i1} < 0, \cdots, Z_{iT} < 0) \\ &= \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} f(\beta, \Sigma_i, Z_i) dZ_{i1} \cdots dZ_{iT}, \end{aligned} \tag{3}$$

where

$$f(\beta, \Sigma_i, Z_i) = \frac{1}{(2\pi)^{T/2} |\Sigma_i|^{1/2}} \exp\{-\frac{1}{2}(Z_i - \mu_i(\beta))^t \Sigma_i^{-1} (Z_i - \mu_i(\beta))\}.$$

Other joint probabilities can be defined similarly. Equation (3) demonstrates why a straight forward likelihood analysis of the multivariate probit model is intractable. If one is interested in achieving logistic margins, one can, as suggested by Cessie and Houwelingen (1994), set $\mu(\beta)$ to

$$\mu(\beta) = (-\Phi^{-1}(\frac{\exp(\eta_{i1}(\beta))}{1 + \exp(\eta_{i1}(\beta))}),, \cdots, -\Phi^{-1}(\frac{\exp(\eta_{iT}(\beta))}{1 + \exp(\eta_{iT}(\beta))}))^t \tag{4}$$

in (2).

The specification of the dependence structure $\Sigma_i$ allows for a wide range of association models. We present now some possibilities:

**(i)** Covariate independence: $\Sigma_i = \Sigma$

**(ii)** Serial correlation pattern with covariate independence:
$Cor(Z_{is}, Z_{it}) = \rho^{|s-t|}$.

**(iii)** Exchangeable correlation pattern with covariate independence:
$Cor(Z_{is}, Z_{it}) = \rho$.

Pattern (ii) has been used by Fitzmaurice and Lipsitz (1995) for odds ratios. It is also appropriate, when the binary responses are measured at unequally spaced time points.

Since the covariance matrix $\Sigma_i$ has unit diagonal entries, $\Sigma_i$ is the correlation matrix of the latent vector $Z_i$, therefore the (s,t)th element of $\Sigma_i$, denoted by $\rho_{ist}$, is restricted to the interval [-1,1]. It is easier to consider a transformation of $\rho_{ist}$ to the real line for incorporating covariate dependence of the correlation structure. Cessie and Houwelingen (1994) used the following one-to-one transformation

$$\tau_{ist} = \log(\frac{1+\rho_{ist}}{1-\rho_{ist}}).$$

A regression model for $\tau_{ist}$ can now be assumed, for example

$$\tau_{ist} = \alpha_{st0} + \alpha_{st1} W_i, \tag{5}$$

where $W_i$ is an appropriate covariate. Additional covariates for the association structure can be incorporated in the same way. Marginal parameters as defined in (2) or (4) will be denoted by $\boldsymbol{\beta}$, while the association parameters defined in (5) will be denoted by $\boldsymbol{\alpha}$. Since the covariance matrices $\Sigma_i$ depend on $\boldsymbol{\alpha}$, we will denote them with $\Sigma_i(\boldsymbol{\alpha})$.

## 3.2 Bayesian Inference using Monte Carlo Markov Chain Methods

For the Bayesian analysis, we assume that the response $Y_i$ given the regression parameters $\boldsymbol{\beta}$ and the association parameters $\boldsymbol{\alpha}$ follow the multivariate probit model as specified in (1),(2) and (5). A model for logistic margins is achieved by using (4) instead of (2). The prior information about $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is summarized in a joint density of the form $\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \pi(\beta) \times \pi(\alpha)$. Noninformative and multivariate normal priors can be used.

MCMC methods allow to draw a sample from the posterior distribution $[\beta, \boldsymbol{\alpha}, Z|Y]$, where $Z = (Z_1, \cdots, Z_n)^t$ and $Y = (Y_1, \cdots, Y_n)^t$. Here, $[u|w]$ denotes the conditional distribution of u given w. A Metropolis within Gibbs approach (Müller (1994)) is now taken, since the conditional distributions $[Z_i|Y_i, \boldsymbol{\beta}, \boldsymbol{\alpha}]$ and $[\boldsymbol{\beta}|\alpha, Z, Y]$ are known when (2) holds, while $[\alpha|\boldsymbol{\beta}, Z, Y]$ and $[\boldsymbol{\beta}|\boldsymbol{\alpha}, Z, Y]$ are known only up to a normalizing constant when (4) holds, thus requiring a Metropolis-Hastings step. The reader unfamilar with MCMC methods can consult Gilks et al. (1996) for an introduction to the Gibbs sampler and the Metropolis Hastings algorithm.

It is easy to see that, $[Z_i|Y_i, \boldsymbol{\beta}, \boldsymbol{\alpha}]$ is a truncated multivariate normal distribution with mean vector $\mu(\boldsymbol{\beta})$ and covariance matrix $\Sigma_i(\alpha)$ truncated to the rectangular area given by $[log(1-y_{i1}), -log(y_{i1})] \times \cdots \times [log(1-y_{iT}), -log(y_{iT})]$. Note, that $\eta_i(\boldsymbol{\beta})$ and $\Sigma_i(\boldsymbol{\alpha})$ are determined by $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, respectively. For the generation of truncated multivariate random variables, we followed the approach of Robert (1995) (see also Geweke (1991)). It is MCMC based and uses a Gibbs sampling scheme to simulate from univariate truncated conditionals. An accept-reject algorithm for the tails of the univariate truncated normals is then utilised. This is a different approach as the one proposed in Czado (1996), which resulted in highly biased association estimates in simulations, when high correlations were present.

