Discussion Paper No. 107

Efficiency Improvement by Local Moments in Grouped Data Analysis

Naoya Sueishi, Qing-Feng Liu,
Kohtaro Hitomi
and Yoshihiko Nishiyama

June, 2006

21COE
Interfaces for Advanced Economic Analysis
Kyoto University

Efficiency Improvement by Local Moments in Grouped Data Analysis *

Naoya Sueishi
Kyoto University

Qing-Feng Liu
Kyoto University

Kohtaro Hitomi
Kyoto Institute of Technology

Yoshihiko Nishiyama
Kyoto Institute of Economic Research

Abstract

This paper proposes an efficient density estimation method for analyzing grouped data when local moments are given. We use the generalized method of moments (GMM) estimator of Hansen (1982) to incorporate the information contained in the local moments. We show that our estimator is more efficient than the classical maximum likelihood estimator for grouped data. We also construct a specification test statistic based on moment conditions. Monte Carlo experiments suggest that our estimator performs remarkably well and the specification test has good size properties even in finite samples.

Keywords: Grouped data; GMM.

1 Introduction

The purpose of this paper is to improve the efficiency of the estimators when the observations are grouped. We investigate the properties of Hansen's (1982) generalized method of moments (GMM) estimator applied to grouped data analysis. Economic data are often provided in a grouped form. Typical example is

*We would like to thank seminar participants at Kyoto University for significant comments. This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grand-in-Aid for 21st Century COE Program "Interfaces for Advanced Economic Analysis".

personal income data reported by government. The range of income distribution is divided into some intervals and only summary statistics corresponding to each interval are observable.

Firstly we consider the following simple example. We want to estimate the underlying distribution of the data. We have the density function $f(x, \boldsymbol{\theta})$ depending on the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. The intervals $B_1, B_2, \ldots, B_L$ are given and we observe only the frequency n_i falling in B_i. Naive maximum likelihood estimator (MLE) is given by the solution of the equation

$$\sum_{i=1}^{L} n_i \frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{1.1}$$

where $P_i(\boldsymbol{\theta}) = \int_{B_i} f(x, \boldsymbol{\theta}) dx$. Asymptotic properties of the naive MLE have been examined in several papers (see, for example, Tallis (1967)). Note that a set of frequencies $(n_1, n_2, \ldots, n_L)$ is equivalent to a sample from multinomial distribution; that is, $P(n_1, n_2, \ldots, n_L) = \frac{n}{n_1! \ldots n_L!} P_1(\boldsymbol{\theta})^{n_1} \ldots P_L(\boldsymbol{\theta})^{n_L}$, where $n = \sum_{i=1}^{L} n_i$.

Victoria-Feser and Ronchetti (1997) considered more general estimators. Victoria-Feser and Ronchetti (1997) investigated the family of minimum power divergence estimators (MPEs) of multinomial distribution.[1] The MPEs are defined by the solution of the equation

$$G(\mathbf{p}; \mathbf{P}(\boldsymbol{\theta}))^{\lambda} = \sum_{i=1}^{L} \left(\frac{p_i}{P_i(\boldsymbol{\theta})} \right)^{\lambda+1} \frac{\partial P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{1.2}$$

where $-\infty < \lambda < \infty$ is a fixed parameter, and $p_i = n_i/n$ is the relative frequency. The value $\lambda = 0$ corresponds to the naive MLE. Victoria-Feser and Ronchetti (1997) investigated the robustness properties of the estimators for various values of λ.

Instead of emphasis on robustness, we put emphasis on improving the efficiency of the estimators using of the local moments. In this paper, we consider the case where the local moments such as sample mean in each interval are also available. Although several studies have been made on the estimators based on the frequencies, there is little argument on the estimator which incorporate the information of the local moments. The information is given in the form of the local moment conditions. The local moment conditions can be written as

$$E\left[I(X \in B_i)(X - \mu_i(\boldsymbol{\theta}_0))\right] = 0, \tag{1.3}$$

where $\mu_i(\boldsymbol{\theta}) = \int_{B_i} x \frac{f(x,\boldsymbol{\theta})}{\int_{B_i} f(x,\boldsymbol{\theta}) dx} dx$ is the local mean.

