Table 6.3 Bayesian estimates of parameters and their standard error estimates for a three-factor model: ICPSR data set.

Par	EST	SE	Par	EST	SE	Par	EST	SE
λ_{11}	0.859	0.075	λ_{62}	0.086	0.043	$\psi_{\epsilon 4}$	0.976	0.088
λ_{22}	−0.304	0.130	λ_{63}	−0.271	0.043	$\psi_{\epsilon 5}$	0.616	0.061
λ_{31}	1.507	0.099	λ_{71}	0.269	0.090	$\psi_{\epsilon 6}$	0.915	0.057
λ_{32}	−0.443	0.144	λ_{73}	0.098	0.058	$\psi_{\epsilon 7}$	0.366	0.050
λ_{33}	0.344	0.087	λ_{81}	0.077	0.049	$\psi_{\epsilon 8}$	0.858	0.051
λ_{41}	0.122	0.051	λ_{82}	0.382	0.044	$\psi_{\epsilon 9}$	0.535	0.076
λ_{42}	0.016	0.056	λ_{83}	0.055	0.046	$\alpha_{6,2}$	0.328	0.034
λ_{43}	0.230	0.060	λ_{91}	−0.104	0.052	$\alpha_{7,2}$	1.116	0.052
λ_{51}	−0.487	0.059	λ_{92}	0.036	0.073	$\alpha_{7,3}$	1.657	0.062
λ_{52}	0.316	0.078	$\psi_{\epsilon 1}$	4.190	0.195	$\alpha_{8,2}$	−0.107	0.034
λ_{53}	−0.111	0.059	$\psi_{\epsilon 2}$	1.798	0.115	$\alpha_{8,3}$	0.472	0.036
λ_{61}	−0.048	0.039	$\psi_{\epsilon 3}$	2.604	0.224	$\alpha_{9,2}$	−0.782	0.046
						$\alpha_{9,3}$	0.162	0.046

6.6 APPLICATION 2: BAYESIAN ANALYSIS OF QUALITY OF LIFE DATA

There is increasing recognition that measures of quality of life (QOL) and/or health-related QOL have great value for clinical work and the planning and evaluation of health care as well as for medical research. It has been generally accepted that QOL is a multidimensional concept (Staquet, Hayes and Fayes, 1998) that is best evaluated by a number of different latent constructs such as physical function, health status, mental status and social relationships. As these latent constructs often cannot be measured objectively and directly, they are treated as latent variables in QOL analysis. The most popular method that is used to assess a latent construct is by a survey which incorporates a number of related items that are intended to reflect the underlying latent construct of interest.

EFA has been used as a method for exploring the structure of a new QOL instrument (The WHOQOL Group, 1998; Fayer and Machin, 1998), while CFA has been used to confirm the factor structure of the instrument. SEMs that are based on continuous observations with a normal distribution have also been applied to QOL analyses (Power, Bullingen and Hazper, 1999).

Items in a QOL instrument are usually measured on an ordered categorical scale, typically with three- to five-points. The discrete ordinal nature of the items also attracts much attention in QOL analysis (Fayer and Machin, 1998; Fayer and Hand, 1997). It has been pointed out that non-rigorous treatments of the ordinal items as continuous can be subjected to criticism (Glonek and McCullagh, 1995), and models such as the item response model and ordinal regression that take into account the ordinal nature are more appropriate (Olschewski and Schumacker, 1990). The aim of this section is to apply the Bayesian methods for analyzing a common QOL instrument with ordered categorical items.