We derive now the conditional distribution $[\boldsymbol{\beta}|\boldsymbol{\alpha}, Z, Y]$ when mean specification (2) holds. First note that $[\beta|\boldsymbol{\alpha}, Z, Y] = [\beta|\boldsymbol{\alpha}, Z]$ since $Z$ determines $Y$. But finding $[\beta|\boldsymbol{\alpha}, Z]$ is now equivalent to finding the posterior distribution of the regression parameters in a linear regression model (see for example Lee (1997)). It is multivariate normal with mean vector $-X\boldsymbol{\beta}$, where $X$ is a block diagonal matrix with ith block given by

$$X_i = \begin{pmatrix} 1 & x_{i11} & \cdots & x_{i1p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{iT1} & \cdots & x_{iTp} \end{pmatrix}$$

and block diagonal covariance matrix $\Sigma(\boldsymbol{\alpha})$ with ith block given by $\Sigma_i(\boldsymbol{\alpha})$. In the case of a multivariate normal prior for $\beta$ with mean vector $\beta_p$ and covariance matrix $\Sigma_p$, it is straight forward to determine that $[\beta|Z, \boldsymbol{\alpha}]$ is again multivariate normal with mean vector

$$-(\Sigma_p^{-1} + X^t\Sigma(\boldsymbol{\alpha})^{-1}X)^{-1}(\Sigma_p^{-1}\beta_p + X^t\Sigma(\boldsymbol{\alpha})^{-1}Z) \tag{6}$$

and covariance matrix

$$(\Sigma_p^{-1} + X^t\Sigma(\boldsymbol{\alpha})^{-1}X)^{-1}. \tag{7}$$

For a flat prior the terms involving the prior parameters $\beta_p$ and $\Sigma_p$ vanish.

For the case of logistic margins (4), $[\beta|\boldsymbol{\alpha}, Z, Y]$ is only known up to a normalizing constant, thus requiring a Metropolis-Hastings Update. Since

$$\Phi^{-1}(\frac{\exp(z)}{1+\exp(z)}) \simeq \frac{\sqrt{2\pi}}{4}z$$

we approximate the distribution of $Z_i$ by a multivariate normal with mean vector $-\frac{\sqrt{2\pi}}{4}\eta_i(\boldsymbol{\beta})$. Therefore we choose as proposal distribution for $\boldsymbol{\beta}$ a multivariate normal cdf with mean and covariance given by (6) and (7), respectively, where the design matrix X is changed to $\frac{\sqrt{2\pi}}{4}X$.

For updating the association parameters $\alpha$, we also require a Metropolis-Hastings update. Here, the density of $[\boldsymbol{\alpha}|\boldsymbol{\beta}, Z, Y]$ is proportional to the density of $[Z|\boldsymbol{\alpha}, \boldsymbol{\beta}]$ considered as function of $\alpha$. A normal proposal density with same mode as $[\boldsymbol{\alpha}|\boldsymbol{\beta}, Z, Y]$ and a user controlled covariance matrix is used for the corresponding Metropolis-Hastings step.

Using the above conditionals, an approximate sample from the posterior can be drawn and point and interval estimates of the parameters can be calculated using this sample. It should be noted that the algorithm can also be used for data with varying cluster sizes. This approach was first suggested by Czado (1996).

# 4 Multivariate Regression Analysis of the Unemployment Data

Using the findings of the explanatory analysis we fitted the following mean specifications for the multivariate probit model ((1), (2) and (5)):

$$\begin{aligned}\eta_{it}(\beta) = {}& \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Education}_i + \beta_3 \text{Gender}_i + \beta_4 \text{Time}_t \\ & + \beta_5 \text{Time}_t^2 + \beta_6 \text{Education*Age}_i, \end{aligned} \tag{8}$$

where Time is coded as 0 to 10 for 1982 to 1992. This mean specification takes into effect that the explanatory analysis indicates a quadratic time effect and an interaction effect between education level and age. For a second mean specification we allowed individual year effects. For this we defined the following dummy indicators

$$It = \begin{cases} 1 \text{ if year is } t \\ 0 \text{ otherwise} \end{cases}$$

for $t = 1983, \cdots, 1992$. Using this we investigated also the following mean specification

$$\begin{aligned}\eta_{it}(\beta) = {}& \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Education}_i + \beta_3 \text{Gender}_i \\ & + \beta_4 \text{I1983}_t + \cdots \beta_{13} \text{I1992}_t + \beta_{14} \text{Education*Age}_i. \end{aligned} \tag{9}$$

We would like to note that the nonlinear time effect could also be modelled by semiparametric approaches. Nonparametric smoothing methods as studied by Rice and Silverman (1991) or additive models as proposed by Hastie and Tibsherani (1990) could possibly be adopted to this situation. However, we feel that the binary time series involved are too short to warrant such modelling in this situation.

