Supplementary Material of Additional Simulations and Proofs

1 Simulation result on all pairwise comparisons

In Section 4 of the main manuscript, we provide simulation results under various settings for many-to-one comparisons. We investigate many methods including MNQ, MNQ naive, Bonferroni, Dunn-Sidák, Holm and Schéffe. In the manuscript, only MNQ, MNQ naive, and Bonferroni are reported. In this supplementary file, we report the result for all six methods in Table 1-4.

We also provide additional simulations for all pairwise comparisons. We simulate multivariate normal, multiviate probit and quadratic exponential models as described in Section 4 of the paper. The global null hypothesis, sample size and the two alternative configurations are the same as those used in many-to-one comparisons. We perform all pairwise comparisons with $m = 4$ and $p = 10$. For the multivariate normal, multivariate probit, and quadratic exponential models, we consider $\rho = 0$, or 0.5. For the quadratic exponential, we consider $w = 0, 0.5$. The results are summarized in Table 5. We again observe that the MNQ approach has the best performance. MNQ maintains good control of the FWER except for the case of quadratic exponential model with $\rho = 0.5$, where it is slightly above 0.05. The naive MNQ either has either very large FWER or very small FWER, indicating its poor control of the error rate. Among all the methods which maintain good control of FWER, the MNQ method achieves the highest power. In addition, we consider the Tukey approach, as it is a commonly used testing procedure in all pairwise comparisons.

Table 1: Simulations results for multivariate normal with exchangeable Σ

	ρ	m	p	MNQ	naive	Bonf	S-D	Holm	Scheffé
FWER	0	4	10	0.0545	0.0553	0.0419	0.0427	0.0419	0.0007
a_1				0.8164	0.8166	0.7894	0.7918	0.7894	0.2801
a_2				0.8057	0.8053	0.7617	0.7738	0.7617	0.2226
ind a_2				0.1816	0.1816	0.1683	0.1701	0.1770	0.0448
FWER			20	0.0511	0.0502	0.0352	0.0363	0.0352	0.0000
a_1				0.7487	0.7476	0.7062	0.7086	0.7062	0.0259
a_2				0.7150	0.7134	0.6687	0.6628	0.6687	0.0162
ind a_2				0.1540	0.1535	0.1416	0.1401	0.1504	0.0032
FWER		10	10	0.0479	0.0471	0.0375	0.0378	0.0375	0.0001
a_1				0.9983	0.9983	0.9979	0.9980	0.9979	0.9284
a_2				0.9993	0.9993	0.9986	0.9990	0.9986	0.8792
ind a_2				0.2964	0.2963	0.2844	0.2857	0.2979	0.1787
FWER			20	0.0487	0.0485	0.0363	0.0373	0.0363	0.0000
a_1				0.9981	0.9980	0.9967	0.9969	0.9967	0.5428
a_2				0.9978	0.9977	0.9963	0.9957	0.9963	0.4137
ind a_2				0.2688	0.2681	0.2552	0.2555	0.2653	0.0828
FWER	0.5	4	10	0.0513	0.0977	0.0390	0.0398	0.0390	0.0007
a_1				0.7235	0.8129	0.6867	0.6904	0.6867	0.1947
a_2				0.6922	0.8042	0.6615	0.6571	0.6615	0.1497
ind a_2				0.1514	0.1878	0.1442	0.1415	0.1507	0.0300
FWER			20	0.0510	0.1029	0.0377	0.0385	0.0377	0.0000
a_1				0.6420	0.7526	0.5950	0.5985	0.5950	0.0140
a_2				0.6031	0.7322	0.5437	0.5508	0.5437	0.0076
ind a_2				0.1274	0.1607	0.1135	0.1150	0.1222	0.0015
FWER		10	10	0.0520	0.2079	0.0410	0.0417	0.0410	0.0000
a_1				0.9570	0.9936	0.9466	0.9469	0.9466	0.6125
a_2				0.9555	0.9982	0.9438	0.9431	0.9438	0.5062
ind a_2				0.2383	0.3381	0.2273	0.2274	0.2375	0.1019
FWER			20	0.0459	0.2271	0.0328	0.0337	0.0328	0.0000
a_1				0.9403	0.9938	0.9224	0.9243	0.9224	0.1408
a_2				0.9222	0.9948	0.8932	0.8968	0.8932	0.0871
ind a_2				0.2118	0.3072	0.1981	0.1997	0.2061	0.0174

