Testing Parameter Constancy Across Many Groups

Shinichi Sakata*
University of British Columbia
Department of Economics
997-1873 East Mall
Vancouver, BC V6T 1Z1, *Canada*
Email: ssakata@interchange.ubc.ca

September 17, 2008

Abstract

The dependence of an individual's behavior on his affiliation with a group is often suspected. Given a parametric model, the dependence of the behavior on the group affiliation means the constancy of the model parameters across groups, which is often tested by the Wald test (the Chow test), the Lagrange multiplier (LM) test, or the generalized likelihood ratio (GLR) test. In this paper, we consider the familiar problem of parameter constancy testing in the setup in which the population consists of many groups, each of which offers only a few observations. The parameter constancy test in such problem set up has not been studied at our best knowledge.

In our problem setup, the parameters cannot be accurately estimated for each group, due to the small sample size of the group. This rules out use of the Wald and GLR test. In the LM test, the Lagrange multiplier vector is a large vector consisting of the average scores taken over each group. The small group sample sizes make the normal approximation to the distribution of the Lagrange multiplier vector unreliable and the weighting matrix in the LM test statistic very unstable. Thus, our problem setup requires an approach different from the familiar large sample tests.

*The author acknowledges support from the SSHRC Standard Research Grant Program.

We assume that the researcher has a well-thought-out model in hand, which is supposed to capture individuals' behavior in all groups with a parameter value common to all groups. Without having any expectation on the way how the parameter constancy could be violated, he desires to test the hypothesis that the parameter of interest stays the same across the groups against the alternative hypothesis that the parameter of interest is unequal between at least one pair of groups.

This paper develops a novel test of parameter constancy that works in the above-mentioned setup and studies its statistical properties in detail. Out test uses a statistic obtained by taking a weighted average of an *unbiasedly* estimated quadratic form of the group mean score, evaluated with the parameter estimate obtained imposing the parameter constancy. Once suitably standardized, the test statistic is approximately distributed with the standard normal distribution under the null hypothesis. Our test rejects the null hypothesis when the standardized statistic exceeds the critical value determined based on the standard normal approximation. The proposed test only requires that an M-estimator accurately estimate the model parameters under the null. The estimation of the parameter for each group is unnecessary. The sizes of group samples can be as small as two for the proposed method to work. To illustrate use of the proposed test, we apply the proposed method to test constancy of the wage regression function across workplaces, using the Workplace and Employee Survey from Statistics Canada.

Key Words: Score test; M-estimation; Workplace and Employee Survey

1. INTRODUCTION

The dependence of an individual's behavior on his affiliation with a group is often suspected. In a typical statistical analysis, an individual's behavior is captured by the parameters in a statistical model. If the individual's affiliation does matter to his behavior, the parameters should vary from one group to another. This paper proposes a method to test the constancy of the parameters across groups.

It has been been extensively studied how to test parameter constancy in various setups. Most existing papers on this topic assume that many observations are available for each of the groups. The large group sample sizes are essential for the Wald and generalized likelihood ratio (GLR) tests, because they need an accurate estimate of the parameter for each group. In the Lagrange multiplier (LM) test of the parameter constancy, the Lagrange multiplier vector is a large vector consisting of the average scores taken over each group. The large group sample sizes are again a key requirement. If the group sample sizes are small, the normal approximation to the distribution of the Lagrange multiplier vector is unreliable. Even worse, the estimated autocovariance matrix of the Lagrange multiplier vector, whose inverse is the weighting matrix in the LM test statistic, is nearly or exactly singular.

Our paper, on the other hand, focuses on the situation in which the constancy of parameters across *many groups* is questioned, and only a *small number of observations from each group* (as few as two observations) are available. A situation with many groups and small group sample sizes can arise in two scenarios. In the first scenario, a researcher models the behavior of individuals in the population, given a data set. The researcher then realizes that affiliation with certain groups that form a fine partition of the population may matter for the behavior of the individuals. For example, the researcher may be worried that a worker's behavior may be different from one workplace to another, or alternatively, a worker's behavior may be affected by which high school he attended and when he graduated. In either way, the population is split into many groups based on such factors, and each group would have a small sample size. It would be useful for the researcher to have a test of parameter constancy suitable in this case, because it allows him to check if he really needs to extend the model to incorporate the possible effects of the groups before spending much of his precious time.

In the second scenario, a researcher has a data set in hand, which contains a decent number of observations as a whole but contains only a small number of observations per group. For instance, the data set may contain the information on individuals for studying what affects propagation of human immunodeficiency virus. Trying to estimate a given model satisfactorily for each group formed based on the gender, race, age, area of residence, etc., a researcher may suspect that the parameters of interest could vary across the groups. If the parameters are constant across groups, he could estimate the model satisfactorily with the currently available data set; otherwise, he would need to spend his scarce financial resources to collect more observations per group. In this scenario, it would be again useful to have a suitable test of the parameter constancy.

We assume that the researcher has a well-thought-out model in hand, which is supposed to capture individuals' behavior in all groups with a parameter value common to all groups. Without having any expectation on the way how the parameter constancy could be violated, he desires to test the hypothesis that the parameter of interest stays the same across the groups against the alternative hypothesis that the parameter of interest is unequal between at least one pair of groups.

This paper develops a novel test of parameter constancy that works in the above-mentioned setup and studies its statistical properties in detail. The proposed test uses a statistic obtained by taking a weighted average of an *unbiasedly* estimated quadratic form of the group mean score, evaluated with the parameter estimate obtained imposing the parameter constancy. Once suitably standardized, the test statistic is approximately distributed with the standard normal distribution under the null hypothesis. Our test rejects the null hypothesis when the standardized statistic exceeds the critical value determined based on the standard normal approximation. The proposed test only requires that an M-estimator accurately estimate the model parameters under the null. The estimation of the parameter for each group is unnecessary. *The sizes of group samples can be as small as two for the proposed method to work.*

The general theme of this paper is not new, as mentioned above. Chow (1960) proposes a test of the equality of the set of coefficients of a classical linear regression model between two groups, imposing normality and homoskedasticity on the error terms. Fisher (1970) gives some discussion on the technical aspect of the test of Chow (1960). The test is now called the Chow test among economists. Also, Kullback and Rosenblatt (1957) consider how to test the equality of linear regression parameters across multiple groups. Toyoda (1974) studies how the Chow test is affected if the error term has different variances between the two groups. Weerahandi (1987) provides a remedy to the problem in the Chow test caused by unequal variances of the error term between the two groups.

Dufour (1982) considers an interesting generalization of the Chow test. Pointing out that the group sample sizes can be smaller than the number of the regressors in some applications, he demonstrates how the Chow test could be modified to handle the case with small group sample sizes. The problem setup of Dufour (1982) is similar to ours in this sense. Nonetheless, Dufour's analysis heavily depends on the normality of the error term. Also, it does not seem to extend to any models beyond the linear regression model. Our approach, on the other hand, readily covers a wide range of situations and does not rely on stringent assumptions such as the normality of the error distribution. The generality of our method, however, does not come free. First, our approach is only justified by the large sample theory, unlike Dufour's method whose finite sample property is known under the assumed conditions. Second, our approach does require that the number of groups be large, unlike Dufour's approach.

Pesaran and Yamagata (2008) is another work related to the current paper. It considers a test of parameter constancy across individuals in the panel analysis. If we label groups as "individuals" and individuals in each group as "time periods for the individual", the model considered in our analysis looks like a panel model. In this view, Pesaran and Yamagata (2008) and the current paper are considering similar problems. Nevertheless, the translated version of our problem assumes a short panel (with a small number of time periods per individual), while the method proposed by Pesaran and Yamagata (2008) treats long panels, even though they focus on the situation with small numbers of periods per individual relative to the number of individuals. The long panel setup allows them to estimate the parameters of interest for each individual (each group in our problem) separately with decent accuracy, while estimation of the parameters per individual (i.e., per group) is out of question in our problem setup. Thus, their approach is inapplicable in the problem considered in the current paper, while our approach might not be effective in their problem setup.

The rest of the paper is organized as follows. In Section 2, we precisely define the null hypothesis of interest and show an implication of the null hypothesis on the group mean scores. In Section 3, we propose a method to test the null hypothesis and derive its large sample properties. In Section 4, we conduct Monte Carlo simulations to assess the finite sample performance of the proposed test. We also provide an example to illustrative use of the proposed method in Section 5. We then conclude the paper with remarks in Section 6. The proofs of theorems are collected in the Appendices.

We employ the following convention and symbols throughout this paper. Limits are taken along the sequence of numbers of groups (denoted G) growing to infinity, unless otherwise indicated. For each matrix A, $|A|$ denotes the Frobenius norm of A, i.e., $|A| \equiv \sqrt{\mathrm{tr}(A'A)}$, and A^+ the Moore-Penrose (MP) inverse of A. By applying the MP inverse in division by scalars, we rule that division by zero equals zero. We use the MP inverses of random square matrices instead of the regular inverses to avoid technical problems caused by the singularity of the random matrices that occurs with a low probability. The reader can safely replace the MP inverses with the regular inverses, when applying the formulas in this paper in practice.