For the association models we studied the serial (see (ii)) and exchangeable (see (iii)) correlation. Finally we assumed a flat prior for both the mean and association parameters. 2000 iterations of the Markov Chain Sampler described in Section 3.2 were run for mean specification (8) and (9), respectively. We monitored the convergence of the sampler using the diagnostic measures implemented in the Splus library coda() of Best et al (1995) and described in more detail in Cowles and Carlin (1995). They show a very small burn in effect ($< 10$ iterations). A slower mixing of the chains especially for the estimation of $\rho$ was observed (lag 1 autocorrelations for $\rho > .8$ for all models). Therefore subsampling of every 5th iteration after first 50 iterations were discarded was applied. The results are presented in Table 4.1 and 4.3. The tables give the posterior mean estimates and a 95% Bayes credible interval based on estimates of the 2.5% and 97.5% quantiles. For comparison we present the corresponding results of GEE analyses for the exchangeable correlation. For the GEE analysis the robust estimates of the regression parameters were used for the interval estimate. The GEE2 analysis was performed using the Gauss program of Y.Qu (see Qu et al. 1995).

| | **Serial Correlation** | | | **Exchangeable Correlation** | | |
|---|---|---|---|---|---|---|
| | Estimate | 95% Bayes CI | | Estimate | 95% Bayes CI | |
| | | lower limit | upper limit | | lower limit | upper limit |
| Intercept | -2.760 | -3.840 | -1.780 | -2.660 | -3.810 | -1.470 |
| Age | 0.066 | 0.038 | 0.097 | 0.063 | 0.028 | 0.093 |
| Education | 0.217 | 0.131 | 0.311 | 0.208 | 0.106 | 0.306 |
| Gender | -0.074 | -0.188 | 0.036 | -0.065 | -0.163 | 0.053 |
| Time | 0.219 | 0.172 | 0.268 | 0.215 | 0.173 | 0.256 |
| Time$^2$ | -0.015 | -0.019 | -0.010 | -0.014 | -0.018 | -0.010 |
| Education*Age | -0.004 | -0.007 | -0.002 | -0.004 | -0.007 | -0.001 |
| $\rho$ | 0.649 | 0.578 | 0.717 | 0.384 | 0.284 | 0.471 |

Table 4.1: Mean and Association Parameter Estimates fitting the Multivariate Probit Model for the Unemployment Data based on Mean Specification with Quadratic Time Effect (see (8))

From the results of the multivariate probit analyses we see strong evidence for a nonlinear time trend, since the interval estimate for the quadratic time effect does not include zero. The year specific mean formulation (9) allows us to assess the effects of individual years separately. In general, one can say that the chances of employment increases until about 1989, while they drop in 1990, recover in 1991 and drop again in 1992 to 1986/87 levels for the study population.

As expected the effects of age and education are positive thus showing that employment chances improve over age and education level, however there is some evidence of an interaction between education level and age, thus decreasing the chances of getting employed for well educated older individuals. To see this effect, Figure 4.1 gives the contour lines of the fitted odds of being employed for several years based on the model (8) with exchangeable correlation structure. For models with no interaction the contour levels would have been parallel lines.

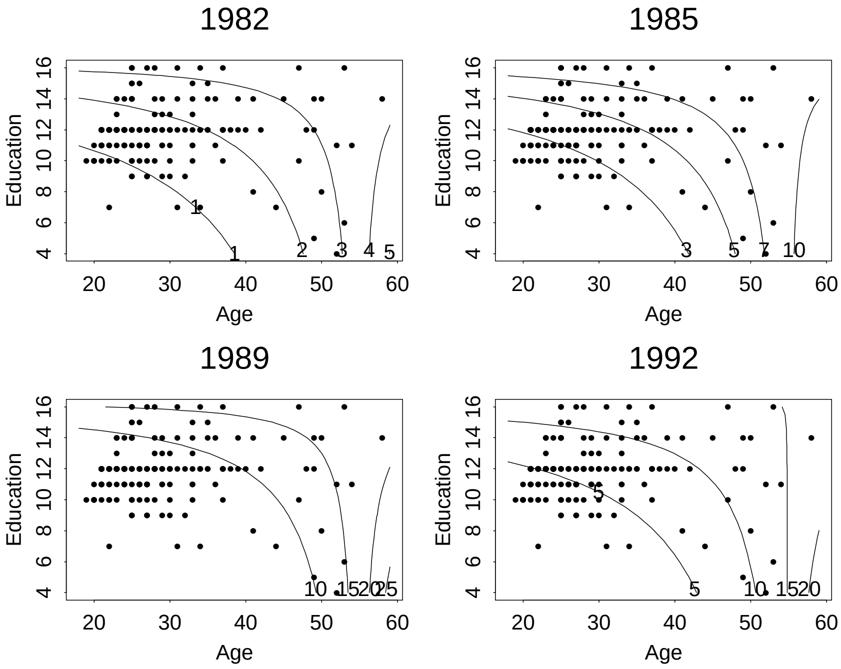

Figure 4.1: Contour Lines for the fitted Odds of being Employed as a Function of Age and Gender for the Quadratic Time Effects Model with Exchangeable Correlation.