2 A skewed distribution example

Here, we consider a multivariate gamma distribution which has marginal univariate gamma distribution and a covariance structure. To generate a multivariate gamma model, let g_1 be $m \times 1$ independent vectors from a gamma distribution with shape parameters γ_1, a positive vector of dimension m. Define $G = Kg_1$, where K is a full rank matrix

Table 2: Simulations results for multivariate normal with unstructured Σ (many-to-one)

	m	p	MNQ	naive	Bonf	S-D	Holm	Scheffé
FWER	4	10	0.0464	0.0729	0.0345	0.0358	0.0345	0.0004
a_1			0.6348	0.7089	0.5962	0.5992	0.5962	0.1358
a_2			0.5123	0.6045	0.4763	0.4714	0.4763	0.0614
ind a_2			0.1100	0.1341	0.1019	0.1022	0.1071	0.0123
FWER		20	0.0390	0.0664	0.0285	0.0290	0.0285	0.0000
a_1			0.5205	0.6081	0.4694	0.4736	0.4694	0.0046
a_2			0.3913	0.4864	0.3378	0.3428	0.3378	0.0011
ind a_2			0.0811	0.1018	0.0687	0.0702	0.0749	0.0002
FWER	10	10	0.0472	0.0407	0.0360	0.0367	0.0360	0.0004
a_1			0.6310	0.6102	0.5906	0.5940	0.5906	0.1198
a_2			0.5025	0.4779	0.4560	0.4599	0.4560	0.0537
ind a_2			0.1088	0.1028	0.0974	0.0982	0.1032	0.0107
FWER		20	0.0361	0.0302	0.0262	0.0267	0.0262	0.0000
a_1			0.5078	0.4865	0.4585	0.4615	0.4585	0.0025
a_2			0.3668	0.3448	0.3148	0.3167	0.3148	0.0010
ind a_2			0.0742	0.0698	0.0637	0.0641	0.0692	0.0002

called the incidence matrix with all entries equal to either zero or one that follows certain properties (Ronning, 1977). Then G has a multivariate gamma distribution with shape parameter $\alpha = K\gamma_1$ and covariance matrix $\Sigma = K\Gamma_1 K^T$, where the (diagonal) matrix Γ_1 is the variance matrix of g_1.

Given n independent multivariate gamma vectors $Y = (y_1, y_2, \ldots, y_n)^T$, with $y_i = (y_{i1}, \ldots, y_{im})^T$. The univariate composite loglikelihood function can be formulated as

$$
\begin{aligned}
& cl(\beta; Y) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} \left(-\frac{\nu y_{ij}}{\mu_{ij}} - \nu \log \mu_{ij} + \nu log \nu + (\nu - 1)\log y_{ij} - \log \Gamma(\nu)\right),
\end{aligned}
$$

where $\mu_{ij} = E(y_{ij})$, ν is the shape parameter, and μ_{ij}/ν is the scale parameter. We used the log link to define the mean parameter: $\mu_{ij} =$

Table 3: Simulation results for the probit model (many-to-one)