6.6.1 A Synthetic Illustrative Example

This instrument WHOQOL-100 (Power, Bullingen and Hazper, 1999) was established to evaluate four latent constructs. The first seven items (Q3 to Q9) are intended to address physical health, the next six items (Q10 to Q15) are intended to address psychological health, the three items (Q16, Q17, Q18) that follow are for social relationships, and the last eight items (Q19 to Q26) are intended to address environment. In addition to the 24 ordered categorical items, the instrument also includes two ordered categorical items for the overall QOL (Q1) and general health (Q2), giving a total of 26 items. All of the items are measured with a five-point scale (1 = 'not at all/very dissatisfied'; 2 = 'a little/dissatisfied'; 3 = 'moderate/neither'; 4 = 'very much/satisfied'; 5 = 'extremely/very satisfied'). The sample size of the whole data set is extremely large. To illustrate the Bayesian methods, we only analyze a synthetic data set with sample size $n = 338$. The frequencies of all the ordered categorical items are presented in Table 6.4. As can be seen from the table, many items are skewed to the right. Treating these ordered categorical data as coming from normal is not correct. Hence, our Bayesian approach that takes into account the discrete nature of the data is applied to analyze this ordered categorical data set.

To illustrate the path sampling procedure, we compare an SEM with four exogenous latent variables with another SEM with three exogenous latent variables, see Lee, Song, Skevington and Hao (2005). Let M_1 be the SEM whose measurement equation is defined by

$$y_i = \mathbf{\Lambda}_1 \boldsymbol{\omega}_{1i} + \boldsymbol{\epsilon}_i, \tag{6.25}$$

where $\boldsymbol{\omega}_{1i} = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3}, \xi_{i4})^T$, $\boldsymbol{\epsilon}_i$ is distributed according to $N[0, \mathbf{\Psi}_{\epsilon 1}]$, and

$$\mathbf{\Lambda}_1^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \lambda_{42} & \cdots & \lambda_{92} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{11,3} & \cdots & \lambda_{15,3} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{17,4} & \lambda_{18,4} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & \lambda_{20,5} & \cdots & \lambda_{26,5} \end{bmatrix}.$$

The structural equation of M_1 is defined by

$$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \gamma_4 \xi_4 + \delta, \tag{6.26}$$

where the distributions of $(\xi_1, \xi_2, \xi_3, \xi_4)^T$ and δ are independently distributed as $N[\mathbf{0}, \mathbf{\Phi}_1]$, and $N[0, \sigma_{1\delta}^2]$, respectively. Let M_2 be the SEM whose measurement is defined by

$$y_i = \mathbf{\Lambda}_2 \boldsymbol{\omega}_{2i} + \boldsymbol{\epsilon}_i, \tag{6.27}$$

Table 6.4 Frequencies of the questions in the WHOQOL data set.

	1	2	3	4	5
Q1 Overall QOL	2	34	75	160	67
Q2 Overall health	25	89	71	117	36
Q3 Pain and discomfort	16	49	78	111	84
Q4 Medical treatment dependence	17	48	65	83	125
Q5 Energy and fatigue	16	53	107	86	76
Q6 Mobility	13	33	62	95	135
Q7 Sleep and rest	23	62	73	116	64
Q8 Daily activities	9	55	63	158	53
Q9 Work capacity	19	71	79	116	53
Q10 Positive feeling	8	22	93	165	50
Q11 Spirituality/personal beliefs	8	29	99	137	65
Q12 Memory and concentration	4	22	148	133	31
Q13 Bodily image and appearance	3	30	106	112	87
Q14 Self-esteem	7	38	104	148	41
Q15 Negative feeling	4	35	89	171	39
Q16 Personal relationship	5	16	59	165	93
Q17 Sexual activity	25	48	112	100	53
Q18 Social support	7	6	73	164	88
Q19 Physical safety and security	4	20	147	129	38
Q20 Physical environment	7	20	142	126	43
Q21 Financial resources	15	34	140	87	62
Q22 Daily life information	4	22	102	154	56
Q23 Participation in leisure activity	15	76	102	108	37
Q24 Living condition	4	12	35	173	114
Q25 Health accessibility and quality	4	20	59	205	50
Q26 Transportation	5	16	43	188	86
Total	269	960	2326	3507	1726