Turning now to the effect of gender on unemployment, all analyses show that there is no significant effect of gender for the study population. Note that this does not mean that there is no effect of gender in general. Recall that housekeepers were removed from the study and women are likely to represent the majority of this group. However no gender differences were detected among this group of unemployed individuals who stayed available to the labor market during the whole study period.

| | **Exchangeable Correlation Structure** | | | | | |
|---|---|---|---|---|---|---|
| | **GEE** | | | **GEE2** | | |
| | Estimate | 95% Robust CI | | Estimate | 95% CI | |
| | | lower limit | upper limit | | lower limit | upper limit |
| Intercept | -3.340 | -5.940 | -0.740 | -3.151 | -5.728 | -0.574 |
| Age | 0.080 | 0.015 | 0.146 | 0.076 | 0.011 | 0.141 |
| Education | 0.268 | 0.047 | 0.490 | 0.253 | 0.033 | 0.472 |
| Gender | -0.077 | -0.335 | 0.182 | -0.090 | -0.346 | 0.165 |
| Time | 0.213 | 0.133 | 0.293 | 0.212 | 0.133 | 0.292 |
| Time$^2$ | -0.014 | -0.022 | -0.006 | -0.014 | -0.022 | -0.006 |
| Education*Age | -0.005 | -0.011 | 0.000 | -0.005 | -0.011 | 0.001 |
| $\rho$ | 0.202 | - | - | 0.397 | 0.313 | 0.480 |

Table 4.2: Mean and Association Parameter Estimates using the GEE Approach for the Unemployment Data based on Mean Specification with Quadratic Time Effect (see (8))

We are also aware that the results presented are based on a very limited set of economic determinants. Further, the number of observations are quite small, thus a sample selection bias cannot be excluded. Therefore the interpretation of these results has to proceed with caution.

With regard to the correlation among the response variables, the results for the multivariate probit analysis with exchangeable correlation structure shows moderate correlation, while with serial correlation structure the correlation is even larger. The association parameter estimate for the GEE2 approach is of the same magnitude as the one from the multivariate probit, while the GEE approach gives much lower estimates. This is also an indication that the original GEE approach is inefficient for the estimation of the association parameters.

The results for the two different correlation structures differ little with respect to the estimated mean parameters and their interval estimates, thus indicating some measure of robustness with regard to the specification of the correlation structure.

Comparing the results of the multivariate probit models to the corresponding ones of the GEE models, one sees very little difference between the estimated parameter values. The GEE interval estimates are much wider than the multivariate probit ones, which is to be expected since the GEE ones are robust with regard to the misspecification of the working correlation, while the Bayes interval estimates assume that the correlation among the latent variables is correctly specified. The length of the GEE2 interval estimates are somewhat lower than the ones for the GEE approach, but still much larger than the ones from the multivariate probit analysis. Even though the GEE analysis gives similar results, we remind the reader that the GEE approach does not fully specify a statistical model, while the multivariate probit does.

Finally we compare our results to the ones achieved assuming independence among the response variables (see Table 4.5). We observe that the regression parameter estimates are of the magnitude than the ones from the multivariate probit analysis. In this example we do not observe an overestimating of the strength of covariate effects, when the correlation is ignored, since interval estimates are of comparable lengths.

| | **Serial Correlation** | | | **Exchangeable Correlation** | | |
|---|---|---|---|---|---|---|
| | Estimate | 95% Bayes CI | | Estimate | 95% Bayes CI | |
| | | lower limit | upper limit | | lower limit | upper limit |
| Intercept | -2.710 | -3.980 | -1.500 | -2.680 | -3.860 | -1.550 |
| Age | 0.066 | 0.034 | 0.102 | 0.064 | 0.034 | 0.101 |
| Education | 0.213 | 0.109 | 0.320 | 0.210 | 0.117 | 0.311 |
| Gender | -0.073 | -0.174 | 0.034 | -0.060 | -0.171 | 0.042 |
| I1983 | 0.133 | 0.009 | 0.258 | 0.140 | -0.006 | 0.270 |
| I1984 | 0.408 | 0.258 | 0.558 | 0.413 | 0.250 | 0.557 |
| I1985 | 0.535 | 0.375 | 0.703 | 0.540 | 0.392 | 0.695 |
| I1986 | 0.624 | 0.462 | 0.776 | 0.617 | 0.477 | 0.756 |
| I1987 | 0.567 | 0.410 | 0.706 | 0.565 | 0.414 | 0.715 |
| I1988 | 0.782 | 0.601 | 0.942 | 0.798 | 0.649 | 0.959 |
| I1989 | 0.869 | 0.720 | 1.040 | 0.895 | 0.728 | 1.070 |
| I1990 | 0.752 | 0.579 | 0.913 | 0.745 | 0.577 | 0.918 |
| I1991 | 1.020 | 0.835 | 1.210 | 1.010 | 0.824 | 1.220 |
| I1992 | 0.641 | 0.496 | 0.810 | 0.642 | 0.490 | 0.790 |
| Education*Age | -0.004 | -0.007 | -0.001 | -0.004 | -0.007 | -0.001 |
| $\rho$ | 0.663 | 0.605 | 0.725 | 0.404 | 0.321 | 0.491 |