2. NULL HYPOTHESIS

Suppose that there are G groups in the population under study. We have a random sample from each of the G groups. The size of the sample from group g is denoted n_g. The size of the whole sample is thus $N_G \equiv \sum_{g=1}^{G} n_g$. The ith observation from group g is denoted X_{gi}, which is $v \times 1$. For each group g, the parameter of interest $\theta_g^\dagger$ is characterized as a maximizer of $\bar{q}_g(\theta) \equiv \mathrm{E}[q_g(X_{g1}, \theta)]$ with respect to θ over a parameter space Θ, where q_g is a known function.

Example 2.1 Partition X_{gi} as $X_{gi} = (Y_{gi}, Z'_{gi})'$, where Y_{gi} is a random variable, and Z_{gi} is a $(v-1) \times 1$ random vector ($v \geq 2$). Assume that

$$Y_{gi} = Z'_{gi}\theta^\dagger_g + U_{gi}, \quad i = 1, 2, \ldots, n_g,\ g = 1, 2, \ldots, G,$$

where $\theta^\dagger_g \in \Theta = \mathbb{R}^{v-1}$; and $\mathrm{E}[U_{g1}|Z_{g1}] = 0$ (the conditional mean restriction) or $\mathrm{E}[Z_{g1}U_{g1}] = 0$ (exogeneity). Then $\theta^\dagger_g$ is a maximizer of $\bar{q}_g(\theta) \equiv \mathrm{E}[q_g(X_g, \theta)]$, where for each $x = (y, z')' \in \mathbb{R} \times \mathbb{R}^{v-1}$, each $\theta \in \Theta$, and each $g \in \{1, 2, \ldots, G\}$,

$$q_g(x, \theta) \equiv -(y - z'\theta)^2.$$

Example 2.2 The random variable X_{gi} is partitioned in the same way as in Example 2.1, and Y_{gi} is a binary random variable. Assume that

$$P[Y_{gi} = 1 \,|\, Z_{gi}] = \Phi(Z'_{gi}\theta^\dagger_g), \quad i = 1, 2, \ldots, n_g,\ g = 1, 2, \ldots, G,$$

where $\theta^\dagger_g \in \Theta \subset \mathbb{R}^{v-1}$, and Φ is the standard normal cdf. Then $\theta^\dagger_g$ is a maximizer of $\bar{q}_g(\theta) \equiv \mathrm{E}[q_g(X_{g1}, \theta)]$, where for each $x = (y, z')' \in \mathbb{R} \times \mathbb{R}^{v-1}$, each $\theta \in \Theta$, and each $g \in \{1, 2, \ldots, G\}$,

$$q_g(x, \theta) \equiv y \log \Phi(z'\theta) + (1 - y) \log(1 - \Phi(z'\theta)).$$

Example 2.3 The random variable X_{gi} is partitioned in the same way as in Example 2.1, and Y_{gi} is a count variable. Assume that

$$P[Y_{gi} = 1 \,|\, Z_{gi}] = \exp(Z'_{gi}\theta^\dagger_g), \quad i = 1, 2, \ldots, n_g,\ g = 1, 2, \ldots, G, \tag{1}$$

where $\theta^\dagger_g \in \Theta \subset \mathbb{R}^{v-1}$. Then $\theta^\dagger_g$ is a maximizer of $\bar{q}_g(\theta) \equiv \mathrm{E}[q_g(X_{g1}, \theta)]$, where for each

$x = (y, z')' \in \mathbb{R} \times \mathbb{R}^{v-1}$, each $\theta \in \Theta$, and each $g \in \{1, 2, \ldots, G\}$,

$$q_g(x, \theta) \equiv -\exp(z'\theta) + yz'\theta.$$

Here, q_g is obtained from the log-likelihood function of the Poisson regression model, dropping the term that does not depend on θ. The validity of this choice of q_g to characterize $\theta_g^\dagger$ only requires (1). It holds even if Y_{gi} is not conditionally distributed with the Poisson distribution given Z_{gi}. See Gouriéroux et al. (1984).

The null hypothesis we attempt to test is that H_0: $\theta_1^\dagger = \theta_2^\dagger = \cdots = \theta_G^\dagger$. Because we are considering the situation in which a researcher has a well-thought-out model in hand, the researcher should not expect that H_0 can be violated in any particular manner (if he did, he would have already incorporated it into the model). The alternative hypothesis is therefore the violation of the equalities in H_0 *in any form*, i.e., at least between one pair of groups, g_1 and g_2, $\theta_{g_1}^\dagger \neq \theta_{g_2}^\dagger$.

We consider how to test H_0 in the situation with a large G and small n_g's. Our asymptotic analysis incorporates this feature via the large-G asymptotics, in which the sequence $\{n_g\}_{g\in\mathbb{N}}$ is bounded. As is the case in most asymptotic analysis, our asymptotics is a fiction used for approximating various finite sample phenomena. Given the likely heterogeneity across groups, however, we do not have a natural and simple mechanism to create more and more groups together with their population distributions. In this sense, our large-G asymptotics is more fictional than the familiar ones. Some of the assumptions made in this paper are not verifiable for this reason.

The basic setup for our large sample analysis is formalized in the following assumption.

Assumption 1 (a) *The sequence $\{n_g \in \mathbb{N}\}_{g\in\mathbb{N}}$ satisfies that for each $g \in \mathbb{N}$, $2 \leq n_g \leq \bar{n}$, where $\bar{n}$ is a natural number. The data are a realization of an independent, rowwise identically distributed double array of $v \times 1$ random vectors $\{X_{gi} : i \in \mathbb{I}_g, g \in \mathbb{N}\}$ on a probability space $(\Omega, \mathscr{F}, P)$, where $\mathbb{I}_g \equiv \{i \in \mathbb{N} : i \leq n_g\}$ for each $g \in \mathbb{N}$.*

(b) *Θ is a subset of $\mathbb{R}^p$, and $\{q_g : \mathbb{R}^v \times \Theta \to \mathbb{R}\}_{g \in \mathbb{N}}$ is a sequence of functions measurable-$(\mathscr{B}^v \otimes \mathscr{B}(\Theta))/\mathscr{B}$ such that for each $g \in \mathbb{N}$ and each $x \in \mathbb{R}^v$, $q_g(x, \cdot) : \Theta \to \mathbb{R}$ is twice continuously differentiable.*

(c) *For each $g \in \mathbb{N}$ and each $\theta \in \Theta$,*

$$\mathrm{E}[|q_g(X_{g1}, \theta)|] + \mathrm{E}[|\nabla q_g(X_{g1}, \theta)|] < \infty,$$

and

$$\nabla \mathrm{E}[q_g(X_{g1}, \theta)] = \mathrm{E}[\nabla q_g(X_{g1}, \theta)],$$

where ∇ denotes the gradient operator with respect to the parameters.

(d) *For almost all $G \in \mathbb{N}$, $\bar{Q}_G : \Theta \to \mathbb{R}$ defined by*

$$\begin{aligned} \bar{Q}_G(\theta) &\equiv \mathrm{E}\Big[N_G^{-1} \sum_{g=1}^{G} \sum_{i=1}^{n_g} q_g(X_{gi}, \theta)\Big] \\ &= N_G^{-1} \sum_{g=1}^{G} n_g \bar{q}_g(\theta), \quad \theta \in \Theta,\ G \in \mathbb{N} \end{aligned} \tag{2}$$

has a unique maximizer $\theta_G^* \in \Theta$, where $N_G \equiv \sum_{g=1}^{G} n_g$, $G \in \mathbb{N}$.

The requirement in Assumption 1(a) that $\{X_{gi}\}$ be rowwise identically distributed is satisfied if observations are drawn from each group by simple random sampling. The differentiability imposed in Assumption 1(b) is mild, though it clearly rules out the quantile regression analysis Koenker and Bassett (1978), in which q_g would be a "check function" of the regression residual. The moment conditions imposed in Assumption 1(c) are weaker than those typically employed in proving the asymptotic normality of the M-estimator defined to be a maximizer of the average of $q_g(X_{gi}, \theta)$ taken over the entire sample. The interchangeability of the gradient operator and the expectation operator is also innocuous.