We also propose the specification test of the underlying distribution based on moment conditions. The over-identifying restriction test of Hansen (1982) can be used as the specification test.

[1]For a discussion of power divergence statistics, see Cressie and Read (1984).

This paper is organized as follows. In section 2 we describe the estimation procedure utilizing the local moments. Section 3 discusses the asymptotics of the resulting estimator. Section 4 provides the specification test of the underlying distribution. Section 5 presents the results of the Monte Carlo experiments. In section 6 some concluding remarks are made. All proofs are in the Appendix.

2 The estimator

In this section we present the GMM estimator based on the local moment conditions. The estimation environment we study is as follows. Let the range of random variable X be devided into a set of fixed disjoint intervals $B_1, B_2, \ldots, B_L$. We assume that a random sample of size n is drawn from a population with density function $f(x, \boldsymbol{\theta})$, but we can observe only frequencies n_i and local sample moments:

$$\bar{x^s}_i = \frac{1}{n_i} \sum_{j \in B_i} x_j^s \tag{2.1}$$

for $i = 1, \ldots, L$ and $s = 1, \ldots, k$. For notational simplicity, henceforth, we consider only the case $k = 1$, namely, local mean is available. Extension to the higher order moments is straightforward. The moment conditions can be written as

$$E[\mathbf{g}(X; \boldsymbol{\theta}_0)] \equiv E\left[\begin{pmatrix} \mathbf{g}_1(X; \boldsymbol{\theta}_0) \\ \mathbf{g}_2(X; \boldsymbol{\theta}_0) \end{pmatrix}\right] = \mathbf{0}, \tag{2.2}$$

where

$$\mathbf{g}_1(x; \boldsymbol{\theta}) = \sum_{i=1}^{L} I(x \in B_i) \frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{2.3}$$

$$\mathbf{g}_2(x; \boldsymbol{\theta}) = \begin{pmatrix} I(x \in B_1)(x - \mu_1(\boldsymbol{\theta})) \\ I(x \in B_2)(x - \mu_2(\boldsymbol{\theta})) \\ \vdots \\ I(x \in B_L)(x - \mu_L(\boldsymbol{\theta})) \end{pmatrix}, \tag{2.4}$$

and $\boldsymbol{\theta}_0$ is the true parameter value. The sample counterpart of the moment conditions, $\hat{\mathbf{g}}(\boldsymbol{\theta}, \mathcal{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}(X_i, \boldsymbol{\theta})$, is given by

$$\hat{\mathbf{g}}_1(\boldsymbol{\theta}; \mathcal{X}_n) \equiv \frac{1}{n} \sum_{j=1}^{n} \mathbf{g}_1(x_j; \boldsymbol{\theta}) = \sum_{i=1}^{L} \frac{n_i}{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log P_i(\boldsymbol{\theta}) \tag{2.5}$$

$$\hat{\mathbf{g}}_2(\boldsymbol{\theta}; \mathcal{X}_n) \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_2(x_j; \boldsymbol{\theta}) = \begin{pmatrix} \frac{n_1}{n}(\bar{x}_1 - \mu_1(\boldsymbol{\theta})) \\ \frac{n_2}{n}(\bar{x}_2 - \mu_2(\boldsymbol{\theta})) \\ \vdots \\ \frac{n_L}{n}(\bar{x}_L - \mu_L(\boldsymbol{\theta})) \end{pmatrix}. \tag{2.6}$$

The efficient GMM estimator is given by

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\theta}[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n)]'\hat{\mathbf{S}}_n^{-1}[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n)], \tag{2.7}$$

where $\hat{\mathbf{S}}_n$ is the estimator of the asymptotic variance $\mathbf{S}$ of $\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n)$.