Table 4.3: Mean and Association Parameter Estimates of the Multivariate Probit Model for the Unemployment Data based on Mean Specification with Year Specific Time Effects (see (9))

| | **Exchangeable Correlation Structure** | | | | | |
|---|---|---|---|---|---|---|
| | **GEE** | | | **GEE2** | | |
| | Estimate | 95% Robust CI | | Estimate | 95% CI | |
| | | lower limit | upper limit | | lower limit | upper limit |
| Intercept | -3.469 | -6.046 | -0.890 | -3.168 | -5.749 | -0.586 |
| Age | 0.085 | 0.021 | 0.150 | 0.077 | 0.012 | 0.142 |
| Education | 0.280 | 0.060 | 0.500 | 0.255 | 0.036 | 0.475 |
| Gender | -0.077 | -0.336 | 0.180 | -0.090 | -0.347 | 0.166 |
| I1983 | 0.133 | -0.083 | 0.350 | 0.133 | -0.082 | 0.348 |
| I1984 | 0.413 | 0.172 | 0.650 | 0.409 | 0.169 | 0.649 |
| I1985 | 0.529 | 0.264 | 0.790 | 0.529 | 0.266 | 0.791 |
| I1986 | 0.618 | 0.354 | 0.881 | 0.617 | 0.356 | 0.879 |
| I1987 | 0.571 | 0.295 | 0.847 | 0.571 | 0.297 | 0.845 |
| I1988 | 0.789 | 0.493 | 1.090 | 0.790 | 0.496 | 1.084 |
| I1989 | 0.883 | 0.597 | 1.170 | 0.880 | 0.596 | 1.163 |
| I1990 | 0.737 | 0.447 | 1.030 | 0.738 | 0.449 | 1.026 |
| I1991 | 1.023 | 0.671 | 1.380 | 1.030 | 0.680 | 1.380 |
| I1992 | 0.631 | 0.336 | 0.926 | 0.636 | 0.342 | 0.929 |
| Education*Age | -0.006 | -0.011 | -0.000 | -0.005 | -0.011 | 0.001 |
| $\rho$ | 0.202 | - | - | 0.401 | 0.486 | 0.317 |

Table 4.4: Mean and Association Parameter Estimates using the GEE Approach for the Unemployment Data based on Mean Specification with Year Specific Time Effects (see (9))

# 5 Model Checking and Discussion

The application presented in the last section demonstrates that MCMC methods can be used to achieve a tractable analysis of the multivariate probit model. This allows us to fit models with a wide range of mean and association parameter specifications. Therefore, it is important to be able to check the fit of a particular model to the data. In the context of the unemployment data we are interested in assessing and comparing the fit of the two mean specifications (year specific or quadratic time effect) as well as the two association structure specifications (serial or exchangeable correlation). For this we discuss now how a Bayesian test of model fit based on the posterior predictive distribution can be used in the context of the multivariate probit model. The use of posterior predictive distributions for model checking was first proposed and applied by Guttman (1967) and Rubin (1981, 1984). An introduction and general discussion of this method is given by Gelman and Meng in Gilks et al. (1996, Chapter 11) and Gelman et al. (1996).

| | **Quadratic Time Effect** | | |
|---|---|---|---|
| | Estimate | 95% CI | |
| | | lower limit | upper limit |
| Intercept | -2.834 | -4.229 | -1.439 |
| Age | 0.067 | 0.030 | 0.104 |
| Education | 0.229 | 0.110 | 0.348 |
| Gender | -0.101 | -0.243 | 0.040 |
| Time | 0.213 | 0.136 | 0.289 |
| Time$^2$ | -0.014 | -0.021 | -0.006 |
| Education*Age | -0.004 | -0.008 | -0.001 |

| | **Year Specific Time Effect** | | |
|---|---|---|---|
| | Estimate | 95% CI | |
| | | lower limit | upper limit |
| Intercept | -2.815 | -4.218 | -1.413 |
| Age | 0.067 | 0.030 | 0.104 |
| Education | 0.228 | 0.109 | 0.348 |
| Gender | -0.101 | -0.243 | 0.040 |
| I1983 | 0.133 | -0.147 | 0.412 |
| I1984 | 0.411 | 0.123 | 0.699 |
| I1985 | 0.528 | 0.235 | 0.821 |
| I1986 | 0.617 | 0.319 | 0.914 |
| I1987 | 0.569 | 0.274 | 0.865 |
| I1988 | 0.788 | 0.479 | 1.096 |
| I1989 | 0.881 | 0.565 | 1.198 |
| I1990 | 0.735 | 0.430 | 1.040 |
| I1991 | 1.022 | 0.692 | 1.352 |
| I1992 | 0.631 | 0.332 | 0.930 |
| Education*Age | -0.004 | -0.008 | -0.001 |

Table 4.3: Mean Parameter Estimates for the Unemployment Data assuming Independence among the Responses

For posterior predictive model checking we require the specification of a discrepancy measure, which in contrast to classical test statistics can depend on unknown model parameters in addition to the observed data. The discrepancy measure is chosen to assess the fit of the model with regard to particular aspects of the data. Following Gilks (1996, p. 190), let $\mathbf{Y}$ be the observed data and $\theta$ the vector of unknown model parameters. From the Markov Chain simulation we obtained draws $\theta_1, \cdots, \theta_R$ from the posterior. We now simulate R hypothetical replications of the data denoted by $\boldsymbol{Y}_1^{rep}, \cdots, \boldsymbol{Y}_R^{rep}$, where $Y_i^{rep}$ is drawn from the sampling distribution of $\boldsymbol{Y}$ given the simulated parameter $\theta_r$.