where $\boldsymbol{\omega}_{2i} = (\eta_i, \xi_{i1}, \xi_{i2}, \xi_{i3})^T$. $\boldsymbol{\epsilon}_i$ is distributed according to $N[0, \boldsymbol{\Psi}_{\epsilon 2}]$, and

$$\boldsymbol{\Lambda}_2^T = \begin{bmatrix} 1 & \lambda_{21} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \lambda_{42} & \cdots & \lambda_{92} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{11,3} & \cdots & \lambda_{15,3} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \lambda_{17,4} & \cdots & \lambda_{26,4} \end{bmatrix}.$$

The structural equation of M_2 is defined by

$$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \gamma_3 \xi_3 + \delta, \tag{6.28}$$

where the distributions of $(\xi_1, \xi_2, \xi_3)^T$ and δ are independently distributed as $N[\mathbf{0}, \boldsymbol{\Phi}_2]$, and $N[0, \sigma_{2\delta}^2]$, respectively. Bayesian analyses are conducted using the conjugate prior distributions. The hyperparameter values corresponding to the prior distributions of the unknown loadings in $\mathbf{\Lambda}_1$ or $\mathbf{\Lambda}_2$ are all taken to be 0.8; those corresponding to $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ are {0.6, 0.6, 0.4, 0.4}; those corresponding to $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are $\rho_0 = 30$, $\boldsymbol{R}_0^{-1} = 8\mathbf{I}_4$ and $\boldsymbol{R}_0^{-1} = 8\mathbf{I}_3$, respectively; $\mathbf{H}_{0yk} = 0.25\mathbf{I}_{26}$, $\mathbf{H}_{0\omega k} = 0.25\mathbf{I}_4$, or $\mathbf{H}_{0\omega k} = 0.25\mathbf{I}_3$, $\alpha_{0\epsilon k} = \alpha_{0\delta k} = 10$, and $\beta_{0\epsilon k} = \beta_{0\delta k} = 8$. In the path sampling procedure in computing the Bayes factor, we take $S = 10$ and $J = 2000$ after a 'burn-in' phase of 1000 iterations.

We first compare M_1 with the following simple model M_0:

$$M_0 : \mathbf{y}_i = \boldsymbol{\epsilon}_i$$

where $\boldsymbol{\epsilon}_i \overset{D}{=} N[\mathbf{0}, \mathbf{\Psi}_\epsilon]$ and $\mathbf{\Psi}_\epsilon$ is a diagonal matrix. The measurement equation of the linked model is defined by $M_t : \mathbf{y}_i = t\mathbf{\Lambda}_1\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i$. We obtain $\widehat{\log B_{10}} = 81.05$. Clearly, M_1 is better than M_0. Similarly, M_2 and M_0 can be compared via the path sampling procedure. We find that $\widehat{\log B_{20}} = 57.65$, which suggests that M_2 is better than M_0. From the above results, we can obtain an estimate of $\log B_{12}$, which is equal to 23.40. Hence, M_1, the SEM with one endogenous and four exogenous latent variables is selected. Bayesian estimates of the unknown structural parameters in M_1 are presented in Figure 6.4. The less interesting threshold estimates are not presented. All the factor loading estimates, except $\hat{\lambda}_{17,4}$ that associates with the indicator 'sexual activity', are high. This indicates a strong association between each of the latent variables and their corresponding indicators. From the meaning of the items, η, ξ_1, ξ_2, ξ_3 and ξ_4 can be interpreted as the overall QOL, physical health, psychological health, social relationship and environment, respectively. The estimates of correlations among the exogenous latent variables are equal to $\{0.680, 0.432, 0.632, 0.685, 0.736, 0.698\}$. As expected, these correlations are high. The estimated structural equation that addresses the relations of QOL with the latent constructs about physical and psychological health, social relationship and environment is

$$\eta = 0.724\xi_1 + 0.319\xi_2 + 0.169\xi_3 - 0.041\xi_4.$$

Hence, physical health has the most important effect on QOL, followed in turn by psychological health and social relationship, while the effect of environment is not important.