Unfortunately we cannot estimate $\mathbf{S}$ directry from the given data. Though Continuous-Updating Estimator (CUE) of Hansen, Heaton, and Yaron (1996) can be obtained, we will calculate $\hat{\mathbf{S}}$ through the computer simulation instead. We do not have individual data here and the estimate $\hat{\mathbf{S}}$ cannot be obtained in the conventional manner. The CUE is still valid under this circumstance because, as we will show later, the asymptotic variance can be written as the function of $\boldsymbol{\theta}$. In spite of the theoretical validity it has, though, it is computationally burdensome. We will explore the simulation method to earn the value because it is easier than the CUE.

The spirit underlying this method is that since we know the functional form of density function $f(x,\boldsymbol{\theta})$, if some value of $\boldsymbol{\theta}$ is given, we can generate random samples repeatedly. The estimation method is as follows.

1. Obtain preliminary estimator $\tilde{\boldsymbol{\theta}}_n$ by naive MLE or GMM using an identity matrix.
2. Generate random samples from $f(x,\tilde{\boldsymbol{\theta}}_n)$.
3. Estimate the weighting matrix $\tilde{\mathbf{S}}_n$ based on simulated data.

Simulation results suggest that this method works satisfactory well.

3 Large sample theory

We now consider large sample properties of the GMM estimator given by (2.7). We give a set of regularity conditions to help us doing asymptotic analysis. Although conditions given below are slightly stronger than necessary, they are satisfied for a large class of the distributions.

In the following, we use $\|\cdot\|$ to denote Euclidean norm.

Assumption 3.1
$\mathbf{S}$ *is positive semi-definite and for* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \mathbf{S}^{-1}E[\mathbf{g}(X;\boldsymbol{\theta})] \neq \mathbf{0}$.

Assumption 3.1 guarantees the identification of $\boldsymbol{\theta}_0$. Under the assumption, the objective function of GMM attains unique minimum at $\boldsymbol{\theta}_0$. Next assumption is useful to show consistency of the estimator.

Assumption 3.2
(i) The parameter space $\boldsymbol{\Theta}$ *is a compact subset of* $\mathbb{R}^p$.
(ii) $\frac{\partial}{\partial\boldsymbol{\theta}}\log P_i(\boldsymbol{\theta})$ *and* $\mu_i(\boldsymbol{\theta})$ *exist and are continuous at each* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *for* $i = 1,\ldots,L$.
(iii) $\max_{1\leq i\leq L}\left\{\int_{B_i}\sup_{\theta\in\Theta}|x-\mu_i(\boldsymbol{\theta})|f(x;\boldsymbol{\theta}_0)dx\right\} \leq M_1 < \infty$.

With these conditions, following result can be established.

Proposition 3.1
Suppose that $\hat{\mathbf{S}}_n \xrightarrow{p} \mathbf{S}$, *and Assumption 3.1 and 3.2 are satisfied. Then* $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$.

Next, we give the asymptotic distribution of the estimator.

Assumption 3.3
(i) $\boldsymbol{\theta}_0$ *is an interior point of* $\boldsymbol{\Theta}$.
(ii) $\frac{\partial}{\partial \boldsymbol{\theta}} \log P_i(\boldsymbol{\theta})$ *and* $\mu_i(\boldsymbol{\theta})$ *are continuously differentiable in a neighborhood* $\mathcal{N}$ *of* $\boldsymbol{\theta}_0$ *for* $i = 1, \ldots, L$.
(iii) $\int_{B_i} (x - \mu_i(\boldsymbol{\theta}_0))^2 f(x; \boldsymbol{\theta}_0) dx$ *is finte for* $i = 1, \ldots, L$.
(iv) $\sup_{\theta \in \mathcal{N}} \left\| \frac{\partial^2 \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \leq M_2 < \infty$ *and* $\max_{1 \leq i \leq L} \sup_{\theta \in \mathcal{N}} \left\| \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq M_3 < \infty$.
(v) $\mathbf{D}\mathbf{S}^{-1}\mathbf{D}'$ *is nonsingular for* $\mathbf{D} = E[\frac{\partial}{\partial \boldsymbol{\theta}'} \mathbf{g}(X, \boldsymbol{\theta})\big|_{\theta=\theta_0}]'$.

