Supplementary Appendix to "Nonparametric Identification and Estimation of the Number of Components in Multivariate Mixtures"

Hiroyuki Kasahara
Department of Economics
University of British Columbia
hkasahar@mail.ubc.ca

Katsumi Shimotsu
Faculty of Economics
University of Tokyo
shimotsu@e.u-tokyo.ac.jp

January 10, 2013

This supplementary appendix contains the following details omitted from the main paper due to space constraints: (A) proof of the results in the paper, (B) the asymptotic variance of a multivariate contingency table when its categories are defined by the empirical quantiles of the marginal data (from Borkowf (2000)), (C) tables of simulation results, and (D) additional results from empirical examples.

A Proofs of the results in the main text

A.1 Proof of Proposition 2

First, note that $M_0 = \text{rank}_+(P(\theta))$ holds from the definition of Θ. Define P_x, P_y, and V as $P_x = [p_x^1, \ldots, p_x^{M_0}]'$, $P_y = [p_y^1, \ldots, p_y^{M_0}]'$, and $V = \text{diag}(\pi^1, \ldots, \pi^{M_0})$, respectively, where P_x is $M_0 \times |\Delta_x|$ and P_y is $M_0 \times |\Delta_y|$. Then, $P(\theta)$ is written as $P(\theta) = P_x' V P_y$. Applying the Frobenius inequality to the right-hand side of $\text{rank}(P(\theta)) = \text{rank}(P_x' V P_y)$ and noting that $\text{rank}(P_x' V) = \text{rank}(P_x')$, $\text{rank}(V P_y) = \text{rank}(P_y)$, and $\text{rank}(V) = \text{rank}_+(P(\theta))$, we obtain $\text{rank}(P(\theta)) \geq \text{rank}(P_x) + \text{rank}(P_y) - \text{rank}_+(P(\theta))$. Suppose that $\text{rank}_+(P(\theta)) > \text{rank}(P(\theta))$. Then, we have $2\text{rank}_+(P(\theta)) > \text{rank}(P_x) + \text{rank}(P_y)$, and thus, either $\text{rank}(P_x)$ or $\text{rank}(P_y)$ must be strictly smaller than $\text{rank}_+(P(\theta))$. Because $\text{rank}_+(P(\theta)) \leq \min\{|\Delta_x|, |\Delta_y|\}$, either P_x or P_y does not have full rank. Note that the elements of a rank-deficient matrix must satisfy a set of polynomial restrictions, and hence, must lie in a zero set of a finite collection of polynomials. Therefore, in the space of $M_0 \times |\Delta_x|$ matrices that represents the space of $\{p_x^m\}_{m=1}^{M_0}$, the set of P_x's that do not have full rank has zero Lebesgue measure (see Allman et al. 2009, p. 3105), and a similar argument holds for P_y. This proves the stated result. $\square$

A.2 Proof of Proposition 3

From Lemma 17 of Allman et al. (2009), there exists a positive integer κ and real numbers $x_1 < x_2 < \cdots < x_{\kappa-1}$ such that the vectors $\{(F_x^m(x_1), \ldots, F_x^m(x_{\kappa-1}), 1)\}_{1 \le m \le M}$ are linearly independent. Therefore, it is possible to construct Δ such that $P_x = [p_x^1, \ldots, p_x^M]'$ has rank M. Similarly, it is possible to construct Δ such that $P_y = [p_y^1, \ldots, p_y^M]'$ has rank M. Write P_Δ as $P_\Delta = P_x' V P_y$, where $V = \text{diag}(\pi^1, \ldots, \pi^M)$. Applying the Frobenius inequality to the right-hand side of $\text{rank}(P_\Delta) = \text{rank}(P_x' V P_y)$ and noting that $\text{rank}(P_x' V) = \text{rank}(P_x')$, $\text{rank}(V P_y) = \text{rank}(P_y)$, and $\text{rank}(V) = M$, we obtain $\text{rank}(P_\Delta) \ge M$. The stated result of part (a) then follows because $\text{rank}(P_\Delta) \le M$. Part (b) follows straightforwardly from Proposition 2. □