	ρ	m	p	MNQ	naive	Bonf	S-D	Holm	Scheffé
FWER	0	4	10	0.0530	0.0506	0.0413	0.0424	0.0413	0.0001
a_1				0.8700	0.8705	0.8477	0.8496	0.8477	0.3420
a_2				0.9114	0.9109	0.8885	0.8907	0.8885	0.3572
ind a_2				0.2166	0.2156	0.2039	0.2061	0.2136	0.0718
FWER			20	0.0528	0.0503	0.0389	0.0395	0.0389	0.0000
a_1				0.8258	0.8232	0.7902	0.7924	0.7902	0.0460
a_2				0.8547	0.8511	0.8149	0.8159	0.8149	0.0410
ind a_2				0.1887	0.1878	0.1769	0.1765	0.1862	0.0082
FWER		10	10	0.0526	0.0515	0.0423	0.0428	0.0423	0.0005
a_1				0.9996	0.9996	0.9995	0.9995	0.9995	0.9641
a_2				1.0000	1.0000	1.0000	1.0000	1.0000	0.9695
ind a_2				0.3330	0.3319	0.3168	0.3188	0.3332	0.2005
FWER			20	0.0527	0.0508	0.0364	0.0375	0.0364	0.0000
a_1				0.9993	0.9995	0.9985	0.9985	0.9985	0.6596
a_2				1.0000	0.9999	0.9997	0.9997	0.9997	0.6603
ind a_2				0.2973	0.2956	0.2811	0.2812	0.2925	0.1321
FWER	0.5	4	10	0.0508	0.0793	0.0393	0.0404	0.0393	0.0003
a_1				0.8102	0.8601	0.7808	0.7841	0.7808	0.2726
a_2				0.8530	0.9038	0.8305	0.8258	0.8305	0.2689
ind a_2				0.1970	0.2206	0.1864	0.1867	0.1954	0.0542
FWER			20	0.0585	0.0915	0.0406	0.0415	0.0406	0.0000
a_1				0.7578	0.8196	0.7082	0.7106	0.7082	0.0264
a_2				0.7891	0.8534	0.7365	0.7428	0.7365	0.0247
ind a_2				0.1727	0.1942	0.1571	0.1593	0.1666	0.0049
FWER		10	10	0.0513	0.1437	0.0402	0.0412	0.0402	0.0005
a_1				0.9900	0.9979	0.9871	0.9876	0.9871	0.8017
a_2				0.9952	0.9997	0.9966	0.9939	0.9966	0.8038
ind a_2				0.2815	0.3585	0.2704	0.2710	0.2831	0.1629
FWER			20	0.0543	0.1622	0.0382	0.0389	0.0382	0.0000
a_1				0.9862	0.9974	0.9784	0.9787	0.9784	0.3081
a_2				0.9935	0.9998	0.9894	0.9883	0.9894	0.3006
ind a_2				0.2575	0.3250	0.2450	0.2444	0.2555	0.0601

$\exp\{\vec{x}_{ij}\beta\}$. Denote $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^T$. Under this set up, we have

$$cl^{(1)}(\beta;Y) \;=\; \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T V(\mu)_i^{-1}(y_i - \mu_i),$$

Table 4: Simulation results for the quadratic exponential model (many-to-one)