Under H_0, $\bar{q}_g$ is maximized at the same parameter value, and so is $\bar{Q}_G$ defined in (2). It follows that for each $G \in \mathbb{N}$, $\theta_G^* = \theta_1^\dagger = \cdots = \theta_G^\dagger$ under H_0. Assumption 1(d) requires that the uniqueness of θ_G^* also hold under the alternative. Under suitable additional assumptions, the pseudo-true parameter θ_G^* can be consistently estimated by the M-estimator that maximizes the average of $q_g(X_{gi}, \theta)$ taken over the entire sample, which appears in the square brackets in (2), while $\theta_g^\dagger$ is not estimable for each $g \in \mathbb{N}$, given the boundedness of $\{n_g\}$. Note that we do not require that the maximizer of $\bar{q}_g$ be unique for each $g \in \mathbb{N}$. This allows for situations in which $\theta_g^\dagger$ is not unique for some groups, due to some variables taking on the same values over a group (e.g., an indicator variable for a region, to which some groups belong).

Because $\bar{q}_g$ is maximized at θ_G^* for each group $g \in \mathbb{N}$ under H_0, we have that under H_0, $\mathrm{E}[\nabla q_g(X_{g1}, \theta_G^*)] = 0$ for each $(g, G) \in \mathbb{G} \equiv \{(g, G) \in \mathbb{N}^2 : g \leq G\}$, by the first order condition for the maximization of $\bar{q}_g$. On the other hand, $\mathrm{E}[\nabla q_g(X_{q1}, \theta_G^*)]$ is likely to be nonzero, when H_0 does not hold, though it is not impossible for it to be zero, as θ_G^* can happen to be a local maximum or a saddle point of $\bar{q}_g$.

We construct a test of H_0 based on whether or not the data indicates that $\mathrm{E}[\nabla q_g(X_{g1}, \theta_G^*)] \neq 0$. We allow the test to focus on a subvector of $\mathrm{E}[\nabla q_g(X_{g1}, \theta_G^*)]$, if the user desires so. Without loss of generality, we assume that the user chooses the first m elements $(0 < m \leq p)$. Define $\{s_{gi} : \Omega \times \Theta \to \mathbb{R}^m : i \in \mathbb{I}_g,\ g \in \mathbb{N}\}$ by

$$s_{gi}(\cdot, \theta) \equiv \nabla_{\theta^1} q_g(X_{gi}, \theta), \quad \theta \in \Theta,\ i \in \mathbb{I}_g,\ g \in \mathbb{N},$$

where ∇_{θ^1} denote the operator of the partial differentiation with respect to θ^1, the first m components of θ. Also write

$$s_{Ggi}^* \equiv s_{gi}(\cdot, \theta_G^*), \quad i \in \mathbb{I}_g,\ (g, G) \in \mathbb{G}.$$

Then it holds that $\mathrm{E}[s_{Gg1}^*] = 0$ for each $(g, G) \in \mathbb{G}$ under H_0.

A challenge is that $\mathrm{E}[s^*_{Gg1}]$ cannot be estimated well, given that the size of each group sample is small. Its averaged version, $N_G^{-1}\sum_{g=1}^G n_g \mathrm{E}[s^*_{Gg1}]$, is also useless, because it is equal to zero, whether or not H_0 is true, and so is its sample counterpart.

3. FORMULATION OF A TEST

We now state a key fact used in our construction of a test of H_0. Let $\mathbb{S}^m$ denote the set of all $m \times m$ symmetric matrices. Assume:

Assumption 2 (a) *The sequence $\{W_G \in \mathbb{S}^m\}_{G\in\mathbb{N}}$ is bounded.*

(b) *The sequence $\{W_G\}_{G\in\mathbb{N}}$ is uniformly positive definite.*

Then for each $(g, G) \in \mathbb{G}$, $\mathrm{E}[s^*_{Gg1}]'W_G\mathrm{E}[s^*_{Gg1}] \geq 0$; and $\mathrm{E}[s^*_{Gg1}]'W_G\mathrm{E}[s^*_{Gg1}] = 0$ if and only if $\mathrm{E}[s^*_{Gg1}] = 0$. It follows that $\alpha^*_G \equiv \alpha_G(\theta^*_G, W_G)$ is nonnegative, where

$$\begin{aligned}\alpha_G(\theta, W) &\equiv N_G^{-1}\sum_{g=1}^G n_g \mathrm{E}[s_{g1}(\cdot,\theta)]'W\mathrm{E}[s_{g1}(\cdot,\theta)] \\ &= N_G^{-1}\sum_{g=1}^G n_g \mathrm{tr}(W\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]'), \quad \theta\in\Theta,\ W\in\mathbb{S}^m; \end{aligned} \tag{3}$$

and $\alpha^*_G = 0$ if and only if $\mathrm{E}[s^*_{Gg1}] = 0$. Thus, it holds that $\alpha^*_G = 0$ under H_0. We base our test on this simple fact. Namely, our test rejects H_0 if an estimate of α^*_G is positive and far from zero.

Our formulation of a test of H_0 requires consistent estimation of α^* (in the large-G asymptotics). We first consider how to estimate $\alpha_G(\theta, W)$ with fixed $\theta \in \Theta$ and $W \in \mathbb{S}^m$. Note that $\alpha_G(\theta, W)$ is a weighted sum of $\mathrm{tr}(W\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]')$ over the first G groups. If we can estimate $\mathrm{tr}(W\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]')$ unbiasedly for each $g \in \mathbb{N}$, plugging the unbiased estimator into $\mathrm{tr}(W\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]')$ in (3) generates an estimator that converges to $\alpha_G(\theta, W)$ in probability-P by the law of large numbers for independent but not identically distributed processes.

For the unbiased estimation of $\mathrm{tr}(W\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]')$, we use the fact that

$$\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]' = \mathrm{E}[s_{g1}(\cdot,\theta)s_{g1}(\cdot,\theta)'] - \mathrm{var}[s_{g1}(\cdot,\theta)], \quad (g,G) \in \mathbb{G}. \tag{4}$$

We know that the sample second moment and sample covariance matrix of $s_{g1}(\cdot,\theta)$ taken over the observations from the gth group unbiasedly estimate $\mathrm{E}[s_{g1}(\cdot,\theta)s_{g1}(\cdot,\theta)']$ and $\mathrm{var}[s_{g1}(\cdot,\theta)]$, respectively. Plugging them into the right-hand side of (4) yields an unbiased estimator of $\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]'$, with which we can also estimate $\mathrm{tr}(W\mathrm{E}[s_{g1}(\cdot,\theta)]\mathrm{E}[s_{g1}(\cdot,\theta)]')$ unbiasedly.

Define $\{\tilde{s}_g : \Omega \times \Theta \to \mathbb{R}^m\}_{g\in\mathbb{N}}$ by

$$\tilde{s}_g \equiv n_g^{-1}\sum_{i=1}^{n_g} s_{gi}, \quad g \in \mathbb{N}.$$

Also, define $\check{S}_G : \Omega \times \Theta \to \mathbb{R}^{m\times m}$ and $\check{\Sigma}_G : \Omega \times \Theta \to \mathbb{R}^{m\times m}$ by

$$\check{S}_G(\cdot,\theta) \equiv N_G^{-1}\sum_{g=1}^{G}\sum_{i=1}^{n_g} s_{gi}(\cdot,\theta)s_{gi}(\cdot,\theta)', \quad \theta \in \Theta,\, G \in \mathbb{N} \quad \text{and}$$

$$\begin{aligned}\check{\Sigma}_G(\cdot,\theta) \equiv N_G^{-1}\sum_{g=1}^{G}\Bigg(&\frac{n_g}{n_g-1}\\ &\times\sum_{i=1}^{n_g}(s_{gi}(\cdot,\theta)-\tilde{s}_g(\cdot,\theta))(s_{gi}(\cdot,\theta)-\tilde{s}_g(\cdot,\theta))'\Bigg), \qquad \theta \in \Theta,\, G \in \mathbb{N}.\end{aligned}$$

Then the consistent estimator $\check{\alpha}_G(\cdot,\theta,W) : \Omega \to \mathbb{R}$ of $\alpha_G(\theta,W)$ is

$$\check{\alpha}_G(\cdot,\theta,W) \equiv \mathrm{tr}(W_G(\check{S}_G(\cdot,\theta) - \check{\Sigma}_G(\cdot,\theta))), \quad \theta \in \Theta,\, G \in \mathbb{N}. \tag{5}$$

Given $\check{\alpha}_G(\cdot,\theta,W)$, it is natural to replace θ with an estimator $\hat{\theta}_G$ of θ^*_G and W with W_G to obtain an estimator of α^*_G. Also, one may want to use a data-dependent weighting matrix

$\hat{W}_G$ in place of W_G in the actual implementation of this estimation strategy. We assume:

Assumption 3 (a) *The sequence* $\{\hat{\theta}_G : \Omega \to \Theta\}_{G\in\mathbb{N}}$ *consists of* Θ*-valued random vectors on* $(\Omega, \mathscr{F}, P)$ *such that for each* $G \in \mathbb{N}$, $Q_G(\cdot, \hat{\theta}_G) = \sup_{\theta\in\Theta} Q_G(\cdot, \theta)$, *where* $Q_G : \Omega \times \Theta \to \mathbb{R}$ *is defined by*

$$Q_G(\cdot, \theta) \equiv N_G^{-1} \sum_{g=1}^{G} \sum_{i=1}^{n_g} q_g(X_{gi}, \theta), \quad \theta \in \Theta,\, G \in \mathbb{N}.$$

The sequence $\{|\hat{\theta}_G - \theta^*_G|\}_{G\in\mathbb{N}}$ *converges to zero in probability-*P.