A.3 Proof of Proposition 4

As in the proof of Theorem 4 of Allman et al. (2009; p. 3119), given an $n \times a_1$ matrix A_1 and an $n \times a_2$ matrix A_2, define an $n \times a_1 a_2$ matrix $A = A_1 \otimes^{\text{row}} A_2$ as the row-wise tensor product, so that $A(i, a_2(j-1)+k) = A_1(i,j) A_2(i,k)$. Let $P_j = [p_j^1, \ldots, p_j^{M_0}]'$ be an $(M_0 \times |\Delta_j|)$ matrix collecting the distribution of W_j on Δ_j across all the components. Define $P_{x^\alpha} = \otimes_{j \in S_x(\alpha)}^{\text{row}} P_j$, and then, from Lemma 12 of Allman et al. (2009), we have that P_{x^α} collects the distribution of X^α on Δ_{x^α} across all the components. Define P_{y^α} similarly, and then, $P^\alpha(\theta)$ may be written as $P^\alpha(\theta) = P_{x^\alpha}' V P_{y^\alpha}$, where $V = \text{diag}(\pi^1, \ldots, \pi^{M_0})$. From Lemma 13 of Allman et al. (2009), we have $\text{rank}(P_{x^\alpha}) = \min\{M_0, |\Delta_{x^\alpha}|\}$ and $\text{rank}(P_{y^\alpha}) = \min\{M_0, |\Delta_{y^\alpha}|\}$ for generic P_j's; that is, all the P_j's except for a set of Lebesgue measure zero. Because $|\Delta_{x^\alpha}|, |\Delta_{y^\alpha}| \ge M_0$ by assumption, we have $\text{rank}(P_{x^\alpha}) = \text{rank}(P_{y^\alpha}) = M_0$ for generic P_j's. Therefore, $\text{rank}(P^\alpha(\theta)) = M_0$ holds for generic P_j's, from the Frobenius inequality and proceeding as in the proof of Proposition 2. □

A.4 Proof of Proposition 6

First, we show that $\Pr(\tilde{r} < r_0) \to 0$. If $\tilde{r} < r_0$, $Q(r) < Q(r_0)$ for some $r < r_0$. Thus, $\Pr(\tilde{r} < r_0) \le \sum_{r=1}^{r_0-1} \Pr(Q(r) < Q(r_0))$. Observe that $\Pr(Q(r) < Q(r_0)) = \Pr(\text{rk}(r) - \text{rk}(r_0) - f(N)g(r) + f(N)g(r_0) < 0) = \Pr(N\hat{\lambda}_r'\hat{\Omega}_r^{-1}\hat{\lambda}_r - N\hat{\lambda}_{r_0}'\hat{\Omega}_{r_0}^{-1}\hat{\lambda}_{r_0} + f(N)(g(r_0) - g(r)) < 0)$. For any $r < r_0$, this probability tends to 0 as $N \to \infty$ because $f(N)/N \to 0$, $\hat{\lambda}_r'\hat{\Omega}_r^{-1}\hat{\lambda}_r \to_p \lambda_r'\Omega_r^{-1}\lambda_r > 0$, and $\hat{\lambda}_{r_0}'\hat{\Omega}_{r_0}^{-1}\hat{\lambda}_{r_0} \to_p \lambda_{r_0}'\Omega_{r_0}^{-1}\lambda_{r_0} = 0$.

Second, we show that $\Pr(\tilde{r} > r_0) \to 0$. As above, we have $\Pr(\tilde{r} > r_0) \le \sum_{r=r_0+1}^{t^*} \Pr(Q(r) < Q(r_0))$ and $\Pr(Q(r) < Q(r_0)) = \Pr(N\hat{\lambda}_r'\hat{\Omega}_r^{-1}\hat{\lambda}_r - N\hat{\lambda}_{r_0}'\hat{\Omega}_{r_0}^{-1}\hat{\lambda}_{r_0} + f(N)(g(r_0) - g(r)) < 0)$. For any $r > r_0$, this probability tends to 0 as $N \to \infty$ because both $N\hat{\lambda}_r'\hat{\Omega}_r^{-1}\hat{\lambda}_r$ and $N\hat{\lambda}_{r_0}'\hat{\Omega}_{r_0}^{-1}\hat{\lambda}_{r_0}$ converge to a chi-square random variable, $f(N) \to \infty$, and $\Pr(g(r_0) - g(r) > 0) \to 1$ as $N \to \infty$. □

B Asymptotic variance of vec($\hat{P}_\Delta$) when the empirical quantiles are used for partitioning

In simulations in Section 4, we use the empirical quantiles of the marginal data to construct the partitions in forming $\hat{P}_\Delta$. When the categories of a multivariate contingency table are defined by the empirical quantiles of the marginal data, the resulting contingency table does not follow a multinomial distribution, but a distribution called the empirical multivariate quantile-partitioned (EMQP) distribution. Borkowf (2000) derives the asymptotic variance of the EMQP distribution, which we summarize in the following. See Borkowf (2000) for further details.

Let $W = (W_1, \ldots, W_k)'$ be a k-dimensional random variable with the distribution function $F(w)$. Let d_j be the number of categories in the j-th dimension, and define an ordered set of population cumulative marginal proportions in the j-th dimension as $\{\gamma_j(i)\}_{i=0}^{d_j}$ with $0 = \gamma_j(0) < \gamma_j(1) < \cdots < \gamma_j(d_j) = 1$. Let $F_j(w_j)$ denote the marginal distribution in the j-th dimension, and let $\xi_j(i) = F_j^{-1}(\gamma_j(i))$ denote the population quantile for the i-th category of the j-th dimension. For a vector of indices $a = (a_1, \ldots, a_k)'$ of categories with $1 \le a_j \le d_j$, let $\xi(a) = (\xi_1(a_1), \ldots, \xi_k(a_k))'$ be the $k \times 1$ vector of population quantiles, and let $\phi(a) = F(\xi(a))$ denote a cumulative joint proportion.