The hypothetical replications should look similar to the observed data $\boldsymbol{Y}$, when the model is fitting the data. Since the discrepancy measure $D(Y, \theta)$ will have an extreme value if the data is in conflict with the chosen model, the proportion of cases where the simulated discrepancy measure $D(Y_r^{rep}, \theta_r)$ exceeds the realized value $D(Y, \theta_r)$, estimates the p-value of this Bayesian model test.

If the formulated model provides a good fit to the particular aspect of the data as measured by $D(Y, \theta)$, we expect that half of the points in a scatter plot of $D(\boldsymbol{Y}, \theta_r)$ versus $D(Y_r^{rep}, \theta_r)$ are falling above the $45^o$ line and half falling below, i.e. a estimated p-value of .5 indicates no lack of fit.

For the unemployment data we are primarily concerned about the marginal fit, which can be measured by a $\chi^2$ discrepancy statistic

$$D_{\chi^2}(\boldsymbol{Y}, \boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{(y_{it} - \pi_{it})^2}{\pi_{it}(1 - \pi_{it})}$$

where $\pi_{it} = \Phi(\eta_{it}(\beta))$. If we are also interested in assessing jointly the fit of the regression parameters as well as the association parameter, the Mahalanobis discrepancy measure might be used:

$$D_M(\boldsymbol{Y}, \boldsymbol{\beta}, \rho) = \sum_{i=1}^{N} (y_i - p_i)^t \Sigma_{Y_i}^{-1}(\boldsymbol{\beta}, \rho)(y_i - p_i)$$

where $Y_i = (Y_{i1}, \cdots, Y_{iT})^t$ and $\pi_i = (\pi_{i1}, \cdots, \pi_{iT})^t$. Here $\Sigma_{Y_i}(\beta, \rho)$ denotes the variance covariance matrix of the random vector $Y_i$. In particular we have

$$\Sigma_{Y_i}(\boldsymbol{\beta}, \rho)_{tt} = \pi_{it}(1 - \pi_{it}) \text{ and } \Sigma_{Y_i}(\boldsymbol{\beta}, \rho)_{ts} = P(Y_{it} = 1, Y_{is} = 1) - \pi_{it}\pi_{is}.$$

Table 5.1 gives the estimated p-values for the two discrepancy measures for all models considered. With regard to the $\chi^2$ discrepancy the quadratic time specification(8) is sufficient. The serial correlation structure for the year specific time specification (9) is less appropriate. When one is interested in jointly assessing the fit of the mean specification and correlation structure, the quadratic time specification (8) with exchangeable correlation is preferred over all other models considered.

|  | Quadratic Time Effects |  | Year Specific Time Effects |  |
|---|---|---|---|---|
|  | Exchangeable | Serial | Exchangeable | Serial |
| $D_{\chi^2}(\boldsymbol{Y},\boldsymbol{\beta})$ | .417 | .438 | .413 | .338 |
| $D_M(\boldsymbol{Y},\boldsymbol{\beta},\rho)$ | .446 | .338 | .359 | .237 |

Table 5.1: Estimated p-values of the posterior predictive model checking

To gain more insight into the behavior of these model diagnostics we present the results of a small simulation study. Here 100 data sets were generated according to a multivariate probit model with either serial or exchangeable correlation. As mean specification we used

$$\eta_{it}(\boldsymbol{\beta}) = \beta_0 + \beta_1 x_i + \beta_2 t \text{ for } i = 1, \cdots, 100 \text{ and } t = 1, \cdots, 5.$$

The covariate $x_i$ is chosen to be equally spaced between -1 and 1. 500 iterations of the MCMC algorithm for each data set were run and the reported estimates are based on the last 200 iterations. Table 5.2 presents parameter and p-value estimates together with their standard errors given below.

| True Corr. | Fitted Corr. | $\hat{\rho}$ | $\hat{\alpha}$ | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\beta_2$ | $D_{\chi^2}$ | $D_M$ |
|---|---|---|---|---|---|---|---|---|
| Serial | Serial | 0.7881 | 2.187 | -1.031 | 1.046 | 0.5138 | 0.470 | 0.456 |
|  |  | 0.0055 | 0.030 | 0.018 | 0.019 | 0.0061 | 0.014 | 0.014 |
| Ex. | Serial | 0.8271 | 2.444 | -1.022 | 1.023 | 0.5119 | 0.462 | 0.375 |
|  |  | 0.0055 | 0.037 | 0.018 | 0.021 | 0.0061 | 0.015 | 0.016 |
| Ex. | Ex. | 0.7894 | 2.206 | -1.023 | 1.021 | 0.5123 | 0.483 | 0.452 |
|  |  | 0.0058 | 0.032 | 0.015 | 0.020 | 0.0051 | 0.016 | 0.015 |
| Serial | Ex. | 0.7175 | 1.871 | -1.014 | 1.009 | 0.5036 | 0.503 | 0.550 |
|  |  | 0.0081 | 0.036 | 0.020 | 0.019 | 0.0062 | 0.017 | 0.019 |

Table 5.2: Multivariate probit estimates, standard errors and model diagnostics based on 100 simulations (True parameter values: $\rho = .8, \alpha = \log(\frac{1+\rho}{1-\rho}) = 2.197, \beta_0 = -1, \beta_1 = 1, \beta_2 = .5$)

The correct correlation structure is not essential for the estimation of the regression parameter estimates. This is shown by the unbiasedness of the regression parameter estimates as well as that the estimated p-values corresponding to the $\chi^2$ discrepancy $D_{\chi^2}(\boldsymbol{Y},\boldsymbol{\beta})$ are close to .5 for all models fitted.