	w	m	p	MNQ	naive	Bonf	S-D	Holm	Scheffé
FWER	0	4	10	0.0514	0.0562	0.0400	0.0403	0.0400	0.0001
a_1				0.5390	0.5534	0.5010	0.5046	0.5010	0.0777
a_2				0.7067	0.7240	0.6573	0.6636	0.6573	0.0935
ind a_2				0.1956	0.2057	0.1765	0.1778	0.1875	0.0197
FWER			20	0.0561	0.0767	0.0404	0.0412	0.0404	0.0000
a_1				0.4551	0.4853	0.3990	0.4021	0.3990	0.0025
a_2				0.6040	0.6365	0.5237	0.5403	0.5237	0.0027
ind a_2				0.1546	0.1669	0.1303	0.1336	0.1406	0.0005
FWER		10	10	0.0491	0.0549	0.0381	0.0384	0.0381	0.0001
a_1				0.5391	0.5535	0.5010	0.5046	0.5010	0.0779
a_2				0.7066	0.7239	0.6573	0.6636	0.6573	0.0934
ind a_2				0.1956	0.2057	0.1765	0.1778	0.1875	0.0197
FWER			20	0.0561	0.0767	0.0404	0.0412	0.0404	0.0000
a_1				0.4548	0.4849	0.3989	0.4020	0.3989	0.0026
a_2				0.5971	0.6309	0.5255	0.5361	0.5255	0.0013
ind a_2				0.1536	0.1663	0.1305	0.1338	0.1409	0.0003
FWER	0.5	4	10	0.0521	0.0000	0.0417	0.0424	0.0417	0.0002
a_1				0.7864	0.0307	0.7546	0.7582	0.7546	0.2329
a_2				0.9050	0.0444	0.8800	0.8772	0.8800	0.2531
ind a_2				0.3027	0.0090	0.2820	0.2818	0.2983	0.0551
FWER			20	0.0509	0.0000	0.0377	0.0383	0.0377	0.0000
a_1				0.7214	0.0158	0.6739	0.6769	0.6739	0.0178
a_2				0.8460	0.0148	0.7998	0.7976	0.7998	0.0132
ind a_2				0.2580	0.0030	0.2306	0.2322	0.2428	0.0027
FWER		10	10	0.0521	0.0000	0.0417	0.0424	0.0417	0.0002
a_1				0.7864	0.0307	0.7546	0.7582	0.7546	0.2329
a_2				0.9141	0.0407	0.8800	0.8855	0.8800	0.2518
ind a_2				0.3065	0.0083	0.2820	0.2852	0.2983	0.0549
FWER			20	0.0509	0.0000	0.0378	0.0384	0.0378	0.0000
a				0.7202	0.0161	0.6731	0.6760	0.6731	0.0178
a_2				0.8460	0.0148	0.7998	0.7976	0.7998	0.0132
ind a_2				0.2580	0.0030	0.2306	0.2322	0.2428	0.0027

where $V_i = \mathrm{diag}(\mu_{i1}^2, \cdots, \mu_{im}^2)/\nu$, and

$$\begin{aligned} H(\beta) &= n^{-1} \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1} \left(\frac{\partial \mu_i}{\partial \beta} \right), \\ J(\beta) &= n^{-1} \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1} \mathrm{cov}(y_i) V_i^{-1} \left(\frac{\partial \mu_i}{\partial \beta} \right). \end{aligned}$$

Table 5: Simulations results of all pairwise comparisons for three multivariate distributions

model		ρ	MNQ	naive	Bonf	S-D	Scheffé	Tukey
normal	FWER	0	0.0537	0.0562	0.0411	0.0420	0.0038	0.0536
	a1		0.9274	0.9266	0.9096	0.9113	0.6115	0.9256
	a2		0.9800	0.9807	0.9735	0.9740	0.8173	0.9792
	FWER	0.5	0.0484	0.1101	0.0358	0.0365	0.0032	0.0489
	a1		0.8611	0.9245	0.8325	0.8346	0.4769	0.8587
	a2		0.9492	0.9775	0.9346	0.9361	0.6854	0.9482
probit	FWER	0	0.0534	0.0494	0.0409	0.0412	0.0026	0.0524
	a1		0.9792	0.9790	0.9745	0.9747	0.7972	0.9791
	a2		0.9961	0.9961	0.9946	0.9946	0.9321	0.9959
	FWER	0.5	0.0523	0.0864	0.0394	0.0394	0.0023	0.0514
	a1		0.9586	0.9754	0.9467	0.9484	0.6991	0.9577
	a2		0.9885	0.9938	0.9842	0.9848	0.8707	0.9884
quad. exp.	FWER	0	0.0534	0.0631	0.0399	0.0407	0.0018	0.0530
	a1		0.7710	0.7869	0.7270	0.7301	0.3224	0.7678
	a2		0.9706	0.9741	0.9613	0.9621	0.7348	0.9701
	FWER	0.5	0.0548	0.0000	0.0388	0.0393	0.0014	0.0535
	a1		0.9360	0.0197	0.9199	0.9213	0.6417	0.9356
	a2		0.9976	0.2855	0.9957	0.9958	0.9408	0.9974

The dispersion parameter is $\frac{1}{\nu} = \frac{D(6(n-p)+nD)}{6(n-p)+2nD}$, where

$$D = \frac{2}{nm-p} \sum_{i,j} \left(\frac{y_{ij} - \mu_{ij}}{\mu_{ij}} + \log \frac{\mu_{ij}}{y_{ij}} \right).$$