Given the iid observations of $W_i = (W_{i1}, \ldots, W_{ik})'$ for $i = 1, \ldots, n$ from $F(w)$, define the joint and marginal empirical distribution functions as $\hat{F}(w) = n^{-1} \sum_{i=1}^n \prod_{j=1}^k I\{W_{ij} \le w_j\}$ and $\hat{F}_j(w_j) = n^{-1} \sum_{i=1}^n I\{W_{ij} \le w_j\}$. Let $u_j(i) = \inf\{u : \gamma_j(i) \le \hat{F}_j(u)\}$ denote the empirical quantile for the i-th category of the j-th dimension, and let $u(a) = (u_1(a_1), \ldots, u_k(a_k))'$ be a vector of empirical quantiles. In our simulations, we estimate P_Δ by the proportion of observations in each cell of the partition Δ, which is constructed using the empirical quantiles. Let $\hat{p}(a)$ be the proportion of observations in the a-th cell, then $\hat{p}(a)$ can be derived from $\hat{F}(a)$ by the relation (Borkowf (2000), equation (2.1))

$$\hat{p}(a) = \sum_{b_1=a_1-1}^{a_1} \cdots \sum_{b_k=a_k-1}^{a_k} \left[\prod_{j=1}^{k} (-1)^{a_j - b_j} \right] \hat{F}(u(b)). \tag{12}$$

Borkowf (2000, equation (3.10)) shows that the asymptotic covariance between $n^{1/2}\hat{F}(u(a))$ and $n^{1/2}\hat{F}(u(b))$ is given by

$$\lambda(a)'\Omega(a,b)\lambda(b), \quad \lambda(a) = [-\eta(a)' \; 1]', \tag{13}$$

where $\eta(a) = (\eta_1(a), \ldots, \eta_k(a))'$ is a $k \times 1$ vector of conditional proportions with $\eta_j(a) = \Pr[\bigcap_{\ell=1,\ell\neq j}(W_\ell \le \xi_\ell(a_\ell)) | W_j = \xi_j(a_j)]$ denoting the probability that $W_\ell \le \xi_\ell(a_\ell)$ for all $\ell \neq j$ given that $W_j = \xi_j(a_j)$. For $\Omega(a,b) = \{\{\omega_{j\ell}\}_{\ell=1}^{k+1}\}_{j=1}^{k+1}$, its first k diagonal elements are $\omega_{jj} = \gamma_j(\min\{a_j, b_j\}) - \gamma_j(a_j)\gamma_j(b_j)$. The (j,ℓ)-th off-diagonal element in the upper left $k \times k$

submatrix is $\omega_{j\ell} = \phi(m^a) - \gamma_j(a_j)\gamma_\ell(b_\ell)$, where m^a is a $k \times 1$ vector with $m^a_j = a_j$, $m^a_\ell = b_\ell$, and $m^a_i = d_i$ for $i \neq \{j, \ell\}$. The first k elements in the last row are $\omega_{(k+1)j} = \phi(m^b) - \phi(a)\gamma_j(b_j)$, where $m^b_j = \min\{a_j, b_j\}$ and $m^b_i = a_i$ for $i \neq j$, and the elements in the last column are similar, but with the indices transposed. Finally, the lower right cell is $\omega_{(k+1)(k+1)} = \phi(m^c) - \phi(a)\phi(b)$, where $m^c_j = \min\{a_j, b_j\}$.

The asymptotic covariance of $\hat{p}(a)$ and $\hat{p}(b)$ is derived from the asymptotic covariance between $n^{1/2}\hat{F}(u(a))$ and $n^{1/2}\hat{F}(u(b))$, and relation (12). To estimate $\lambda(a)'\Omega(a, b)\lambda(b)$ in (13), one needs to estimate the conditional distribution function $\eta_j(a) = \Pr[\bigcap_{\ell=1,\ell\neq j}(W_\ell \leq \xi_\ell(a_\ell))|W_j = \xi_j(a_j)]$ for $j = 1, \ldots, k$ nonparametrically by a kernel estimator. Alternatively, one may estimate the asymptotic variance of the EMQP distribution by bootstrap.

C Additional simulation results

Tables 3, 4, and 5 report the simulation results from 2-variable, 3-component normal, chi-squared, and gamma mixtures, respectively. Table 6 reports the simulation result from 4-variable, 3-component normal mixture while Tables 7-8 report the simulation results from 8-variable, 3-component normal mixtures. Figure 1 is based on Tables 3-5 while Figure 2 is based on Tables 6-7.

We also consider an 8-variable, 3-component normal mixture in which the distribution of $W_1, \ldots, W_5$ is the same as in the previous experiment, but W_6, W_7, and W_8 are set to have identical distributions across sub-populations with $\mu^1 = (0, 0, 0, 0, 0, 0, 0, 0)'$, $\mu^2 = (1.0, 2.0, 0.5, 1.0, 0.75, 0, 0, 0)'$, and $\mu^3 = (2.0, 1.0, 1.0, 0.5, 1.25, 0, 0, 0)'$. This is a challenging setup because only five out of eight variables can be used to identify the number of components. As shown in Table 8, the performance of our procedures in this experiment is generally worse than the performance in the experiment in Figure 2(b) but the max-rk$^+$ statistic successfully chooses the correct M at $N = 8000$.