With regard to the estimation of the correlation the correct correlation structure of course matters. It is interesting to note that the bias in $\rho$ is larger when the parameters are estimated using an exchangeable correlation structure when the true correlation struction is serial compared to the case where the true and fitted correlation structure are interchanged. This effect is also noticable for the estimated p-values of the Mahalanobis distance $D_M(Y, \boldsymbol{\beta}, \rho)$. Further $D_M(Y, \boldsymbol{\beta}, \rho)$ is futher away from .5 for the incorrectly fitted model. This shows that $D_M(Y, \boldsymbol{\beta}, \rho)$ is effective in assessing the correct correlation structure.

In summary, this shows that the analysis of the multivariate probit model using MCMC methods not only allows for a tractable analysis but also allows for model checking of specified aspects of the data.

In discussing the applicability of the multivariate probit analysis using MCMC, we would like to mention that published applications using the likelihood approaches based on the odds ratio (Fitzmaurice and Laird (1993) and Molenberghs and Lesaffre (1994)) involve at most 4 time points.

Recently some progress has been made for the calculation of rectangle normal probabilities (see (3)) in Hajivassiliou et al. (1996). They compare different simulation methods for calculating these probabilities and conclude that the Geweke-Hajivassiliou-Keane (GHK) simulator performs best (see also Geweke et al. (1995)). The GHK simulator also uses a successive generation scheme as used by Czado (1996). It would be interesting to compare the performance of the MCMC algorithm to the performance of the likelihood analysis using the GHK simulator.

We close by mentioning that for longer binary time series dynamic state space models (see Fahrmeir and Tutz (1994, Chapter 8)) are useful. MCMC methods applied to these models were developed by Carlin et al. (1992). However, Carter and Kohn (1994) and Fruehwirth-Schnatter (1994) observe bad mixing and slow convergence behavior if state parameters are not updated in a single step. Recently, a Metropolis-Hastings algorithm based on conditional prior proposals is suggested by Knorr-Held (1996) exhibiting good mixing and convergence properties. These problems were not encountered for the multivariate probit model.

## Acknowledgements

I would like to thank Mr. Wenqing He for his excellent help in translating and extending the original Splus programs into C programs. The author is also grateful to the referrees and the coordinating editor for valuable suggestions on an earlier draft.

## References

Amemiya,T. (1986). *Advanced Econometrics.* Harvard University Press, Cambridge, Mass.

Anderson, J.A. and Pemberton, J.D. (1985). The grouped continuous model for multivariate ordered categorical variables and covariate adjustment, *Biometrics*, **41**, 875-885.

Ashby, M. Neuhaus J.M., Hauck, W.W., Bacchetti P., Heibron, D.C., Jewell, N.P., Segal, M.R. and Fusaro, R.E. (1992) An Annotated Bibliography of Methods for Analyzing Correlated Categorical Data. *Statistics in Medicine*, **11**, 67-99.

Ashford, J.R. and Sowden, R.R. (1970) Multivariate probit analysis, *Biometrics*, **26**, 535-546.

Baltagi, B. H. (1996). *Econometric Analysis of Panel Data*, John Wiley & Sons, New York.

Besag, J., Green P., Hidgon D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems. *Statistical Science*, **10**, No. 1, 3-66.

Best, N. , Cowles, M.K. and Vines, K. (1995). CODA - Convergence Diagnosis and Output Analysis Software, *MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK, email: bugs@mrc-bsu.cam.ac.uk*

Butler, J.S. and Moffit, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model, *Econometrica*, **50**, 761-764.

Carey, V., Zeger, S.L. and Diggle, P.J. (1993) Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, **80**, 517-526.

Carlin, B.P., Polsen, N.G. and Stoffer, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space-modeling. *J. Am. Statist. Ass.*, **87**, 493-500.

Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, **81**, 541-553.

Chamberlain, G. (1984). Comments on "Adaptive estimation of nonlinear regression models", *Econometric Review*,**3**, 199-202.

le Cessie, S and van Houwelingen, J.C. (1994). Logistic Regression for Correlated Binary Data, *Appl. Statist.*, **43**, No. 1, 95-108.

Cowles, M.K. and Carlin, B.P. (1995) Markov chain Monte Carlo convergence diagnostics: a comparative review, *J. Am. Statist. Ass.*, **91**, 883-904.

Cox, D.R. (1972). The analysis of multivariate binary data, *Appl. Statist.*, **21**, 113-120.

Czado, C. (1996). Multivariate Probit Analysis of Binary Time Series Data with Missing Responses, preprint (http://www-m4.mathematik.tu-muenchen.de/m4/Papers/Czado/cc-pubs.html).