Let $\widehat{V}_{in}$ denote the estimator of V_i obtained by substituting $\widehat{\mu}_{ijn}$ for μ_{ij}. We estimate $H(\beta)$ and $J(\beta)$ as

$$\begin{aligned} \widehat{H}_n &= n^{-1} \sum_{i=1}^{n} X_i^T V_{in}^{-1} X_i, \\ \widehat{J}_n &= n^{-1} \sum_{i=1}^{n} X_i^T V_{in}^{-1} \;\; \widehat{\text{cov}}_n(y_i) \;\; V_{in}^{-1} X_i, \end{aligned}$$

with empirical variance $\widehat{\text{cov}}_n(y_i) = (y_i - \widehat{\mu}_{in})(y_i - \widehat{\mu}_{in})^T$, where where $\widehat{\mu}_i$ is the vector $\widehat{\mu}_i = \exp\{X_i \widehat{\beta}_n^c\}$.

In the simulation $\nu = 1$, and under the global null hypothesis H_0, the true value of the regression parameters is set to $\beta = 0.75$, and the power is calculated under two different alternative configurations $\beta_{a_1}^T = (0.75, 0.75, 0.68, 0.75, \ldots, 0.75)$ and $\beta_{a_2}^T = (0.75, 0.80, 0.68, 0.70, 0.79, 0.69, 0.75, \ldots, 0.75)$. We simulate 10000 data sets with $m = 3$, and $p = 10$. We perform many-to-one comparisons with the MNQ, naive MNQ, Bonferroni, Dunn-Sidák, Holm and Scheffé method. We consider both independent and correlated cases. We simulate with the sample size $n = 3000$ as we found that it takes at least $n = 3000$ for the MNQ method to have the FWER fall within 2 standard deviations away from 0.05. This larger sample size is expected for a skewed distribution such as the multivariate gamma. Among all the methods, the MNQ method continues to achieve the highest power and exhibits the best performance. The results are presented in Table 6.

Table 6: FWER and power for multivariate gamma distribution (many-to-one)

		MNQ	naive	Bonf	S-D	Schéffe
FWER	independent	0.0554	0.0507	0.0437	0.0444	0.0003
a_1		0.8763	0.8777	0.8508	0.8531	0.3055
a_2		0.9906	0.9899	0.9856	0.9862	0.4526
FWER	correlated	0.0588	0.3427	0.0468	0.0479	0.0003
a_1		0.8223	0.9883	0.7853	0.7877	0.2378
a_2		0.9778	0.9999	0.9638	0.9653	0.3683

3 Some technical details

Xu and Reid (2011) provided a detailed proof of consistency under misspecification, along with a precise list of required conditions. One can obtain from their work sufficient conditions for consistency even in the well-specified setting. Here, for reference, we give a proof of some asymptotic properties of the composite likelihood estimator provided that the model is correctly specified and data is formed by n independent clusters, each with fixed sample size m.

Regularity conditions:

(A1). The marginal density function of y_{ij}, $f(y;\theta)$ is distinct for different values of y, i.e. if $\theta_1 \neq \theta_2$ then $P(f(y_{ij};\theta) \neq f(y_{ij};\theta)) > 0$, for all $j = 1,\ldots,m$.

(A2). The marginal densities of y_{ij} have common support for all θ.

(A3). The true value θ_0 is an interior point of Ω, the space of possible values of the parameter θ.

(A4). Let α and ∂^α denote the index and partial derivative operator, respectively, as in the standard multi-index notation from multi-variable calculus. The marginal density $\log f$ is three times continuously differentiable in a closed ball around θ_0. Moreover, there exists a constant c and an integrable function $M(y)$ such that

$$\left|(\partial^\alpha \partial^{\theta_i} \log f)(y;\theta)\right| \leq M(y),$$

for all $||\theta - \theta_0||_2 < c$, all $|\alpha| = 2$, and any $i = 1,\ldots,p$. Here, $||\cdot||_2$ denotes the Euclidean norm.