Table 3: Selection Frequencies of the Number of Components for Normal Mixtures: Two Variables

Selection frequencies by the rk statistic using (X, Y) from normal mixtures												
	Partition 1: $t = 4$ with $(q_1, q_2, q_3) = (0.25, 0.5, 0.75)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.021	0.891	0.082	0.006	0.000	0.566	0.414	0.020	0.000	0.008	0.950	0.042
AIC	0.004	0.757	0.215	0.024	0.000	0.317	0.609	0.074	0.000	0.001	0.866	0.133
BIC	0.464	0.533	0.003	0.000	0.000	0.989	0.011	0.000	0.000	0.456	0.543	0.001
HQ	0.092	0.876	0.031	0.001	0.000	0.766	0.226	0.008	0.000	0.043	0.921	0.036
d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
	Partition 2: $t = 4$ with $(q_1, q_2, q_3) = (0.1, 0.5, 0.9)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.024	0.778	0.183	0.015	0.000	0.454	0.512	0.034	0.000	0.003	0.955	0.042
AIC	0.008	0.654	0.287	0.051	0.000	0.246	0.639	0.115	0.000	0.000	0.874	0.126
BIC	0.385	0.593	0.022	0.000	0.000	0.939	0.061	0.000	0.000	0.305	0.692	0.003
HQ	0.088	0.795	0.113	0.004	0.000	0.629	0.353	0.018	0.000	0.020	0.949	0.031
d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
	Partition 3: $t = 6$ with $(q_1, q_2, q_3, q_4, q_5) = (0.1, 0.25, 0.5, 0.75, 0.9)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.027	0.830	0.135	0.008	0.000	0.493	0.483	0.024	0.000	0.000	0.957	0.043
AIC	0.004	0.695	0.297	0.004	0.000	0.281	0.701	0.018	0.000	0.000	0.971	0.029
BIC	0.521	0.472	0.007	0.000	0.000	0.997	0.003	0.000	0.000	0.708	0.292	0.000
HQ	0.133	0.821	0.045	0.001	0.000	0.836	0.164	0.000	0.000	0.034	0.966	0.000
d.f.	25.000	16.000	9.000	0.000	25.000	16.000	9.000	0.000	25.000	16.000	9.000	0.000

Notes: The true number of components is $M = 3$. $(X, Y)'$ follows a 3-component normal mixture distribution, where each component distribution is $N_2(\mu^m, I_2)$ with $\mu^1 = (0, 0)'$, $\mu^2 = (1.0, 2.0)'$, and $\mu^3 = (2.0, 1.0)'$. The mixing proportion is $\pi^1 = \pi^2 = \pi^3 = 1/3$. We set $\Delta = \{(-\infty, z_{q_1}], (z_{q_1}, z_{q_2}), \ldots, (z_{q_{t-1}}, \infty)\}$, where z_q is the $100 \times q$ percentile of the empirical distribution.

Table 4: Selection Frequencies of the Number of Components for Chi-squared Mixtures: Two Variables

Selection frequencies by the rk statistic using (X, Y) from chi-squared mixtures												
	Partition 1: $t = 4$ with $(q_1, q_2, q_3) = (0.25, 0.5, 0.75)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.000	0.714	0.273	0.013	0.000	0.069	0.898	0.033	0.000	0.000	0.958	0.042
AIC	0.000	0.489	0.456	0.055	0.000	0.015	0.860	0.125	0.000	0.000	0.873	0.127
BIC	0.012	0.970	0.018	0.000	0.000	0.655	0.345	0.000	0.000	0.000	0.998	0.002
HQ	0.000	0.826	0.165	0.009	0.000	0.169	0.811	0.020	0.000	0.000	0.969	0.031
d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
	Partition 2: $t = 4$ with $(q_1, q_2, q_3) = (0.1, 0.5, 0.9)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.000	0.663	0.303	0.034	0.000	0.094	0.863	0.043	0.000	0.000	0.965	0.035
AIC	0.000	0.449	0.469	0.082	0.000	0.019	0.857	0.124	0.000	0.000	0.852	0.148
BIC	0.003	0.938	0.058	0.001	0.000	0.726	0.273	0.001	0.000	0.001	0.996	0.003
HQ	0.000	0.771	0.209	0.020	0.000	0.230	0.738	0.032	0.000	0.000	0.977	0.023
d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
	Partition 3: $t = 6$ with $(q_1, q_2, q_3, q_4, q_5) = (0.1, 0.25, 0.5, 0.75, 0.9)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.000	0.691	0.293	0.016	0.000	0.055	0.913	0.032	0.000	0.000	0.965	0.035
AIC	0.000	0.465	0.525	0.010	0.000	0.012	0.969	0.019	0.000	0.000	0.977	0.023
BIC	0.005	0.966	0.029	0.000	0.000	0.873	0.127	0.000	0.000	0.000	1.000	0.000
HQ	0.000	0.877	0.122	0.001	0.000	0.248	0.752	0.000	0.000	0.000	1.000	0.000
d.f.	25.000	16.000	9.000	0.000	25.000	16.000	9.000	0.000	25.000	16.000	9.000	0.000