6.6.2 Application of WinBUGS

For SEMs with ordered categorical variables, the software WinBUGS (Spiegelhalter, Thomas, Best and Lunn, 2003) can produce Bayesian estimates of the structural parameters and latent variable estimates in the model, as well

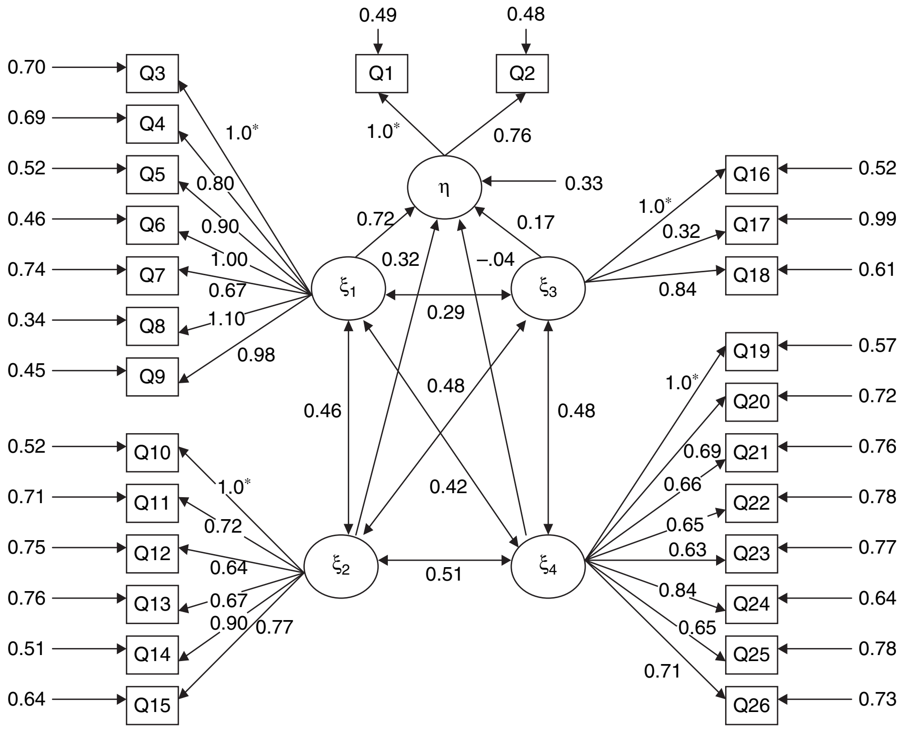

Figure 6.4 Path diagram and Bayesian estimates of parameters in the analysis of the QOL data. Note that Bayesian estimates of ϕ_{11}, ϕ_{22}, ϕ_{33} and ϕ_{44} are 0.648, 0.706, 0.694 and 0.680, respectively.

as their standard error estimates, by means of a sufficiently large number of observations simulated by MCMC methods. In our Bayesian treatment of the ordered categorical variables, we fix the thresholds at both ends in order to solve the identification problem, then the other unknown thresholds are simultaneously estimated with the structural parameters $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}_\epsilon$, $\boldsymbol{\Phi}$, $\boldsymbol{\Lambda}_\omega$ and $\boldsymbol{\Psi}_\delta$ in the model. However, according to our understanding of WinBUGS, it is not straightforward to apply this software to estimate the unknown thresholds and structural parameters simultaneously. Hence, in applying WinBUGS, we first estimate all the thresholds through the method as described in Section 6.2, using the observed frequencies and the distribution function of $N[0, 1]$. Then, the thresholds are fixed in the WinBUGS program in producing the Bayesian solutions. Note that this procedure may underestimate the standard error estimates. Hence, hypothesis testing should be conducted through DIC, rather than the z-score that depends on the standard error estimate.