(A5). $J(\theta_0)$ is well-defined (i.e. exists and is finite) and invertible.

(A6). $H(\theta_0)$ is well-defined (i.e. exists and is finite) and (strictly) positive-definite.

Define the marginal composite log-likelihood function as

$$cl(\theta) \;=\; \log CL(\theta;Y) \;=\; \sum_{i=1}^{n}\sum_{j=1}^{m} \log f(y_{ij};\theta),$$

and let $cl_m(\theta;y_i) = \sum_{j=1}^{m} \log f(y_{ij};\theta))$.

Theorem 3.1. *Under the regularity conditions (A1)-(A6), there exists a solution to the composite likelihood equation, $\widehat{\theta}_n^c$, which satisfies*

$$\sqrt{n}(\widehat{\theta}_n^c - \theta_0) \;\Rightarrow\; G^{-1/2}(\theta_0)Z$$

where $G(\theta) = H(\theta)J^{-1}(\theta)H(\theta)$, and Z is a standard normal random vector.

Proof. The proof is divided into two main steps. We first show that there exists a $\widehat{\theta}_n^c$ which is of order $O(n^{-1/2})$, and then we derive its asymptotic normality.

Let $h(\theta;y) = cl(\theta;y)$. Note that for fixed y, h maps $\mathbb{R}^p$ into $\mathbb{R}$. Then, by a Taylor expansion, we have that

$$h(\theta;y) - h(\theta_0;y) \;=\; (\nabla h)(\theta_0;y)^T(\theta-\theta_0) + (\theta-\theta_0)^T(Dh)(\theta^*;y)(\theta-\theta_0),$$

where θ^* lies on a line joining θ and θ_0. We use ∇, D to denote the gradient and Hessian operators, respectively. Our goal will be to show that there exists a θ in a $n^{-1/2}$ ball of θ_0, the left hand side of the above equation is negative. This in turn will imply that there exists a CMLE which satisfies $\sqrt{n}(\widehat{\theta}_n^c - \theta_0) = O_p(1)$.

To this end, let $\theta - \theta_0 = \xi M/\sqrt{n}$, with $||\xi||_2 = 1$. Assume also that $||\theta-\theta_0||_2 < c$, that is, $M < c\sqrt{n}$. Then, by the above, we have

$$\begin{aligned} &\xi^T\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nabla cl_m)(\theta_0,y_i)\right\} + \xi^T\left\{\frac{1}{n}\sum_{i=1}^{n}(Hcl_m)(\theta^*,y_i)\right\}\xi \\ &\quad\equiv\; \xi^T b_n M + \xi^T B_n \xi M^2, \end{aligned} \tag{3.1}$$

where b_n is a random vector converging to a mean-zero Gaussian RV, and B_n is the random matrix converging to the negative definite matrix $-H(\theta_0)$. The first of these follows by the central limit theorem, along with assumption (A5). The second follows by applying the law of large numbers, along with assumptions (A4) and (A6). Note that the second fact implies also that the eigenvalues of B_n converge almost surely to the eigenvalues of $-H(\theta_0)$.

Let $\lambda_n^{(p)}$ denote the largest eigenvalue of $-B_n$, and let $S = \{\xi : ||\xi||_2 = 1\}$. Since b_n converges as a random Gaussian vector (with mean zero), and $\xi^T b_n$ is uniformly continuous on S, it follows that $\xi^T b_n$ converges to a mean-zero Gaussian process in $C(S)$, the space of continuous functions on S endowed with the uniform metric. This implies that $\xi^T b_n$ is tight in $C(S)$, and hence for all $\varepsilon > 0$, there exists an M_ε, such that

$$\limsup_n P\left(\sup_{\xi\in S} \xi^T b_n/\lambda_n^{(p)} < M_\varepsilon\right) \;\geq\; 1-\varepsilon.$$