Notes: The true number of components is $M = 3$. $(X, Y)'$ follows a 3-component chi-squared mixture distribution, where X^m and Y^m are independently drawn from chi-squared distributions with k_x^m and k_y^m degrees of freedom, respectively, with $(k_x^1, k_y^1) = (1, 1)$, $(k_x^2, k_y^2) = (3, 6)$, and $(k_x^3, k_y^3) = (6, 3)$. The mixing proportion is $\pi^1 = \pi^2 = \pi^3 = 1/3$. We set $\Delta = \{(-\infty, z_{q_1}], (z_{q_1}, z_{q_2}), \ldots, (z_{q_{t-1}}, \infty)\}$, where z_q is the $100 \times q$ percentile of the empirical distribution.

Table 5: Selection Frequencies of the Number of Components for Gamma Mixtures: Two Variables

Selection frequencies by the rk statistic using (X, Y) from gamma mixtures												
	Partition 1: $t = 4$ with $(q_1, q_2, q_3) = (0.25, 0.5, 0.75)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.828	0.161	0.011	0.000	0.203	0.560	0.227	0.010	0.000	0.038	0.919	0.043
AIC	0.653	0.306	0.036	0.005	0.107	0.424	0.415	0.054	0.000	0.006	0.843	0.151
BIC	0.998	0.002	0.000	0.000	0.993	0.007	0.000	0.000	0.339	0.434	0.227	0.000
HQ	0.964	0.036	0.000	0.000	0.689	0.253	0.057	0.001	0.002	0.145	0.826	0.027
d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
	Partition 2: $t = 4$ with $(q_1, q_2, q_3) = (0.1, 0.5, 0.9)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.699	0.240	0.055	0.006	0.108	0.488	0.378	0.026	0.000	0.008	0.946	0.046
AIC	0.540	0.341	0.097	0.022	0.058	0.296	0.564	0.082	0.000	0.002	0.864	0.134
BIC	0.959	0.035	0.006	0.000	0.954	0.039	0.007	0.000	0.090	0.299	0.610	0.001
HQ	0.862	0.117	0.019	0.002	0.510	0.337	0.144	0.009	0.000	0.032	0.930	0.038
d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
	Partition 3: $t = 6$ with $(q_1, q_2, q_3, q_4, q_5) = (0.1, 0.25, 0.5, 0.75, 0.9)$											
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.742	0.216	0.039	0.003	0.142	0.591	0.248	0.019	0.000	0.008	0.961	0.031
AIC	0.556	0.356	0.086	0.002	0.092	0.426	0.470	0.012	0.000	0.000	0.977	0.023
BIC	0.978	0.022	0.000	0.000	0.997	0.003	0.000	0.000	0.535	0.395	0.070	0.000
HQ	0.922	0.073	0.005	0.000	0.770	0.205	0.025	0.000	0.001	0.114	0.885	0.000
d.f.	25.000	16.000	9.000	0.000	25.000	16.000	9.000	0.000	25.000	16.000	9.000	0.000

Notes: The true number of components is $M = 3$. $(X, Y)'$ follows a 3-component chi-squared mixture distribution, where each component distribution is $(X^m, Y^m) \sim (\text{Gamma}(k_1^m, 1), \text{Gamma}(k_2^m, 1))$ with $(k_1^1, k_2^1) = (1, 1)$, $(k_1^2, k_2^2) == (1.5, 3)$, and $(k_1^3, k_2^3) == (3, 1.5)$. The mixing proportion is $\pi^1 = \pi^2 = \pi^3 = 1/3$. We set $\Delta = \{(-\infty, z_{q_1}], (z_{q_1}, z_{q_2}), \ldots, (z_{q_{t-1}}, \infty)\}$, where z_q is the $100 \times q$ percentile of the empirical distribution.