Assumption 3.3(iv) guarantees that for any sequence $\{\boldsymbol{\theta}_n^*\}$ satisfying $\boldsymbol{\theta}_n^* \xrightarrow{p} \boldsymbol{\theta}_0$, we have

$$\operatorname{plim} \left\{ \left. \frac{\partial \hat{\mathbf{g}}(\boldsymbol{\theta}; \mathcal{X}_n)}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_n^*} \right\} = \operatorname{plim} \left\{ \left. \frac{\partial \hat{\mathbf{g}}(\boldsymbol{\theta}; \mathcal{X}_n)}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right\} \equiv \mathbf{D}' \tag{3.1}$$

With these additional conditions, we get the following proposition.

Proposition 3.2
Let $\{\hat{\mathbf{S}}_n\}$ *be a sequence of positive definite matrices such that* $\hat{\mathbf{S}}_n \xrightarrow{p} \mathbf{S}$. *Then under Assumption 3.1-3.3, we have*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{V}),$$

where

$$\mathbf{V} = \left\{ \mathbf{D}_1 \mathbf{S}_{11}^{-1} \mathbf{D}_1' + \mathbf{D}_2 \mathbf{S}_{22}^{-1} \mathbf{D}_2' \right\}^{-1}$$

and

$$\mathbf{D}_1' = \sum_{i=1}^{L} P_i(\boldsymbol{\theta}_0) \left. \frac{\partial^2 \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0},$$

$$\mathbf{D}_2' = \begin{pmatrix} -\int_{B_1} \left. \frac{\partial \mu_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} f(x, \boldsymbol{\theta}_0) dx \\ -\int_{B_2} \left. \frac{\partial \mu_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} f(x, \boldsymbol{\theta}_0) dx \\ \vdots \\ -\int_{B_L} \left. \frac{\partial \mu_L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} f(x, \boldsymbol{\theta}_0) dx \end{pmatrix},$$

$$\mathbf{S}_{11} = \sum_{i=1}^{L} P_i(\boldsymbol{\theta}_0) \left\{ \left. \frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\theta=\theta_0} \right\} \left\{ \left. \frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right\},$$

$$\mathbf{S}_{22} = \begin{pmatrix} \int_{B_1} (x - \mu_1(\boldsymbol{\theta}_0))^2 f(x, \boldsymbol{\theta}_0) dx & & 0 \\ & \ddots & \\ 0 & & \int_{B_L} (x - \mu_L(\boldsymbol{\theta}_0))^2 f(x, \boldsymbol{\theta}_0) dx \end{pmatrix}.$$

Notice that $\left\{ \mathbf{D}_1 \mathbf{S}_{11}^{-1} \mathbf{D}_1' \right\}^{-1}$ is the asymptotic variance of the naive MLE. Proposition 3.2 shows that the asymptotic variance of the GMM is not larger than that of naive MLE, because $\mathbf{D}_2 \mathbf{S}_{22}^{-1} \mathbf{D}_2'$ is positive semi-definite matrix.

4 Specification testing

In this section, we study a specification test of the underlying distribution. Since our moment conditions include the likelihood equation, we can interpret the test of moment conditions as the specification test of the underlying distribution. In our estimation, the number of moment conditions always exceeds the number of parameters by L, which is the number of groups. Hence we can obtain the test of L over-identifying restrictions. By utilizing the Hansen's test of moment conditions, we can construct the test statistic. Let

$$\hat{Q}_n(\hat{\boldsymbol{\theta}}_n) = [\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}_n; \mathcal{X}_n)]' \hat{\mathbf{S}}_n^{-1} [\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}_n; \mathcal{X}_n)], \tag{4.1}$$

where $\{\hat{\mathbf{S}}_n\}$ is a consitent estimator of $\mathbf{S}$. If $\hat{\boldsymbol{\theta}}_n$ is the efficient estimator which is given by (2.7), we obtain the following null asymptotic distribution.

Proposition 4.1
Suppose that $\hat{\mathbf{S}}_n \xrightarrow{p} \mathbf{S}$, and Assumption 3.1-3.3 are satisfied. Then

$$n\hat{Q}_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{L} \chi^2(L),$$

where L is the number of groups.