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling based on Generalized Linear Models.* New Yor, Springer-Verlag.

Fitzmaurice, G.M. and Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses, *Biometrika*, **80**, 1, 141-151.

Fitzmaurice, G.H., Laird, N.H. and Rotnitzky, A.G. (1993). Regression Models for Discrete Longitudinal Responses, *Statist. Sci.*, **8**, 284-309.

Fitzmaurice, G.M. and Lipsitz, S.R. (1995). A model for binary time series data with serial odds ratio patterns. *Appl. Statist.*, **44**, No. 1, 51-61.

Fruehwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *J. of Time Series Analysis*, **15**, 183-202.

Gelfand, A.E. and Smith, A.F.M. (1995). *Bayesian Computation*, New York, Wiley, in preparation.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). *Bayesian Data Analysis*, New York, Chapman and Hall.

Gelman, A., Meng, X.-L. and Stern, H.S. (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion), *Statistica Sinica*, **6**, 733–807.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints, *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface, Seattle, Washington, April 21-24, 1991*, 571-578.

Geweke, J., Keane, K. and Runkle, D. (1995). Recursively Simulating Multinomial Multiperiod Probit Probabilities, *American Statistical Association 1994 Proceedings of the Business and Economic Statistics Section.*

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*, New York, Chapman and Hall.

Guttman, I. (1967) The use of the concept of a future observation in goodness-of-fit problems. *J.R. Statist. Soc. B.* , **29**, 83-100.

Hajivassiliou, V., McFadden, D. and Ruud, P. (1996) Simulation of multivariate normal rectangle probabilities and their derivatives. Theoretical and computational results. *J. of Econometrics*, **72**, 85-134.

Hastie, T.J. and Tibsherani, R.J. (1990). *Generalized Additive Models*, Chapman and Hall, New York.

Heagerty, P.J. and Zeger, S.L. (1996). Marginal regression models for clustered ordinal measurements, *J. Amer. Statist. Soc.* , **91**, 1024-1036.

Heckman, J.J. and Borjas, G. (1980). Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence. *Economica*, **47**, 247-283.

Heumann, C. (1996). Marginal regression modeling of correlated multicategorical response: a likelihood approach, Disscusion paper 19, SFB 386, Seminar für Statistics, Ludwig-Maximilians-Universität, München.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge University Press, Cambridge.

Knorr-Held, L. (1996). Conditional Prior Proposals in Dynamic Models, Discussion Paper 36, SFB 386, LMU Muenchen, Seminar für Statistik, (http://www.stat.uni-muenchen.de/sfb386/publikation.html).

Lee, P.M. (1997). *Bayesian Statistics: An Introduction*, Second Edition, John Wiley & Sons, New York.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13-22.

Liang, K.-Y, Zeger, S.L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with discussion). *J.R. Statist. Soc. B*, **54**, 3-40.

Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association, *Biometrika*, **78**, 153-160.

Lipsitz, S.R., Fitzmaurice, G.M., Sleeper, L. and Zhao, L.P. (1995). Estimation methods for the joint distribution of repeated binary observations, *Biometrics*, **51**, 562-570.

Molenberghs, G. and Lesaffre, E. (1994). Marginal Modeling of Correlated Ordinal Data Using a Multivariate Plackett Distribution. *J. Amer. Statist. Soc.* , **89**, No. 426, 633-644.

Müller, P. (1994) A Generic Approach to Posterior Integration and Gibbs Sampling. to appear in *J. Amer. Stat. Assoc.*

Niesing, W., van Praag, B.M.S. and Veenman, J. (1994). The unemployment of ethnic minority groups in the Netherlands. *J. Econometrics*, **61**, 173-196.

Ochi, Y. and Prentice, R.L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika*, **73**, 531-543.

Pendergast, J.F., Gange, S.J., Newton, M.A., Lindstrom, M.J., Palta, M. and Fisher, M.R. (1996) A survey of methods for analyzing clustered binary response data, *Inter. Statist. Rev.* , 89-118.

Plackett, R.L. (1965). A class of bivariate distributions, *J. Amer. Statist. Ass.*, **60**, 516-522.

Qu, Y., Piedmonte, M.R. and Medendorp, S.V. (1995). Regression models for clustered ordinal data. *Biometrics*,**51**, 268-275.

Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J.R. Statist. Soc. B.*, **53**, 233-243.

Robert, C.P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121-125.

Rubin, D.B. (1981) Estimation in parallel randomized experiments. *J. Educ. Statist.*, **6**, 377-401.

Rubin, D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist*,**12**, 1151-1172.

Spiess, M. and Hamerle, A. (1996). On properties of GEE estimators in the presence of invariant covariates. *Biometrical J.*, **38**, 931-940.

Spiess, M., Nagl, W. and Hamerle, A. (1996) Probit models: Regression parameter estimation using the ML principle despite misspecification of the correlation structure, Discussion Paper 67, SFB 386,
(http://www.stat.uni-muenchen.de/sfb386/publikation.html).

Zhao, L.P. and Prentice, R.L. (1990). Correlated binary regression using a quadratic exponential model, *Biometrika*, **77**, 642-648.