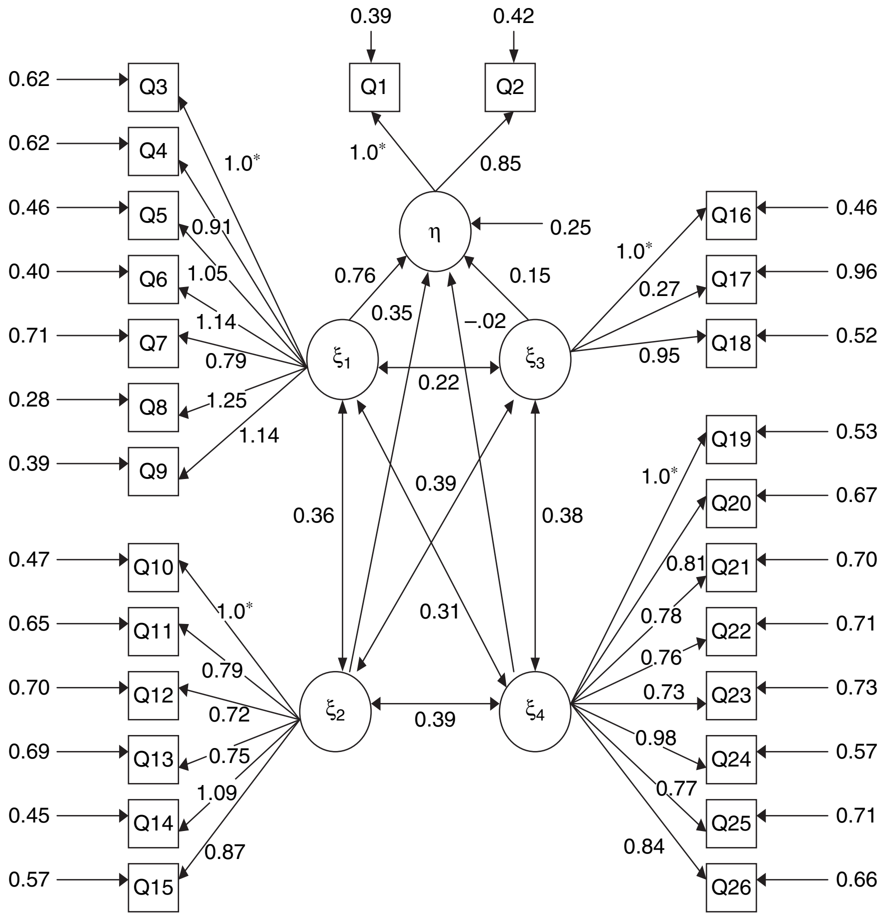

Figure 6.5 Path diagram and Bayesian estimates of parameters in the analysis of the QOL data obtained via WinBUGS. Note that Bayesian estimates of ϕ_{11}, ϕ_{22}, ϕ_{33} and ϕ_{44} are 0.493, 0.584, 0.600 and 0.530, respectively.

We apply WinBUGS to the synthetic QOL data set as discussed in the previous subsection. Bayesian solutions are obtained under the selected model M_1 with the same conjugate prior distributions and hyperparameter inputs. Note that it is desirable to give initial values for $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$ in order to save computer time. The DIC value corresponding to this model is 19 537.9. Bayesian estimates of the unknown structural parameters are presented in Figure 6.5. We observe that the estimates respectively presented in Figures 6.4 and 6.5 are not as close as expected. The possible reason may be that the treatments of the thresholds by the two programs are different. The completely