The proof follows directly from Hansen (1982).

Table 1: Normal Distribution

	$n = 100$			$n = 1000$		
	True MLE	Naive MLE	GMM	True MLE	Naive MLE	GMM
			Standard Error			
μ	0.2967	0.3120	0.3108	0.0955	0.0995	0.0959
σ	0.2109	0.2833	0.2441	0.0684	0.0914	0.0684
			Ratio SE			
μ	—	1.05	1.05	—	1.04	1.00
σ	—	1.34	1.16	—	1.34	1.00

Empirical standard error of the 3000 parameter estimates.

5 Monte Carlo results

In this section we report the results of Monte Carlo experiments to examine the performance of our estimator and test statistic. We study the behavior of the naive MLE (solution of (1.1)), the true MLE (which is computed based on non-grouped observations), and our GMM estimator using the local mean. We use the true MLE as the efficiency bench mark. GMM estimator is calculated using the two-step method explained in Section 2. We employ the naive MLE as a preliminary estimator.

The simulation design is as follows. We consider estimators of the parameters of normal and lognormal distributions. The number of replications is fixed to 3000. We report empirical standard error for each parameter. We also calculate the ratio of the standard error (to be refferd to as Ratio SE) of an estimator relative to that of true MLE to measure the loss of efficiency caused by grouping. The simulations are carried out using GAUSS.

First we estimate a normal distritbution $N(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 3$. We use a small sample size $n = 100$ and a large sample size $n = 1000$. The sample is grouped into five groups with fixed endpoints $(-\infty, -3, -1, 1, 3, \infty)$. The results are summarized in Table 1. From Table 1 we see that GMM estimator dominates the naive MLE in both small and large samples especially when estimating σ. Since local mean contains the information of the location, it can improve the estimate of variance parameter. Table 1 also shows that although Ratio SE of naive MLE is not improved even in large sample, GMM performs on par with true MLE when sample size is large.

Table 2 reports the results of a lognormal distribution $L_N(\mu, \sigma)$ with $\mu = 1$ and $\sigma = 1$. The number of observation is 200. The endpoints are $(0, 3, 6, 9, \infty)$. In this experiment, we also report the GMM using both first and second moments (GMM II). Table 2 shows that when intervals are rather large, naive MLE performs very poorly. On the other hand, GMM works considerably well. This is a preferable feature because it is often the case in practice that the length of intervals are fixed and relatively large. Adding second moment can improve the efficiency of $\hat{\sigma}$, though it does not necessarily reduce the standard error of $\hat{\mu}$. In

Table 2: Lognormal Distribution

	True MLE	Naive MLE	GMM I	GMM II
		Standard Error		
μ	0.0721	0.0928	0.0736	0.0742
σ	0.0505	0.0869	0.0570	0.0540
		Ratio SE		
μ	—	1.29	1.02	1.03
σ	—	1.72	1.13	1.07

Empirical standard error of the 3000 parameter estimates.
GMM I refers to the GMM estimator using local mean.

Table 3: Size and Power

	$n = 100$	$n = 500$	$n = 1000$
size	0.042	0.049	0.054
power	0.202	0.897	0.995

Empirical size and power of the 3000 iterations.
The nominal level of the test is 0.05.

sammary, our GMM performs reasonably well for various distributions even in modest sample size.

We also investigate the performance of our test statistic. Table 3 reports simulation results of the specification test. We calculate the empirical size and power of the test statistic. The null hypothesis is that X follows normal distribution. The simulation design is almost same as that of estimation case. Under the null hypothesis X follows $N(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 3$. The sample is grouped into five groups with fixed endpoints $(-\infty, -3, -1, 1, 3, \infty)$. The nominal level of the test is 0.05. The emprical power is calculated under the alternative hypothesis that X follows t distribution with 5 digrees of freedom. The variance of t random variables is adjusted to be 9. All results are based on 3000 iterations with 100, 500, and 1000 observations. Table 3 shows that the size of our test is quite close to the nominal level even in small sample size. The power of our test is also good when the sample size is modest.