Table 6: Selection Frequencies of the Number of Components: Four Variables

Selection frequencies by the average rk statistic constructed from simultaneously using 3 different groupings												
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
ave-rk ($\alpha = 0.05$)	0.002	0.749	0.244	0.005	0.000	0.041	0.924	0.035	0.000	0.000	0.941	0.059
AIC by ave-rk	0.000	0.483	0.504	0.013	0.000	0.006	0.927	0.067	0.000	0.000	0.876	0.124
BIC by ave-rk	0.052	0.944	0.004	0.000	0.000	0.722	0.278	0.000	0.000	0.000	0.994	0.006
HQ by ave-rk	0.001	0.916	0.083	0.000	0.000	0.132	0.856	0.012	0.000	0.000	0.977	0.023
mean d.f.	11.000	6.992	2.513	0.000	11.000	6.181	2.213	0.000	11.000	6.004	2.015	0.000
Selection frequencies by the MLE-based model selection under parametric multi-dimensional normal distribution												
	$N = 500$				$N = 2000$				$N = 8000$			
	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
AIC by MLE	0.000	0.034	0.592	0.374	0.000	0.000	0.658	0.342	0.000	0.000	0.733	0.267
BIC by MLE	0.000	0.985	0.015	0.000	0.000	0.105	0.895	0.000	0.000	0.000	1.000	0.000
HQ by MLE	0.000	0.553	0.444	0.003	0.000	0.000	0.998	0.002	0.000	0.000	1.000	0.000
d.f.	4	9	14	19	4	9	14	19	4	9	14	19
Selection frequencies by the rk statistic using a single grouping (X^α, Y^α)												
$X^1 = (W_1, W_2)$	$N = 500$				$N = 2000$				$N = 8000$			
$Y^1 = (W_3, W_4)$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.004	0.704	0.274	0.018	0.000	0.071	0.893	0.036	0.000	0.000	0.963	0.037
AIC	0.002	0.490	0.445	0.063	0.000	0.024	0.855	0.121	0.000	0.000	0.849	0.151
BIC	0.297	0.689	0.013	0.001	0.000	0.644	0.354	0.002	0.000	0.000	0.997	0.003
HQ	0.038	0.792	0.158	0.012	0.000	0.154	0.818	0.028	0.000	0.000	0.971	0.029
mean d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
$X^2 = (W_1, W_3)$	$N = 500$				$N = 2000$				$N = 8000$			
$Y^2 = (W_2, W_4)$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.135	0.828	0.033	0.004	0.000	0.888	0.104	0.008	0.000	0.549	0.431	0.020
AIC	0.042	0.825	0.111	0.022	0.000	0.735	0.223	0.042	0.000	0.321	0.590	0.089
BIC	0.775	0.225	0.000	0.000	0.001	0.999	0.000	0.000	0.000	0.991	0.009	0.000
HQ	0.297	0.696	0.007	0.000	0.000	0.969	0.029	0.002	0.000	0.796	0.198	0.006
mean d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000
$X^3 = (W_1, W_4)$	$N = 500$				$N = 2000$				$N = 8000$			
$Y^3 = (W_2, W_3)$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$	$M = 1$	$M = 2$	$M = 3$	$M \geq 4$
SHT ($\alpha = 0.05$)	0.001	0.735	0.247	0.017	0.000	0.123	0.839	0.038	0.000	0.000	0.959	0.041
AIC	0.000	0.535	0.403	0.062	0.000	0.041	0.837	0.122	0.000	0.000	0.861	0.139
BIC	0.166	0.819	0.015	0.000	0.000	0.754	0.244	0.002	0.000	0.006	0.994	0.000
HQ	0.006	0.848	0.139	0.007	0.000	0.261	0.710	0.029	0.000	0.000	0.968	0.032
mean d.f.	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000	9.000	4.000	1.000	0.000

Notes: The true number of components is $M = 3$. $W = (W_1, W_2, W_3, W_4)'$ follows a 3-component normal mixture distribution, where each component distribution is $N_4(\mu^m, I_4)$ for $m = 1, 2, 3$. The parameter values are $\pi^1 = \pi^2 = \pi^3 = 1/3$, $\mu^1 = (0, 0, 0, 0)'$, $\mu^2 = (1.0, 2.0, 0.5, 1.0)$, and $\mu^3 = (2.0, 1.0, 1.0, 0.5)$.

Table 7: Selection Frequencies of the Number of Components: Eight Variables with $\mu^1 = (0,0,0,0,0,0,0,0)'$, $\mu^2 = (1.0,2.0,0.5,1.0,0.75,1.25,0.25,0.5)'$, and $\mu^3 = (2.0,1.0,1.0,0.5,1.25,0.75,0.5,0.25)'$

	$N=500$				$N=2000$				$N=8000$			
Selection frequencies based on the maximum of 70 modified ave-rk statistics												
	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$
max ($\alpha=0.10$)	0.000	0.655	0.344	0.001	0.000	0.022	0.956	0.022	0.000	0.000	0.875	0.125
rk$^+$ ($\alpha=0.05$)	0.000	0.746	0.253	0.001	0.000	0.057	0.933	0.010	0.000	0.000	0.912	0.088
($\alpha=0.01$)	0.000	0.858	0.142	0.000	0.000	0.165	0.835	0.000	0.000	0.000	0.956	0.044
Mean of the selection frequencies across 70 modified ave-rk statistics												
	$N=500$				$N=2000$				$N=8000$			
	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$
mean ($\alpha=0.10$)	0.101	0.817	0.078	0.003	0.003	0.704	0.281	0.012	0.000	0.329	0.634	0.037
of 70 ($\alpha=0.05$)	0.142	0.810	0.047	0.001	0.005	0.776	0.214	0.005	0.000	0.408	0.574	0.019
SHT ($\alpha=0.01$)	0.245	0.740	0.015	0.000	0.012	0.866	0.121	0.001	0.000	0.549	0.446	0.005
mean of AICs	0.012	0.867	0.119	0.003	0.000	0.587	0.399	0.013	0.000	0.222	0.734	0.044
mean of BICs	0.284	0.715	0.001	0.000	0.035	0.942	0.023	0.000	0.000	0.832	0.167	0.000
mean of HQs	0.078	0.909	0.013	0.000	0.004	0.878	0.117	0.001	0.000	0.573	0.424	0.004