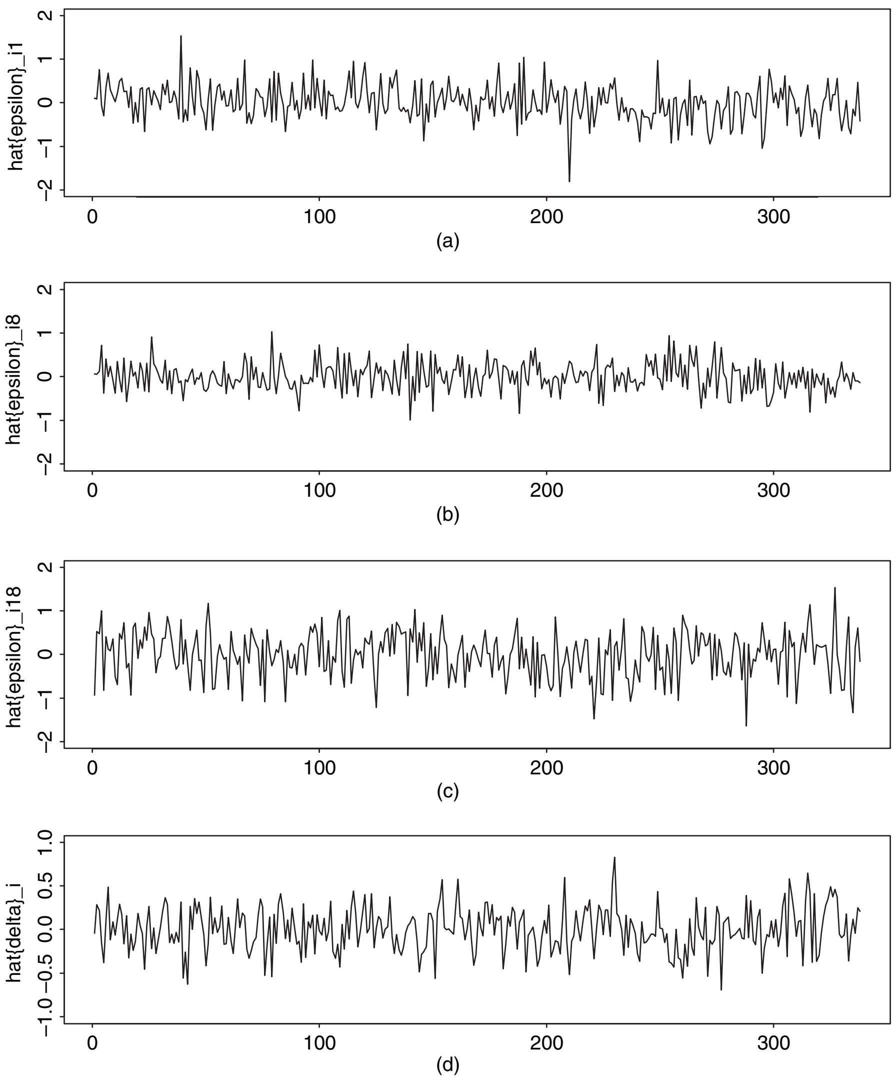

Figure 6.6 Estimated residual plots, (a) $\hat{\epsilon}_{i1}$, (b) $\hat{\epsilon}_{i8}$, (c) $\hat{\epsilon}_{i18}$ and (d) $\hat{\delta}_i$.

standardized solutions corresponding to these two sets of estimates are closer, for example, the estimates of the correlations among the exogenous latent variables are equal to $\{0.664, 0.406, 0.610, 0.661, 0.705, 0.679\}$. Some estimated residual plots, $\hat{\epsilon}_{i1}$, $\hat{\epsilon}_{i8}$, $\hat{\epsilon}_{i18}$ and $\hat{\delta}_i$ versus the case number are displayed in Figure 6.6, while plots of estimated residual $\hat{\epsilon}_{i1}$ and $\hat{\delta}_i$ versus $\hat{\xi}_{i1}$, $\hat{\xi}_{i2}$, $\hat{\xi}_{i3}$, $\hat{\xi}_{i4}$ and $\hat{\eta}_i$ are presented in Figures 6.7 and 6.8 respectively. Other estimated