6 Conclusion

We applied GMM principle to grouped data analysis when we can observe not only count but also local moments for each group. We carry out Monte Carlo experiments to investigate the performance of our estimator and test. Simulation results show that our estimator performs remarkably well even in finite sample, local moments providing great efficiency improvement. We also show that the specification test statistic performs well with good empirical size. Our results suggest that local moments in grouped data could be highly informative if proper methods of grouping and summarizing were chosen.

A Appendix

Here we give brief proofs for the results in Section 3. Our proofs are mainly based on those of Newey and McFadden (1994).

Proof of Proposition 3.1 Let $\hat{Q}_n(\boldsymbol{\theta})$ denote the objective function of the GMM: $[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n)]'\hat{\mathbf{S}}_n^{-1}[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n)]$, and define

$$Q_0(\boldsymbol{\theta}) \equiv E[\mathbf{g}(X,\boldsymbol{\theta})]'\mathbf{S}^{-1}E[\mathbf{g}(X,\boldsymbol{\theta})].$$

To prove consistency we need to establish the following conditions (see Newey and McFadden (1994, Section 2.5)).

(a) $Q_0(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$.

(b) $\boldsymbol{\Theta}$ is compact.

(c) $Q_0(\boldsymbol{\theta})$ is continuous.

(d) $\hat{Q}_n(\boldsymbol{\theta})$ converges to $Q_0(\boldsymbol{\theta})$ in probability uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Condition (a) follows from Assumption 3.1. Condition (b) and (c) hold by Assumption 3.2(i) and (ii). Finally, we need to establish (d). By the triangle and Cauchy-Schwartz inequalities, we have

$$\begin{aligned}
\left|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})\right| \leq\; & \left|[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n) - E[\mathbf{g}(X,\boldsymbol{\theta})]]'\,\hat{\mathbf{S}}_n^{-1}\,[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n) - E[\mathbf{g}(X,\boldsymbol{\theta})]]\right| \\
& + \left|E[\mathbf{g}(X,\boldsymbol{\theta})]'(\hat{\mathbf{S}}_n^{-1} + \hat{\mathbf{S}}_n^{-1})[\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n) - E[\mathbf{g}(X,\boldsymbol{\theta})]]\right| \\
& + \left|E[\mathbf{g}(X,\boldsymbol{\theta})](\hat{\mathbf{S}}_n^{-1} - \mathbf{S}^{-1})E[\mathbf{g}(X,\boldsymbol{\theta})]\right| \\
\leq\; & \|\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n) - E[\mathbf{g}(X,\boldsymbol{\theta})]\|^2\,\|\hat{\mathbf{S}}_n^{-1}\| \\
& + 2\,\|E[\mathbf{g}(X,\boldsymbol{\theta})]\|\,\|\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n) - E[\mathbf{g}(X,\boldsymbol{\theta})]\|\,\|\hat{\mathbf{S}}_n^{-1}\| \\
& + \|E[\mathbf{g}(X,\boldsymbol{\theta})]\|^2\|\hat{\mathbf{S}}_n^{-1} - \mathbf{S}^{-1}\|.
\end{aligned}$$

Therefore, to show $\sup_{\theta\in\Theta}|\hat{Q}_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \xrightarrow{p} 0$, we need to establish $\sup_{\theta\in\Theta}\|\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n) - E[\mathbf{g}(X,\boldsymbol{\theta})]\| \xrightarrow{p} 0$. Since Assumption 3.2(i) and (ii) imply that $\left\|\frac{\partial}{\partial\theta}\log P_i(\boldsymbol{\theta})\right\|$ is bounded on $\boldsymbol{\Theta}$, under Assumption 3.2(i), (ii) and (iii), we obtain

$$\begin{aligned} E\left[\sup_{\theta\in\Theta}\|\mathbf{g}(X,\boldsymbol{\theta})\|\right] &\leq E\left[\sup_{\theta\in\Theta}\|\mathbf{g}_1(X,\boldsymbol{\theta})\|\right]+E\left[\sup_{\theta\in\Theta}\|\mathbf{g}_2(X,\boldsymbol{\theta})\|\right] \\ &\leq \sum_{i=1}^{L} P_i(\boldsymbol{\theta}_0)\sup_{\theta\in\Theta}\left\|\frac{\partial \log P_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \\ &\quad +\sum_{i=1}^{L}\int_{B_i}\sup_{\theta\in\Theta}|x-\mu_i(\boldsymbol{\theta})|f(x,\boldsymbol{\theta}_0)dx \\ &< \infty. \end{aligned}$$