(b) *The sequence of random matrices,* $\{\hat{W}_G : \Omega \to \mathbb{S}^m\}_{G\in\mathbb{N}}$, *satisfies that* $\{|\hat{W}_G - W_G|\}_{G\in\mathbb{N}}$ *converges to zero in probability-*P.

The consistency of $\{\hat{\theta}_G\}$ for $\{\theta^*_G\}$ imposed in Assumption 3(a) is a high-level assumption. We employ it in our analysis to keep the results of the current paper widely applicable. The required consistency is usually not difficult to verify. The framework described in White (1994, Chapter 3), for example, can handle many cases including the setup of Example 2.2. Nevertheless, such framework is not universally applicable. The compactness on the parameter space imposed in the framework rules out the setup of Example 2.1 with $\Theta = \mathbb{R}^{v-1}$, despite that Assumption 3 holds in Example 2.1 if $\{X_{g1}\}_{g\in\mathbb{N}}$ is uniformly $\mathscr{L}_{2+\delta}$-bounded for some real number $\delta > 0$, and $\{N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[X_{g1}X'_{g1}]\}$ is asymptotically uniformly positive definite. Assumption 3(a) let us avoid binding our results to a single set of assumptions employed in the consistency proof.

The convergence of $\{\breve{\alpha}_G(\cdot, \theta, W) - \alpha_G(\theta, W)\}_{G\in\mathbb{N}}$ in probability to zero for each $(\theta, W) \in \Theta \times \mathbb{S}^m$ discussed above can be extended to the convergence uniform in (θ, W) in each compact subset of $\Theta \times \mathbb{S}^m$ under the assumptions imposed below. Our estimator $\{\hat{\alpha}_G \equiv \breve{\alpha}_G(\cdot, \hat{\theta}_G, \hat{W}_G)\}_{G\in\mathbb{N}}$ is thus consistent for $\{\alpha^*_G\}_{G\in\mathbb{N}}$ under the assumptions. We can also establish the large sample distribution of $\{\alpha^*_G\}_{G\in\mathbb{N}}$ using the standard techniques. We now impose the additional conditions to study the large sample behavior of $\{\hat{\alpha}_G\}$.

Assumption 4 *There exists a compact subset Θ_0 of Θ that satisfies the following conditions.*

(a) *The sequence $\{\theta_G^*\}_{G\in\mathbb{N}}$ is uniformly interior to Θ_0.*

(b) *For some $\theta_0 \in \Theta_0$, $\{|\nabla^2 q_g(X_{g1}, \theta_0)|\}_{g\in\mathbb{N}}$ is uniformly $\mathscr{L}_{2+2\delta}$-bounded for some real number $\delta > 0$, where ∇^2 is the Hessian operator with respect to the parameters. There exists a continuous function $h : \mathbb{R} \to [0, \infty)$ such that $h(y) \downarrow 0$ as $y \downarrow 0$ and a Borel measurable function $d_1 : \mathbb{R}^v \to \mathbb{R}$ such that the sequence $\{d_1(X_{g1})\}_{g\in\mathbb{N}}$ is uniformly $\mathscr{L}_{2+2\delta}$-bounded, and for each $(\theta_1, \theta_2) \in \Theta_0^2$ and each $x \in \mathbb{R}^v$,*

$$|\nabla^2 q_g(x, \theta_2) - \nabla^2 q_g(x, \theta_1)| \leq d_1(x)h(|\theta_2 - \theta_1|), \quad g \in \mathbb{N}.$$

(c) *The sequence $\{A_G^* \equiv A_G(\theta_G^*)\}_{G\in\mathbb{N}}$ is asymptotically uniformly nonsingular, where*

$$A_G(\theta) \equiv N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[\nabla^2 q_g(X_{g1}, \theta)], \quad \theta \in \Theta,\, G \in \mathbb{N}.$$

(d) *For some $\theta_0 \in \Theta_0$, $\{|\nabla q_g(X_{g1}, \theta_0)|\}_{g\in\mathbb{N}}$ is uniformly $\mathscr{L}_{4+4\delta}$-bounded, there exists a Borel measurable function $d_2 : \mathbb{R}^v \to \mathbb{R}$ such that $\{d_2(X_{g1})\}_{g\in\mathbb{N}}$ is uniformly $\mathscr{L}_{4+4\delta}$-bounded, and for each $(\theta_1, \theta_2) \in \Theta_0^2$ and each $x \in \mathbb{R}^v$,*

$$|\nabla q_g(x, \theta_2) - \nabla q_g(x, \theta_1)| \leq d_2(x)h(|\theta_2 - \theta_1|), \quad g \in \mathbb{N},$$

where h is as in (b).

(e) *The array* $\{\mathrm{var}[\xi^*_{Gg}] : (g,G) \in \mathbb{G}\}$ *is uniformly positive, where*

$$\begin{aligned}\xi^*_{Gg} \equiv &\frac{G}{N_G}\Biggl(\mathrm{tr}\Biggl(W_G\Bigl(\sum_{i=1}^{n_g} s^*_{Ggi}{s^*_{Ggi}}'\\ &\qquad\qquad - \frac{n_g}{n_g-1}\sum_{i=1}^{n_g}(s^*_{Ggi}-\tilde{s}^*_{Gg})(s^*_{Ggi}-\tilde{s}^*_{Gg})'\Bigr)\Biggr)\\ &- n_g\mathrm{E}[s^*_{Gg1}]'W\mathrm{E}[s^*_{Gg1}]\\ &- {L^*_G}'{A^*_G}^{-1}\sum_{i=1}^{n_g}\nabla q_g(X_{gi},\theta^*_G)\Biggr), \quad (g,G)\in\mathbb{G},\end{aligned}$$

$\tilde{s}^*_{Gg} \equiv \tilde{s}_g(\cdot,\theta^*_G)$, $L^*_G \equiv L_G(\theta^*_G, W_G)$, *and*

$$\begin{aligned}L_G(\theta,W) \equiv 2N_G^{-1}\sum_{g=1}^{G}\mathrm{E}\Biggl[&-\frac{1}{n_g-1}\sum_{i=1}^{n_g}\nabla s_{gi}(\cdot,\theta)W s_{gi}(\cdot,\theta)\\ &+\frac{n_g^2}{n_g-1}\nabla\tilde{s}_g(\cdot,\theta)W\tilde{s}_g(\cdot,\theta)\Biggr], \quad \theta\in\Theta,\ W\in\mathbb{S}^m.\end{aligned}$$

The conditions imposed in Assumptions 4(a)–(d) are similar to those typically imposed in establishing the asymptotic normality of M-estimators, except that the orders in the moment conditions in Assumption 4 are higher, reflecting the fact that our statistic involves products of elements of $s_{gi}(\cdot,\hat{\theta})$ and $\nabla s_{gi}(\cdot,\hat{\theta})$. We now state the large sample properties of $\{\hat{\alpha}_G\}$.

Theorem 3.1 *Suppose that Assumptions 1–4 hold. Then:*

(a) $\hat{\alpha}_G - \alpha^*_G \to 0$ *in probability-P.*

(b) *If* $\alpha^*_G = O(G^{-1/2})$, *then*

$$G^{1/2}(\hat{\alpha}_G - \alpha^*_G) = G^{-1/2}\sum_{g=1}^{G}\xi^*_{Gg} + \mathrm{o}_P(1) \quad \textit{and} \tag{6}$$

$$V_G^{-1/2}G^{1/2}(\hat{\alpha}_G - \alpha^*_G) \overset{A}{\sim} \mathrm{N}(0,1), \quad \textit{where} \tag{7}$$

$$V_G \equiv G^{-1} \sum_{g=1}^{G} \mathrm{var}[\xi^*_{Gg}], \quad G \in \mathbb{N}.$$