Then, by (3.1), if $\xi^T b_n/\lambda_n^{(p)} < M$, then $\xi^T b_n M + \xi^T B_n \xi M^2 < 0$, which in turn implies that

$$\limsup_n P\left(\xi^T b_n M_\varepsilon + \xi^T B_n \xi M_\varepsilon^2 < 0 \;\; \forall \xi \in S\right) \;\geq\; 1-\varepsilon.$$

Note that if $\xi^T b_n M_\varepsilon + \xi^T B_n \xi M_\varepsilon^2 < 0 \quad \forall \xi \in S$, then, by the above and continuity of cl_m, this implies that for sufficiently large n, (with a probability of at least $1-\varepsilon$) there exists at least one local maximum on the set $B_{M_\varepsilon/\sqrt{n}}(\theta_0) \cap B_c(\theta_0)$. This implies that there exists a $\widehat{\theta}_n^c$ which satisfies $\sqrt{n}(\widehat{\theta}_n^c - \theta_0) = O_p(1)$.

Let $g(\theta; y) = cl_m^{(1)}(\theta; y) = \nabla cl_m(\theta; y)$ (this is the vector of first derivatives), then using a multivariate Taylor expansion, we have that

$$\begin{aligned} g(\widehat{\theta}_n^c; y) \;&=\; g(\theta_0; y) + \sum_{|\alpha|\leq 1} (\partial^\alpha g)(\theta_0; y)(\widehat{\theta}_n^c - \theta_0)^\alpha \\ &\quad + \sum_{|\alpha|=2} \frac{2}{\alpha!}(\widehat{\theta}_n^c - \theta_0)^\alpha \int_0^1 (1-t)(\partial^\alpha g)(\theta_0 + t(\widehat{\theta}_n^c - \theta_0); y)dt, \end{aligned}$$

again using the multi-index notation. We take $\widehat{\theta}_n^c$ to be the local maximizer found above. This time, for fixed y, g maps $\mathbb{R}^p$ into $\mathbb{R}^p$, so we have chosen to bound the error term a little differently than above. We let $R_{n,i}$ denote the third term on the right hand side of this equation when y is replaced with y_i. Next, as by definition $\sum_{i=1}^n cl_m^{(1)}(\widehat{\theta}_n^c; y_i) = 0$, we have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (Dcl_m)(\theta; y_i)^T (\widehat{\theta}_n^c - \theta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^n R_{n,i} \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^n f(\theta_0; y_i). \quad (3.2)$$

By condition (A4), we have that

$$\begin{aligned}
&\left| \sum_{|\alpha|=2} \frac{2}{\alpha!} (\widehat{\theta}_n^c - \theta_0)^\alpha \int_0^1 (1-t)(D^\alpha g)(\theta_0 + t(\widehat{\theta}_n^c - \theta_0); y) dt \right| \\
&\qquad \leq \sum_{|\alpha|=2} \frac{1}{\alpha!} |\widehat{\theta}_n^c - \theta_0|^\alpha |M(y)|,
\end{aligned}$$

from which it follows that,

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n R_{n,i} \right| \leq \left\{ \sqrt{n} ||\widehat{\theta}_n^c - \theta_0||_2^2 \right\} \left\{ \frac{1}{n} \sum_{i=1}^n |M(y_i)| \right\}.$$

The first term is then $o_p(1)$ by the first part of this proof, and by the law of large numbers (since M is integrable), the second term is $O_p(1)$. Next, consider

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n cl_m^{(2)}(\theta; y_i) - H(\theta_0) \right\} (\widehat{\theta}_n^c - \theta_0).$$

By similar argument to that above, this is also $o_p(1)$. This allows us to re-write (3.2) as

$$\sqrt{n} H(\theta_0)(\widehat{\theta}_n^c - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n f(\theta_0; y_i) + o_p(1)$$

A straightforward application of the central limit theorem shows that the term on the right hand side has a Gaussian limiting distribution with mean zero and variance $J(\theta_0)$. The full result follows. □

References

Ronning, G.(1977). A simple scheme for generating multivariate gamma distributions with non-negative covariance matrix. *Technometrics* **19**, 179–183.