This implies $\sup_{\theta\in\Theta}\|\hat{\mathbf{g}}(\boldsymbol{\theta};\mathcal{X}_n)-E[\mathbf{g}(X,\boldsymbol{\theta})]\|\xrightarrow{p}0$ by uniform law of large numbers (see, for example, Lemma 2.4 of Newey and McFadden (1994)), and the desired result follows. □

Proof of Proposition 3.2 Using mean value theorem around $\boldsymbol{\theta}_0$, we obtain

$$\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}_n;\mathcal{X}_n)=\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}_0;\mathcal{X}_n)+\left\{\left.\frac{\partial\hat{\mathbf{g}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\theta=\theta_n^*}\right\}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0), \tag{A.1}$$

for some $\boldsymbol{\theta}_n^*$ between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. By Assumption 3.3(i), (ii), and (iii), the first-order condition:

$$2\left\{\left.\frac{\partial\hat{\mathbf{g}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\theta=\hat{\theta}_n}\right\}'\hat{\mathbf{S}}_n^{-1}\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}_n;\mathcal{X}_n)=\mathbf{0} \tag{A.2}$$

is satisfied. Combining (A.1) and (A.2), we have

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0) &= -\left[\left\{\left.\frac{\partial\hat{\mathbf{g}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\theta=\hat{\theta}_n}\right\}'\hat{\mathbf{S}}_n^{-1}\left\{\left.\frac{\partial\hat{\mathbf{g}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\theta=\theta_n^*}\right\}\right]^{-1} \\ &\quad\times\left\{\left.\frac{\partial\hat{\mathbf{g}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right|_{\theta=\hat{\theta}_n}\right\}'\hat{\mathbf{S}}_n^{-1}\sqrt{n}\hat{\mathbf{g}}(\boldsymbol{\theta}_0;\mathcal{X}_n). \end{aligned} \tag{A.3}$$

By Assumption 3.3(ii) and (iv),

$$\begin{aligned} E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial\mathbf{g}(X,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\|\right] &\leq E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial\mathbf{g}_1(X,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\|\right]+E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial\mathbf{g}_2(X,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\|\right] \\ &\leq \sum_{i=1}^{L}P_i(\boldsymbol{\theta}_0)\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial^2\log P_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|+\sum_{i=1}^{L}P_i(\boldsymbol{\theta}_0)\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial\mu_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\| \\ &< \infty, \end{aligned}$$

which implies

$$\text{plim}\left\{ \left. \frac{\partial \hat{\mathbf{g}}_1(\boldsymbol{\theta};\mathcal{X}_n)}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_n^*} \right\} = \text{plim}\left\{ \left. \frac{\partial \hat{\mathbf{g}}_1(\boldsymbol{\theta};\mathcal{X}_n)}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right\} \equiv \mathbf{D}_1' \quad \text{(A.4)}$$

$$\text{plim}\left\{ \left. \frac{\partial \hat{\mathbf{g}}_2(\boldsymbol{\theta};\mathcal{X}_n)}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_n^*} \right\} = \text{plim}\left\{ \left. \frac{\partial \hat{\mathbf{g}}_2(\boldsymbol{\theta};\mathcal{X}_n)}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} \right\} \equiv \mathbf{D}_2' \quad \text{(A.5)}$$

where

$$\mathbf{D}_1' = \sum_{i=1}^{L} P_i(\boldsymbol{\theta}_0) \left. \frac{\partial^2 \log P_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0}$$

$$\mathbf{D}_2' = \begin{pmatrix} -\int_{B_1} \left. \frac{\partial \mu_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} f(x,\boldsymbol{\theta}_0)dx \\ -\int_{B_2} \left. \frac{\partial \mu_2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} f(x,\boldsymbol{\theta}_0)dx \\ \vdots \\ -\int_{B_L} \left. \frac{\partial \mu_L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\theta=\theta_0} f(x,\boldsymbol{\theta}_0)dx \end{pmatrix}$$

(see, for example, Theorem 3.1 of Newey and McFadden (1994)).