Given the result of Theorem 3.1, we now make the "t-statistic" for α^*_G. To formulate an estimator of V_G, we approximate ξ^*_{Gg} by

$$\begin{aligned}\hat{\xi}_{Gg} \equiv& \frac{G}{N_G}\Bigg(\sum_{i=1}^{n_g} \mathrm{tr}\Bigg(\hat{W}_G\Big(\sum_{i=1}^{n_g} s_{gi}(\cdot,\hat{\theta}_G)s_{gi}(\cdot,\hat{\theta}_G)' \\ &- \frac{n_g}{n_g-1}\sum_{i=1}^{n_g}(s_{gi}(\cdot,\hat{\theta}_G)-\tilde{s}_g(\cdot,\hat{\theta}_G))(s_{gi}(\cdot,\hat{\theta}_G)-\tilde{s}_g(\cdot,\hat{\theta}_G))'\Big)\Bigg) \\ &- \hat{L}'_G\hat{A}^+_G\sum_{i=1}^{n_g}\nabla q_g(X_{gi},\hat{\theta}_G)\Bigg), \quad (g,G) \in \mathbb{G}, \quad \text{where}\end{aligned}$$

$$\begin{aligned}\hat{L}_G \equiv 2N_G^{-1}\sum_{g=1}^{G}\Bigg(&-\frac{1}{n_g-1}\sum_{i=1}^{n_g}\nabla s_{gi}(\cdot,\hat{\theta}_G)\hat{W}_G s_{gi}(\cdot,\hat{\theta}) \\ &+ \frac{n_g^2}{n_g-1}\nabla\tilde{s}_g(\cdot,\hat{\theta}_G)\hat{W}_G\tilde{s}_g(\cdot,\hat{\theta}_G)\Bigg), \quad G \in \mathbb{N} \quad \text{and}\end{aligned}$$

$$\hat{A}_G \equiv N_G^{-1}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\nabla^2 q_g(X_{gi},\hat{\theta}_G), \quad G \in \mathbb{N}.$$

We then estimate V_G by

$$\hat{V}_G \equiv G^{-1}\sum_{g=1}^{G}\hat{\xi}^2_{Gg}, \quad G \in \mathbb{N}.$$

Our test statistic is thus

$$\mathcal{T}_G \equiv \frac{G^{1/2}\hat{\alpha}_n}{\hat{V}_G^{1/2}}, \quad G \in \mathbb{N}. \tag{8}$$

The large sample behavior of this statistic is described in the next theorem.

THEOREM 3.2 *Under Assumptions 1, 2(a), 3, and 4, the following hold.*

(a) *The sequence*

$$\left\{ \bar{V}_G \equiv G^{-1} \sum_{g=1}^{G} \mathrm{var}[\xi_{Gg}] + G^{-1} \sum_{g=1}^{G} \left(\frac{G n_g}{N_G} \right)^2 (\mathrm{E}[s^*_{Gg1}]' W_G \mathrm{E}[s^*_{Gg1}])^2 \right\}_{G \in \mathbb{N}}$$

is bounded, and $\{\hat{V}_G - \bar{V}_G\}_{G \in \mathbb{N}}$ *converges to zero in probability-*P*.*

(b) *If in addition Assumption 2(b) hold, and* $\alpha^*_G = O(G^{-1/2})$*, then* $|\hat{V}_G - V_G| \to 0$ *in probability-*P*, and*

$$\mathcal{T}_G - \frac{G^{1/2} \alpha^*_G}{V_G^{1/2}} \overset{A}{\sim} \mathrm{N}(0, 1).$$

(c) *If* $\{\alpha^*_G\}_{G \in \mathbb{N}}$ *is asymptotically uniformly positive, then for each* $c \in \mathbb{R}$*,* $P[\mathcal{T}_G > c] \to 1$*.*

Theorem 3.2 shows that, to perform a level-π test of H_0, we can set the $(1-\pi)$-quantile of the standard normal distribution to the critical value. It also shows that the test has a power approaching 1 as $G \to \infty$, under the asymptotic uniform positiveness of $\{\alpha^*_G\}$, which holds, for example, if $|\mathrm{E}[s^*_{Gg1}]|$ is bounded away from zero for some fixed fraction of individuals in the population, when G grows to infinity. On the other hand, the test does not have much power, if α^*_G is small. For instance, if the variation of $\theta^\dagger_g$ across groups is small, $|\mathrm{E}[s^*_{Gg1}]|$ would be small in most groups, so that α^*_G would be small. Also, if all but only a few groups share the same value for $\theta^\dagger_g$, $|\mathrm{E}[s^*_{Gg1}]|$ would be small in most groups, and α^*_G would be small. Without knowing specifically how the null hypothesis could be violated, it seems difficult to improve the power in such cases.

In using the proposed test, one has to choose the weighting matrix $\hat{W}_G$. Use of $\hat{\Sigma}^+_G \equiv \check{\Sigma}_G(\cdot, \hat{\theta}_G)$ or $\hat{S}^+_G \equiv \check{\Sigma}_G(\cdot, \hat{\theta}_G)$ for $\hat{W}_G$ can be recommended, because these choices make the proposed test invariant to reparametrization of the model, at least, within the set of m parameters corresponding to the m components of the score vector used in the test, provided that the model reparametrized through a continuously differentiable mapping whose Jacobian is nonsingular everywhere. By using Lemma A.4 with Assumption 4(d) and Lemma B.2, it is straightforward to show that $\{\hat{\Sigma}_G\}_{G\in\mathbb{N}}$ and $\{\hat{S}_G\}_{G\in\mathbb{N}}$ are respectively consistent for

$$\left\{\Sigma_G^* \equiv N_G^{-1}\sum_{g=1}^{G}\frac{n_g^2}{n_g-1}\mathrm{E}[(s_{g1}^*-\mathrm{E}[s_{g1}^*])(s_{g1}^*-\mathrm{E}[s_{g1}^*])']\right\}_{G\in\mathbb{N}} \quad \text{and}$$

$$\left\{S_G^* \equiv N_G^{-1}\sum_{g=1}^{G} n_g \mathrm{E}[s_{g1}^* s_{g1}^{*\prime}]\right\}_{G\in\mathbb{N}},$$

both of which are $O(1)$. It follows that setting $\hat{\Sigma}_G^+$ or $\hat{S}_G^+$ to $\hat{W}_G$ satisfies Assumptions 2 and 3(b), provided that $\{\Sigma_G^*\}_{G\in\mathbb{N}}$ is asymptotically uniformly positive definite.

4. MONTE CARLO SIMULATIONS

In this section, we examine the finite sample behavior of the proposed test. Let $\{Z_{gi} : i \in \mathbb{I}_g,\, g\in\mathbb{N}\}$ be an independently and identically distributed array of 4×1 random vectors such that the first component of Z_{gi} is identically equal to one, and the remaining three components are independently distributed with the standard normal distribution. Also, let $\{\epsilon_{gi} : i\in\mathbb{I}_g,\, g\in\mathbb{N}\}$ be an independent array of random variables that are distributed with the standard normal distribution. The arrays $\{Z_{gi}\}$ and $\{\epsilon_{gi}\}$ are independent. In each experiment, we pick a 4×1 constant vector $\bar{\theta}$ and set $\hat{\theta}_g^\dagger = (-1)^g\bar{\theta}$ for each group g. We then generate $\{Y_{gi} : i\in\mathbb{I}_g,\, g\in\mathbb{N}\}$ by

$$Y_{gi} \equiv Z_{gi}'\theta_g^\dagger + U_{gi}, \quad i\in\mathbb{I}_g,\, g\in\mathbb{N},$$

where for some real numbers $\gamma_0 > 0$ and $\gamma_1 \geq 0$,

$$U_{gi} \equiv (\gamma_0 + \gamma_1 Z_{gi2}^2)^{1/2}\epsilon_{gi},$$

and Z_{gi2} is the second component of Z_{gi}. By construction, we have that for each $g\in\mathbb{N}$, $\mathrm{E}[U_{g1}|Z_{g1}] = 0$. The parameter $\theta_g^\dagger$ can be characterized as the maximizer of $\bar{q}(\theta)$, where $\bar{q}$ is

Table 1: Probability of rejection, $G = 50$, $n_g = 5$

Rejection Probability (%), $G = 50$, $n_g = 5$						
	Homoskedastic			Heteroskedastic		
$\bar{\theta}$	10%	5%	1%	10%	5%	1%
(0,0,0,0)'	9.0	3.7	0.4	8.9	3.7	0.3
(0.05,0,0,0)'	9.4	3.9	0.4	9.6	3.8	0.4
(0.1,0,0,0)'	10.7	4.8	0.7	11.0	4.6	0.6
(0.2,0,0,0)'	18.0	9.0	1.3	19.4	9.3	1.3
(0.3,0,0,0)'	34.9	20.5	4.7	37.2	22.7	4.9
(0.4,0,0,0)'	61.5	44.4	15.5	64.1	47.8	18.2
(0.5,0,0,0)'	86.1	74.9	42.1	87.3	77.6	46.7
(0.6,0,0,0)'	97.6	94.0	76.3	98.0	94.9	78.8
(0,0.05,0,0)'	9.4	3.9	0.4	9.0	3.7	0.3
(0,0.1,0,0)'	10.9	4.6	0.5	9.8	4.0	0.4
(0,0.2,0,0)'	17.2	8.0	1.1	12.9	5.4	0.6
(0,0.3,0,0)'	29.0	16.1	2.5	18.5	9.1	1.1
(0,0.4,0,0)'	47.0	29.3	7.1	27.5	14.5	2.1
(0,0.5,0,0)'	65.6	47.9	15.4	40.8	23.6	4.7
(0,0.6,0,0)'	81.3	66.2	27.4	55.8	36.8	9.4

as in Example 2.1. We thus employ the ordinary least squares (OLS) estimator for $\hat{\theta}_G$.