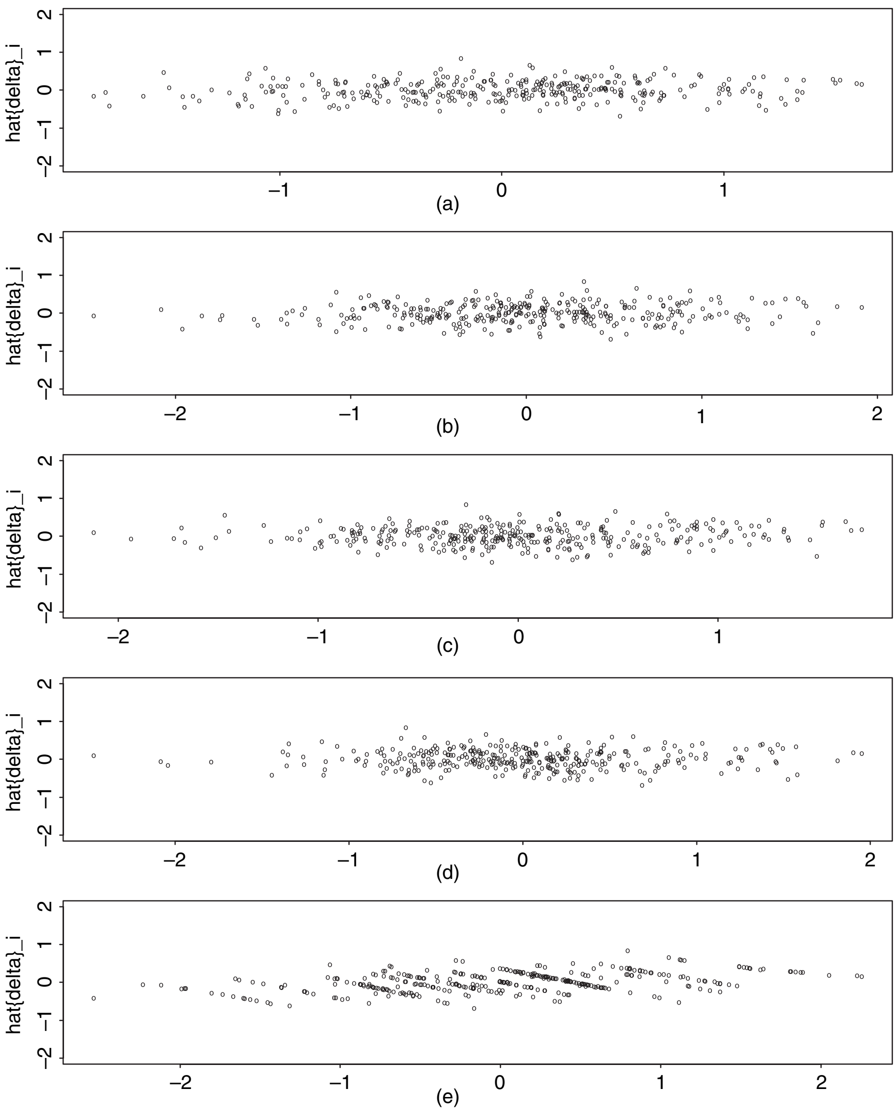

Figure 6.7 Plots of estimated residuals $\hat{\epsilon}_{i1}$ versus (a) $\hat{\xi}_{i1}$, (b) $\hat{\xi}_{i2}$, (c) $\hat{\xi}_{i3}$, (d) $\hat{\xi}_{i4}$ and (e) $\hat{\eta}_i$.

residual plots are similar. These estimated residual plots lie within two parallel horizontal lines that are centered at zero, and no linear or quadratic trends are detected. This indicates that the proposed measurement and structural equations are adequate. The WinBUGS codes and the data are given in the following website: http://www.wiley.com/go/lee_structural.