Notes: The true number of components is $M=3$. W follows a 3-component normal mixture distribution $\sum_{m=1}^{3}\pi^m N_8(\mu^m, I_8)$. The parameter values are $\pi^1=\pi^2=\pi^3=1/3$, $\mu^1=(0,0,0,0,0,0,0,0)'$, $\mu^2 = (1.0,2.0,0.5,1.0,0.75,1.25,0.25,0.5)'$, and $\mu^3=(2.0,1.0,1.0,0.5,1.25,0.75,0.5,0.25)'$. The modified rk statistic with $c=0.01\times\hat{s}_1$ is used, where $\hat{s}_1$ is the estimated largest singular value of Ω_r.

Table 8: Selection Frequencies of the Number of Components: Eight Variables, where Three Variables do not have a mixture distribution with $\mu^1=(0,0,0,0,0,0,0,0)'$, $\mu^2 = (1.0,2.0,0.5,1.0,0.75,0,0,0)'$, and $\mu^3=(2.0,1.0,1.0,0.5,1.25,0,0,0)'$

	$N=500$				$N=2000$				$N=8000$			
Selection frequencies based on the maximum of 70 modified ave-rk statistics												
	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$
max ($\alpha=0.10$)	0.000	0.848	0.152	0.000	0.000	0.200	0.792	0.008	0.000	0.000	0.974	0.026
rk$^+$ ($\alpha=0.05$)	0.000	0.880	0.120	0.000	0.000	0.279	0.718	0.003	0.000	0.000	0.979	0.021
($\alpha=0.01$)	0.003	0.934	0.063	0.000	0.000	0.450	0.550	0.000	0.000	0.000	0.985	0.015
Mean of selection frequencies across the 70 modified ave-rk statistics												
	$N=500$				$N=2000$				$N=8000$			
	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$	$M=1$	$M=2$	$M=3$	$M\geq 4$
mean ($\alpha=0.10$)	0.310	0.655	0.034	0.001	0.127	0.789	0.081	0.004	0.070	0.811	0.112	0.008
of 70 ($\alpha=0.05$)	0.372	0.608	0.019	0.000	0.149	0.789	0.060	0.002	0.075	0.831	0.089	0.004
SHT ($\alpha=0.01$)	0.493	0.501	0.006	0.000	0.192	0.769	0.038	0.000	0.086	0.843	0.070	0.001
mean of AICs	0.077	0.874	0.048	0.001	0.034	0.846	0.117	0.004	0.025	0.806	0.160	0.009
mean of BICs	0.466	0.534	0.000	0.000	0.210	0.780	0.010	0.000	0.101	0.850	0.049	0.000
mean of HQs	0.229	0.766	0.004	0.000	0.108	0.854	0.037	0.000	0.066	0.865	0.068	0.001

Notes: The true number of components is $M=3$. W follows a 3-component normal mixture distribution $\sum_{m=1}^{3}\pi^m N_8(\mu^m, I_8)$. The parameter values are $\pi^1=\pi^2=\pi^3=1/3$, $\mu^1=(0,0,0,0,0,0,0,0)'$, $\mu^2 = (0.5,1.0,0.25,0.5,0.75,0,0,0)'$, and $\mu^3=(1.0,0.5,0.5,0.25,0.25,0,0,0)'$. The modified rk statistic with $c=0.01\times\hat{s}_1$ is used, where $\hat{s}_1$ is the estimated largest singular value of Ω_r.

D Additional results from empirical examples

Tables 9-11 report additional tables for the empirical examples in Section 5.

Table 9: Intergenerational Social Mobility in Great Britain

(1) British Social Mobility Data (8×8 Table)								
Father's	Subject's Status							
Status	1	2	3	4	5	6	7	8
1	50	19	26	8	7	11	6	2
2	16	40	34	18	11	20	8	3
3	12	35	65	66	35	88	23	21
4	11	20	58	110	40	183	64	32
5	2	8	12	23	25	46	28	12
6	12	28	102	162	90	553	230	177
7	0	6	19	40	21	158	143	71
8	0	3	14	32	15	126	91	106

(2) Selected Value of a Lower Bound on M (8×8 Table)			
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
SHT	$M = 7$	$M = 7$	$M = 6$
	AIC	BIC	HQ
Information Criteria	$M = 8$	$M = 5$	$M = 6$
No. of Observations	3497		

Notes: The data are from Table 1.C of Clogg (1981). Occupational categories in Panel (1) are as follows. 1: professional and high administrative; 2: managerial and executive; 3, inspectional, supervisory, and other non-manual (high grade); 4: inspectional, supervisory, and other non-manual (low grade); 5: routine grades of nonmanual; 6: skilled manual; 7: semi-skilled manual; and 8: unskilled manual. Panel (2) reports the result from the 5×5 table in which categories 2 and 3, categories 5 and 6, and categories 7 and 8 were combined.