In each experiment, we set the group sample size g_n equal to 5 for all groups and set G equal to an even natural number. In generating U_{gi}, we set either $(\gamma_0, \gamma_1) = (1, 0)$ (homoskedastic experiments) or $(\gamma_0, \gamma_1) = (0.5, 0.5)$ (heteroskedastic experiments). This makes the variance of U_{gi} is one throughout all experiments. It is straightforward to verify that $\theta^*_G = 0$, regardless of the value of $\bar{\theta}$. To assess the effects of nonconstancy of the intercept and that of slopes separately, we set $\bar{\theta} = (a, 0, 0, 0)'$ in some experiments and $\bar{\theta} = (0, a, 0, 0)'$ in others, where a is a nonnegative real number, which can be interpreted as the magnitude of the variation of the parameters across groups relative to the standard deviation of the error term (which is one). We below tabulate the probability that the proposed test rejects the null hypothesis in experiments with various $\bar{\theta}$ and G. The number of replications is 10,000 in each experiment, the standard errors of the reported probabilities are no larger than 0.5 per cent points $(= \sqrt{0.5^2/10,000})$.

Table 2: Probability of rejection, $G = 100$, $n_g = 5$

Rejection Probability (%), $G = 100$, $n_g = 5$						
	Homoskedastic			Heteroskedastic		
$\bar{\theta}$	10%	5%	1%	10%	5%	1%
(0,0,0,0)'	9.7	4.8	0.7	10.0	4.8	0.8
(0.05,0,0,0)'	10.3	5.1	0.8	10.7	5.2	0.9
(0.1,0,0,0)'	12.6	6.2	1.2	13.2	6.5	1.1
(0.2,0,0,0)'	24.7	13.8	3.2	25.6	14.7	3.4
(0.3,0,0,0)'	51.7	36.0	12.8	53.4	38.0	14.0
(0.4,0,0,0)'	83.5	72.5	42.5	84.6	74.5	45.4
(0.5,0,0,0)'	98.4	95.6	83.0	98.3	95.8	84.0
(0.6,0,0,0)'	100.0	99.9	98.8	99.9	99.8	98.6
(0,0.05,0,0)'	10.1	5.2	0.9	10.3	4.9	0.8
(0,0.1,0,0)'	12.4	6.1	1.1	11.3	5.6	0.9
(0,0.2,0,0)'	23.3	12.2	2.9	15.7	8.3	1.5
(0,0.3,0,0)'	43.5	28.7	8.1	25.8	14.1	3.0
(0,0.4,0,0)'	69.1	53.2	23.0	41.4	26.3	6.9
(0,0.5,0,0)'	88.5	78.0	46.3	60.2	44.1	15.8
(0,0.6,0,0)'	96.9	92.4	71.0	77.9	63.9	30.8

Table 3: Probability of rejection, $G = 500$, $n_g = 5$

Rejection Probability (%), $G = 500$, $n_g = 5$						
	Homoskedastic			Heteroskedastic		
$\bar{\theta}$	10%	5%	1%	10%	5%	1%
(0,0,0,0)'	10.0	4.9	0.9	10.0	5.0	0.8
(0.05,0,0,0)'	11.5	5.9	1.2	11.3	6.0	1.1
(0.1,0,0,0)'	17.4	9.5	2.2	17.5	9.5	2.0
(0.2,0,0,0)'	53.7	38.8	15.8	54.4	39.1	16.1
(0.3,0,0,0)'	95.5	90.7	73.6	95.3	90.8	74.6
(0.4,0,0,0)'	100.0	100.0	99.7	100.0	100.0	99.7
(0.5,0,0,0)'	100.0	100.0	100.0	100.0	100.0	100.0
(0.6,0,0,0)'	100.0	100.0	100.0	100.0	100.0	100.0
(0,0.05,0,0)'	11.5	6.0	1.1	10.7	5.5	0.9
(0,0.1,0,0)'	17.1	9.3	2.1	13.2	6.9	1.3
(0,0.2,0,0)'	48.8	34.6	13.4	26.4	15.9	4.4
(0,0.3,0,0)'	90.2	82.1	57.0	55.7	40.2	16.1
(0,0.4,0,0)'	99.8	99.4	95.5	86.2	76.2	49.2
(0,0.5,0,0)'	100.0	100.0	100.0	98.5	96.4	85.6
(0,0.6,0,0)'	100.0	100.0	100.0	100.0	99.8	98.7

Table 1 shows the probability of rejection in experiments with $G = 50$ and $n_g = 5$. The left three columns report the rejection probabilities of the test with nominal sizes 10%, 5%, and 1% in the homoskedastic experiments, while the right three columns do the same for the heteroskedastic experiments. The experiments with $\bar{\theta} = (0, 0, 0, 0)'$ shows that the test has sizes close to the nominal sizes, though the test tends to somewhat underreject the null hypothesis. The second and third panes of the table show that the power steadily climbs up as the variation of $\theta_g^\dagger$ across the groups becomes larger. Tables 2 and 3 show that with larger G's, the sizes of the test become even closer to the nominal sizes, and the slope of the power curve becomes steeper, as our large sample theory suggests.

5. AN ILLUSTRATIVE EXAMPLE

To assess the effects of schooling and job experience of a worker on his wage, the linear regression of wage in logarithm on years of schooling, years of experience, and other characteristics of the worker is often estimated. If such a regression model properly accounts for all important determinants of the wage rate, the effects of the determinants on the wage rate should be the same at every workplace in the population. If we find that the regression parameters vary from a workplace to another, the model still has a space for improvement, omitting some determinants of the wage rate. In this section, we apply the proposed method to test the constancy of the parameters of a particular wage regression model across workplaces in Canada.

The data set we use in this section is Workplace and Employee Survey (WES) 1999 from Statistics Canada. The WES offers the information on workplaces in Canada and the employees at the workplaces. The information on the employees includes the identification number for each employee's workplace as well as the wage rate, schooling, years of job experience, seniority, ethnicity, gender, marital status, union membership, and occupation type of each employee.