Next, using Lindeberg-Levy central limit theorem, we have

$$\sqrt{n}\begin{pmatrix} \hat{\mathbf{g}}_1(\hat{\boldsymbol{\theta}}_0;\mathcal{X}_n) \\ \hat{\mathbf{g}}_2(\hat{\boldsymbol{\theta}}_0;\mathcal{X}_n) \end{pmatrix} \xrightarrow{L} N\left(\mathbf{0}, \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}\right), \quad \text{(A.6)}$$

where the row k, column l element of $\mathbf{S}_{mn}$ (denote s_{kl}^{mn}) is given by

$$\begin{aligned} s_{kl}^{11} &= E\left[\sum_{i=1}^{L} I(x \in B_i) \left.\frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \theta_k}\right|_{\theta=\theta_0} \left.\frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \theta_l}\right|_{\theta=\theta_0}\right] \\ &= \sum_{i=1}^{L} P_i(\boldsymbol{\theta}_0) \left.\frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \theta_k}\right|_{\theta=\theta_0} \left.\frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \theta_l}\right|_{\theta=\theta_0}, \\ s_{kl}^{12} &= E\left[\left\{\sum_{i=1}^{L} I(X \in B_i) \left.\frac{\partial \log P_i(\boldsymbol{\theta})}{\partial \theta_k}\right|_{\theta=\theta_0}\right\} \{I(X \in B_l)(X - \mu_l(\boldsymbol{\theta}_0))\}\right] \\ &= \left.\frac{\partial \log P_i(\boldsymbol{\theta}))}{\partial \theta_k}\right|_{\theta=\theta_0} E\left[I(X \in B_l)(X - \mu_l(\boldsymbol{\theta}_0))\right] \\ &= 0, \\ s_{kl}^{22} &= E\left[I(X \in B_k) I(X \in B_l)(X - \mu_k(\boldsymbol{\theta}_0))(X - \mu_l(\boldsymbol{\theta}_0))\right] \\ &= \begin{cases} \int_{B_k} (x - \mu_k(\boldsymbol{\theta}_0))^2 f(x,\boldsymbol{\theta}_0)dx & \text{for } k = l \\ 0 & \text{for } k \neq l. \end{cases} \end{aligned}$$

Under Assumption 3.3(v), substituting (A.4), (A.5), and (A.6) into (A.3), we get

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{V}),$$

where

$$\mathbf{V} = \left\{ \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 \end{pmatrix} \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{D}_1' \\ \mathbf{D}_2' \end{pmatrix} \right\}^{-1}.$$

However, since $\mathbf{S}_{12} = \mathbf{S}_{21}' = \mathbf{0}$, it can be reduced to

$$\mathbf{V} = \left\{ \mathbf{D}_1 \mathbf{S}_{11}^{-1} \mathbf{D}_1' + \mathbf{D}_2 \mathbf{S}_{22}^{-1} \mathbf{D}_2 \right\}^{-1},$$

and the proposition is proved. □

References

CRESSIE, N., AND T. R. C. READ (1984): "Multinomial Godness-of-fit Tests," *Journal of the Royal Statistical Society, Ser. B*, 46, 440–464.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimator," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Bussiness and Economic Statistics*, 14, 262–280.

NEWEY, W. K., AND D. L. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, pp. 2111–2245. Amsterdam: Elsevier.

TALLIS, G. M. (1967): "Approximate Maximum Likelihood Estimates From Grouped Data," *Technometrics*, 9, 599–606.

VICTORIA-FESER, M., AND E. RONCHETTI (1997): "Robust Estimation for Grouped Data," *Journal of the American Statistical Association*, 92, 333–340.