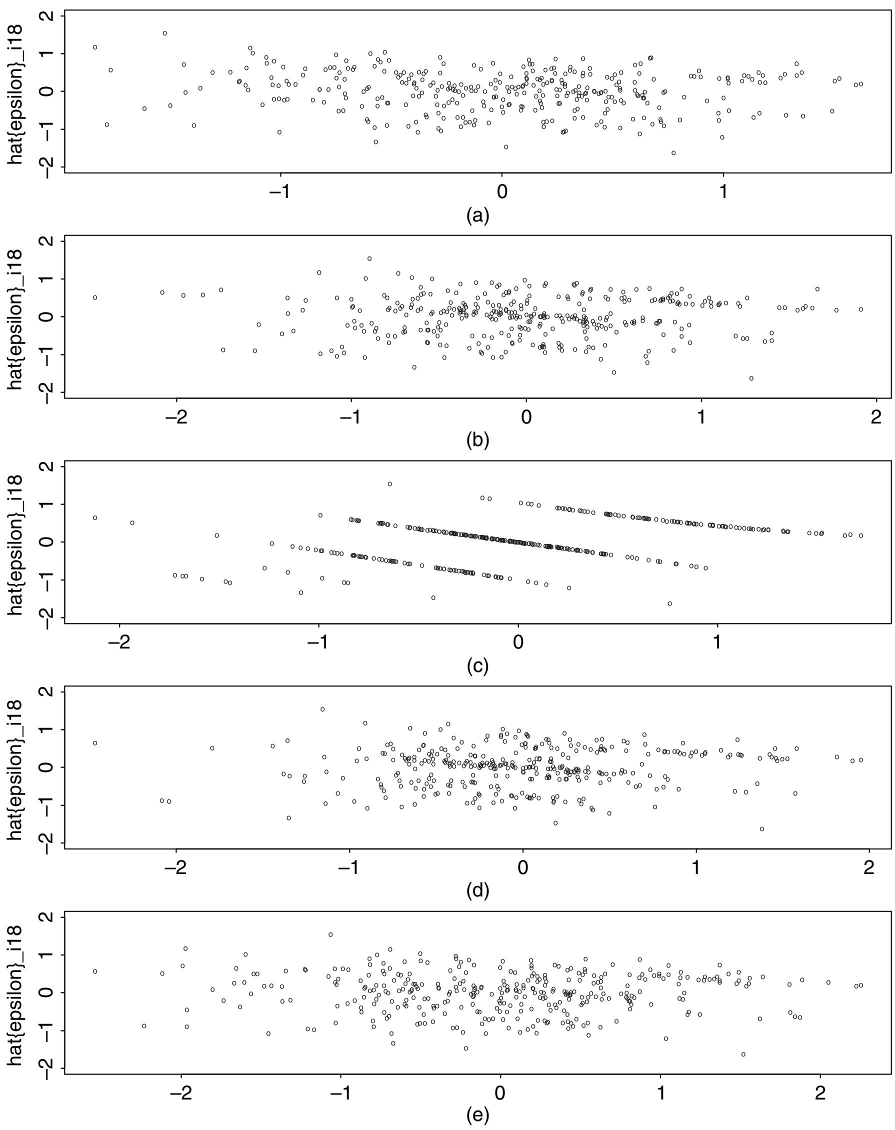

Figure 6.8 Plots of estimated residuals $\hat{\delta}_i$ versus (a) $\hat{\xi}_{i1}$, (b) $\hat{\xi}_{i2}$, (c) $\hat{\xi}_{i3}$, (d) $\hat{\xi}_{i4}$ and (e) $\hat{\eta}_i$.

REFERENCES

Akaike, H (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics*, **22**, 203–217.

Akaike, H. (1987) Factor analysis and AIC. *Psychometrika*, **52**, 317–332.