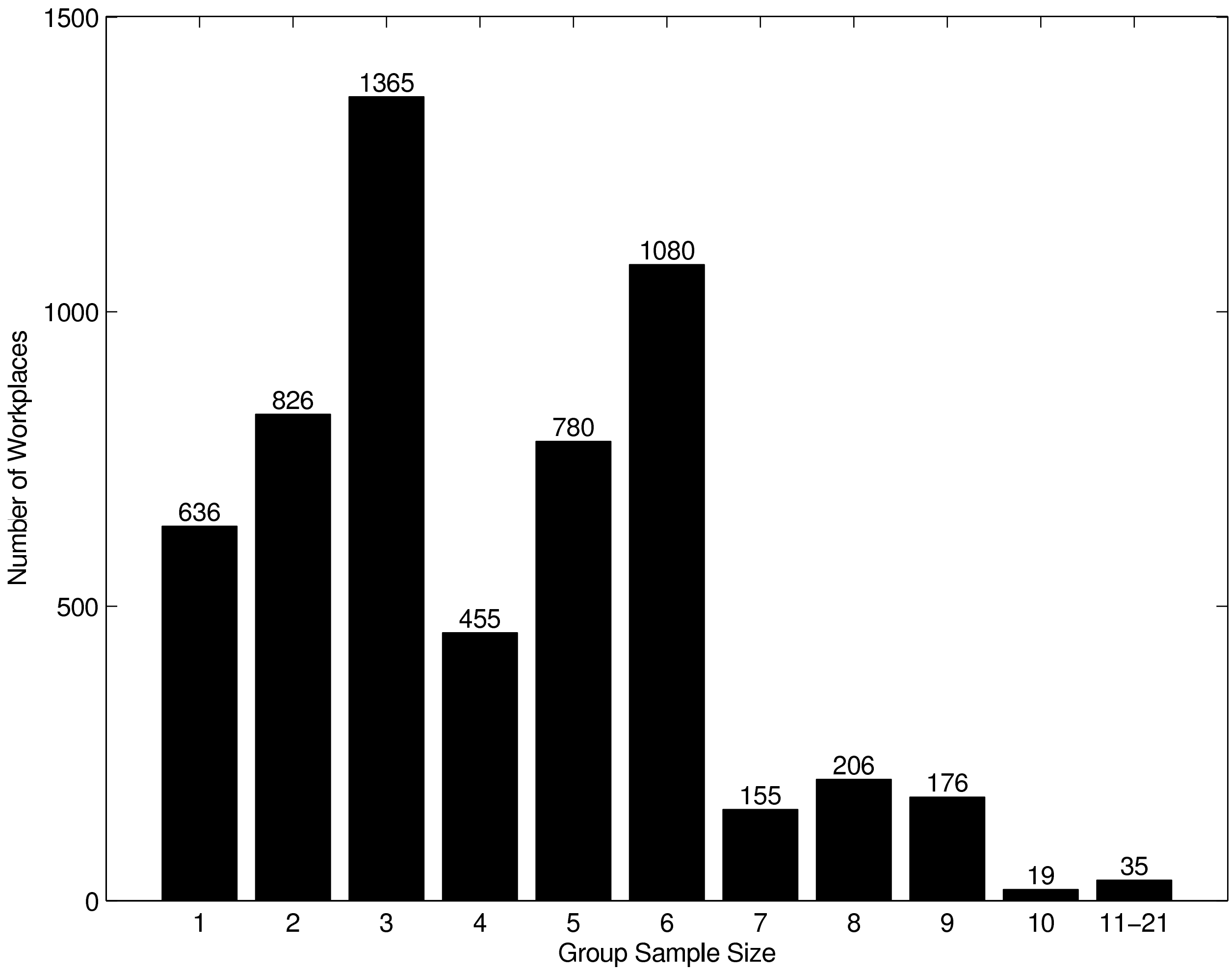

Figure 1: Frequency of work place sample size in the Workplace and Employee Survey 1999
* No more than 21 observations were drawn from any workplace.

The WES 1999 is a large data set containing observations of 23,540 employees from 5,733 workplaces in Canada. Nevertheless, the workplace sample size (i.e. the number of workers sampled from a workplace) ranges between one and 21 and averages only 4.11. Figure 1 shows the frequency of the workplace sample sizes across the 5,733 workplaces. The sample only offers six or fewer observations for 90% of the workplaces and more than ten observations only for 0.6% of the workplaces. Thus, the WES 1999 contains many workplaces but only a few observations from each of the workplaces, as we assumed in developing our test method. In our analysis, we drop the 636 workplaces with only one observation, as our method requires two or more observations from each workplace.

Table 4: Wage regression function estimated by using the Workplace and Employee Survey 1999

Regressor	Estimate	Standard Error
Highest grade completed up to high school	0.0269	0.00206
Additional education and trainng (binary varaibles)		
Trade-vocational training	0.0491	0.00758
Some college	0.0328	0.00852
Completed college	0.0745	0.00672
Some university	0.069	0.0104
Teacher's college	0.0128	0.0563
University below B.Sc.	0.134	0.0198
B.Sc.	0.234	0.00919
University above B.Sc.	0.195	0.0224
M.Sc.	0.186	0.0178
Degree in medicine	0.215	0.0445
Ph.D	0.152	0.0528
Industry certified training or course	0.0515	0.0116
Others	0.0276	0.013
Years of experience	0.0222	0.000901
Years of experience squared	$-3.98 \cdot 10^{-4}$	$2.17 \cdot 10^{-5}$
Seniority	$9.92 \cdot 10^{-4}$	$4.56 \cdot 10^{-5}$
Black (binary variable)	-0.0706	0.0312
Female (binary variable)	-0.157	0.00592
Married (binary variable)	0.072	0.00573
Union membership (binary variable)	0.0848	0.00567
Occupation groups (binary varaibles)		
Managers	0.478	0.0137
Professionals	0.402	0.013
Technical or Trades	0.174	0.0105
Marketing or Sales	-0.0648	0.0181
Clerical or Administrative	0.0865	0.0115
Intercept	1.95	0.0254
R^2	0.369	

(a) The column labeled "Standard Error" shows the heteroskedasticity-robust estimate of the standard error for each parameter estimate.

(b) The occupation groups are exhaustive and mutual exclusive categorization. The base group, which is ommitted from the regression, is "Production workers with no trade/certification, operation and maintenance".

Table 4 reports the regression of the log-wage on schooling, experience, seniority, occupation type, and some other individual characteristics, estimated by the OLS method using the WES 1999. Note that the number of parameters in the regression model exceeds the workplace sample size at most workplaces, so that it is clearly infeasible to estimate the particular regression model for each of the workplaces separately. The estimates in the table qualitatively agree with those found in similar regressions in the literature. All schooling and training variables have positive effects on the wage rate. Years of experience also have positive effects with the decreasing marginal return. Seniority has a tiny positive effect, which is yet statistically significant. The marriage and union membership affect the wage rate positively, while the black and female dummies have negative coefficients. For the occupation dummies, we see that the coefficients are in line with the skill levels required for the occupation groups. The estimated regression function exhibits no obvious problems.

We now apply the method proposed in this paper to test the hypothesis that the coefficients stay the same across workplaces. Using the entire score for s_{gi} and $\hat{\Sigma}_G^+$ for $\hat{W}_G$, the realized value of $\mathcal{T}_G$ is 15.56. The null hypothesis is strongly rejected. It seems that there is still a space for improvement in the regression specification used above.

Unfortunately, the rejection of a model in this form does not tell a researcher specifically how the model should be modified. This feature is not limited to our test. Many of the specification testing methods with general alternative hypotheses share the same property.

6. CONCLUDING REMARKS

The dependence of an individual's behavior on his affiliation with a group is often suspected. Given a parametric model, the dependence of the behavior on the group affiliation means the constancy of the model parameters across groups. In this paper, we consider the parameter constancy testing problem in the setup in which the population consists of many groups, each of which offers only a few observations. The parameter constancy test in such problem set up has not been studied at our best knowledge.

This paper develops a novel test of parameter constancy that works in the above-mentioned setup and studies its statistical properties in detail. The proposed test only requires that an M-estimator accurately estimate the model parameters under the null. The estimation of the parameter for each group is unnecessary. The sizes of group samples can be as small as two for the proposed method to work.

When assessing the power property of the proposed test presented via the large-G asymptotics and Monte Carlo simulations in this paper, one should keep in mind that the alternative hypothesis of the test is the violation of the parameter constancy in any form. A test designed with a more specific alternative in mind may have a better power under the particular alternative than our test but can poorly perform under other alternatives. Suppose, for example, that in the setup of the Monte Carlo simulations in Section 4, a researcher knows that the groups with even g share the same parameter value, and the groups with odd g also share the same parameter value . He can then take the usual Wald, LM, or LR test approach to check the equality of the parameter between the odd and even groups. His test would be undoubtedly more powerful than our test. If, however, the groups are reordered in such a manner that all groups that had odd g's have g between one and $G/2$, and all that had even g's have g between $G/2+1$ and G, then the same Wald, LM, and LR tests have no power in excess of their level. That is, the three tests lose their power under the new unanticipated alternative. Our test, on the other hand, have exactly the same performance under the two alternatives.

One limitation of the proposed method is that we cannot test the constancy of the subset of the parameters across groups (unless the other parameters are known to be constant across groups). This reflects the fact that our setup does not allow us to identify the parameters for each group. In other words, testing constancy of a subset of parameters across groups seems to require that the group sample sizes be large enough (relative to the number of parameters) to allow the parameters to be estimated accurately. If the parameters are estimable in each group, one could capture the variability of the parameters (or a subset of them) across groups in a way similar to the approach of Swamy (1970), as Pesaran and Yamagata (2008) do in their problem setup.

APPENDIX A: A UNIFORM LAW OF LARGE NUMBERS, CENTRAL LIMIT THEOREMS, AND SOME OTHER USEFUL RESULTS

We prove the theorems in the main text in Appendix B. We here establish a few results useful in the proofs, including a uniform law of large numbers, a central limit theorem, and a functional central limit theorem.

Recall that $|\cdot|$ denotes the Frobenius norm. Throughout the appendices, $\|Z\|_r$ denotes the $\mathscr{L}_r$-norm of $|Z|$ for each random matrix Z.