Table 10: Type of Trade and Ethnic Group Data, Amsterdam and Rotterdam

(1) Cross-Classification by Ethnic Group and Type of Trade												
	Amsterdam						Rotterdam					
	Types of Trade						Types of Trade					
Ethnic Group	1	2	3	4	5	Total	1	2	3	4	5	Total
Dutch	382	367	788	113	28	1678	323	209	459	91	153	1235
Turks	14	21	3	8	10	56	29	30	2	15	14	90
Moroccans	12	36	2	5	7	62	8	17	2	13	5	45
Antilleans	8	6	2	1	2	19	5	4	3	4	3	19
Surinamese	44	33	33	17	24	151	35	31	28	19	33	146
Others	208	97	86	26	39	456	82	18	19	16	12	147
Total	668	560	914	170	110	2422	482	309	513	158	220	1682

Notes: The data are from Table 2a of van der Heijden et al. (2002). In the original table, there are 8 ethnic groups but we have merged the "Cape Verdeans" and the "Ghanaians" into the "Other" ethnic group because they are relatively small ethnic minorities. Types of trade in Panel (1) are as follows. 1: wholesale trade; 2: retail trade; 3: producer services; 4: catering and restaurants; 5: personal services.

(2) Selected Value of a Lower Bound on M						
	Amsterdam			Rotterdam		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
SHT	M=4	M=4	M=3	M=3	M=3	M=3
	AIC	BIC	HQ	AIC	BIC	HQ
Information Criteria	M=4	M=2	M=3	M=3	M=3	M=3
No. of Observations	2422			1682		

Table 11: Response Patterns in Five-item Subsets of LSAT and the Estimated Number of Latent Ability Distributions

	Number of Components Selected based on 5 Items					
	LSAT 6			LSAT 7		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
max-rk$^+$ statistic	M = 2	M = 2	M = 2	M = 3	M = 3	M = 2
	Number of Components Selected based on 4 items					
	LSAT 6			LSAT 7		
SHT	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\{W_1, W_2, W_3, W_4\}$	M = 2	M = 2	M = 2	M = 3	M = 2	M = 2
$\{W_2, W_3, W_4, W_5\}$	M = 2	M = 2	M = 2	M = 2	M = 2	M = 2
$\{W_1, W_3, W_4, W_5\}$	M = 2	M = 2	M = 2	M = 2	M = 2	M = 2
$\{W_1, W_2, W_4, W_5\}$	M = 2	M = 2	M = 2	M = 2	M = 2	M = 2
$\{W_1, W_2, W_3, W_5\}$	M = 3	M = 2	M = 2	M = 3	M = 3	M = 2
Model Selection	*AIC*	*BIC*	*HQ*	*AIC*	*BIC*	*HQ*
$\{W_1, W_2, W_3, W_4\}$	M = 2	M = 2	M = 2	M = 3	M = 2	M = 2
$\{W_2, W_3, W_4, W_5\}$	M = 2	M = 2	M = 2	M = 2	M = 2	M = 2
$\{W_1, W_3, W_4, W_5\}$	M = 2	M = 2	M = 2	M = 3	M = 2	M = 2
$\{W_1, W_2, W_4, W_5\}$	M = 2	M = 1	M = 2	M = 2	M = 2	M = 2
$\{W_1, W_2, W_3, W_5\}$	M = 3	M = 1	M = 2	M = 3	M = 2	M = 2
No. of observations	1000			1000		

Notes: The data are from Table 1 of Mislevy (1984).

References

Allman, E. S., Matias, C., and Rhodes, J. A. (2009). "Identifiability of parameters in latent structure models with many observed variables." *The Annals of Statistics*, 37, 3099-3132.

Borkowf, C. B. (2000). "On multidimensional contingency tables with categories defined by the empirical quantiles of the marginal data." *Journal of Statistical Planning and Inference*, 91, 33-51.

Clogg, C. C. (1981). "Latent Structure Models of Mobility." *The American Journal of Sociology*, 86, 836-868.

Cohen, J. E. and Rothblum, U. G. (1993). "Nonnegative ranks, decompositions, and factorizations of nonnegative matrices." *Linear Algebra and its Applications*, 190, 149-168.

Kleibergen, F. and Paap, R. (2006). "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133, 97-126.

Mislevy, R. J. (1984). "Estimating latent distribution." *Psychometrika*, 49, 359-381.

van der Heijden, P. G. M., van der Ark, L. A., and Mooijaart, A. (2002). "Some examples of latent budget analysis and its extensions." In: Hagenaars, J. A. and McCutcheon, A. L. (Eds.), *Applied Latent Class Analysis*, Cambridge: Cambridge University Press, 107-136.