LEMMA A.1 *Let Z_1 and Z_2 be random matrices such that the column dimension of Z_1 equals the row dimension of Z_2. Then for each $p > 0$,*

$$\|Z_1 Z_2\|_p \leq \|Z_1\|_{2p} \|Z_2\|_{2p}.$$

Proof of LEMMA A.1: Because the Frobenius norm is submultiplicative, we have that $|Z_1 Z_2| \leq |Z_1|\,|Z_2|$. It follows that

$$\|Z_1 Z_2\|_p = \mathrm{E}[|Z_1 Z_2|^p]^{1/p} \leq \mathrm{E}[|Z_1|^p\,|Z_2|^p]^{1/p}.$$

By the Cauchy-Schwarz inequality, the right-hand side of this inequality is dominated by

$$
\begin{aligned}
(\mathrm{E}[|Z_1|^{2p}]^{1/2}\mathrm{E}[|Z_2|^{2p}]^{1/2})^{1/p} =& \mathrm{E}[|Z_1|^{2p}]^{1/(2p)}\mathrm{E}[|Z_2|^{2p}]^{1/(2p)} \\
=& \|Z_1\|_{2p}\,\|Z_2\|_{2p}.
\end{aligned}
$$

The desired result therefore follows. $\square$

LEMMA A.2 *Let F_1 be a function from a metric space (Γ, d) to the normed linear space $(\mathbb{R}^{l_1\times l_2}, |\cdot|)$ and F_2 a function from (Γ, d) to $(\mathbb{R}^{l_2\times l_3}, |\cdot|)$. Suppose that there exist real constants c_1 and c_2 and functions h_1 and h_2 from $\mathbb{R}$ to $[0,\infty)$ such that*

$$|F_j(\gamma)| \le c_j, \quad \gamma \in \Gamma,\ j \in \{1,2\}$$

and

$$|F_j(\gamma_2) - F_j(\gamma_1)| \le c_j h_j(d(\gamma_1,\gamma_2)), \quad (\gamma_1,\gamma_2) \in \Gamma^2,\ j \in \{1,2\}.$$

Define $F_3 : \Gamma \to \mathbb{R}^{l_1\times l_3}$ and $h_3 : \mathbb{R} \to [0,\infty)$ by

$$F_3(\gamma) \equiv F_1(\gamma)F_2(\gamma), \quad \gamma \in \Gamma \quad \textit{and}$$

$$h_3(z) \equiv h_1(z) + h_2(z), \quad z \in \mathbb{R}.$$

Then for each $(\gamma_1,\gamma_2) \in \Gamma^2$

$$|F_1(\gamma)F_2(\gamma)| \le c_1c_2, \quad \gamma \in \Gamma,$$

and

$$|F_1(\gamma_2)F_2(\gamma_2) - F_1(\gamma_1)F_2(\gamma_1)| \le (c_1+c_2)\, h_3(d(\gamma_1,\gamma_2)), \quad (\gamma_1,\gamma_2) \in \Gamma^2.$$

Proof of LEMMA A.2: The first inequality immediately follows from the submultiplicativity of the Frobenius norm. For the second inequality, let γ_1 and γ_2 be an arbitrary pair of points in Γ. Because

$$F_3(\gamma_2) - F_3(\gamma_1) = F_1(\gamma_2)(F_2(\gamma_2) - F_2(\gamma_1)) + (F_1(\gamma_2) - F_1(\gamma_1))F_2(\gamma_1),$$

we have that

$$\begin{aligned}
&|F_3(\gamma_2) - F_3(\gamma_1)| \\
&\leq |F_1(\gamma_2)|\,|F_2(\gamma_2) - F_2(\gamma_1)| + |F_2(\gamma_1)|\,|F_1(\gamma_2) - F_1(\gamma_1)| \\
&\leq (c_1 + c_2)(h_1(d(\gamma_1, \gamma_2)) + h_2(d(\gamma_1, \gamma_2))) = (c_1 + c_2)h_3(d(\gamma_1, \gamma_)).
\end{aligned}$$

Thus, the second equality follows. □

Lemma A.3 *For each random matrix Z for which* $\mathrm{E}[Z]$ *exists, it holds that* $|\,\mathrm{E}[Z]\,| \leq \mathrm{E}[\,|Z|\,]$.

Proof of Lemma A.3: Let Z be a $l_1 \times l_2$ random vector such that $\mu \equiv \mathrm{E}[Z]$ exists. Let ϕ denote the function $|\cdot| : \mathbb{R}^{l_1 \times l_2} \to \mathbb{R}$. Because ϕ is convex on $\mathbb{R}^{l_1 \times l_2}$, by the support theorem (Luenberger, 1969, Theorem 2, p. 133), there exists a linear function $L : \mathbb{R}^{l_1 \times l_2} \to \mathbb{R}$ such that for each $z \in \mathbb{R}^{l_1 \times l_2}$, $\phi(z) \geq \phi(\mu) + L(z - \mu)$. It follows that that

$$\begin{aligned}
\mathrm{E}[\,|Z|\,] &= \mathrm{E}[\phi(Z)] \geq \phi(\mu) + \mathrm{E}[L(Z - \mu)] \\
&= \phi(\mu) + L(\mathrm{E}[Z] - \mu) = \phi(\mu) = |\,\mathrm{E}[Z]\,|.
\end{aligned}$$

□

Lemma A.4 *Let $(\Omega, \mathscr{F}, P)$ be a probability space, (Γ, d) a compact metric space, and $\{F_{Gg} : (g, G) \in \mathbb{G}\}$ a double array of functions from $\Omega \times \Gamma$ to $\mathbb{R}^{l_1 \times l_2}$ measurable-$(\mathscr{F} \otimes \mathscr{B}(\Gamma))/\mathscr{B}^{l_1 \times l_2}$. Suppose that for each $\gamma \in \Gamma$, $\{F_{Gg}(\cdot, \gamma)\}$ is a rowwise independent array and that there exists $\gamma_0 \in \Gamma$ and a real number $\delta \geq 0$ such that $\{F_{Gg}(\cdot, \gamma_0)\}$ is uniformly $\mathscr{L}_{1+\delta}$-bounded. Also, suppose that, with the same δ, there exist a continuous function $h : \mathbb{R} \to [0, \infty)$ and a* *uniformly $\mathscr{L}_{1+\delta}$-bounded, rowwise independent array of random variables* $\{D_{Gg} : (g, G) \in \mathbb{G}\}$ *that satisfy that* $h(y) \downarrow 0$ *as* $y \downarrow 0$ *and that*

$$|F_{Gg}(\cdot,\gamma_1) - F_{Gg}(\cdot,\gamma_2)| \leq D_{Gg}\, h(d(\gamma_1,\gamma_2)), \quad (\gamma_1,\gamma_2) \in \Gamma^2,\, (g,G) \in \mathbb{G}. \tag{9}$$

Then

(a) *The set* $\{F_{Gg}(\cdot,\gamma) : \gamma \in \Gamma,\, (g,G) \in \mathbb{G}\}$ *is uniformly* $\mathscr{L}_{1+\delta}$*-bounded.*

(b) *The sequence* $\{\gamma \mapsto G^{-1}\sum_{g=1}^{G} \mathrm{E}[F_{Gg}(\cdot,\gamma)] : \Gamma \to \mathbb{R}\}_{G\in\mathbb{N}}$ *is uniformly bounded and uniformly equicontinuous.*

(c) *If in addition the assumptions of the current theorem hold with some* $\delta > 0$*, then*

$$\sup_{\gamma\in\Gamma}\left|G^{-1}\sum_{g=1}^{G} F_{Gg}(\cdot,\gamma) - G^{-1}\sum_{g=1}^{G} \mathrm{E}[F_{Gg}(\cdot,\gamma)]\right| \to 0$$

as $n \to \infty$ *in probability-*P*.*

Proof of Lemma A.4: Pick $\gamma \in \Gamma$ arbitrarily. We have that for each $(g,G) \in \mathbb{G}$,

$$\begin{aligned}|F_{Gg}(\cdot,\gamma)| &\leq |F_{Gg}(\cdot,\gamma_0)||F_{Gg}(\cdot,\gamma) - F_{Gg}(\cdot,\gamma_0)| \\ &\leq |F_{Gg}(\cdot,\gamma_0)| + D_{Gg}\, h(d(\gamma_1,\gamma_2)) \leq |F_{Gg}(\cdot,\gamma_0)| + \bar{h} D_{Gg},\end{aligned}$$

where $\bar{h} \equiv \sup\{h(|\gamma_2 - \gamma_1|) : \gamma \in \Gamma\} < \infty$, because h is continuous and Γ is compact. It follows that

$$\|F_{Gg}(\cdot,\gamma)\|_{1+\delta} \leq \|F_{Gg}(\cdot,\gamma_0)\|_{1+\delta} + \bar{h}\|D_{Gg}\|_{1+\delta}, \quad \gamma \in \Gamma,\, (g,G) \in \mathbb{G}.$$

Claim (a) follows from this inequality, since $\{F_{gG}(\cdot,\gamma) : \gamma \in \Gamma, (g,G) \in \mathbb{G}\}$ and $\{D_{Gg} : (g,G) \in \mathbb{G}\}$ are uniformly $\mathscr{L}_{1+\delta}$-bounded.

In (b), the uniform boundedness immediately follows from (a). For the uniform equicontinuity, we use (9) to obtain that for each $(\gamma_1, \gamma_2) \in \Gamma^2$ and each $G \in \mathbb{N}$,

$$\begin{aligned}
&\left| G^{-1} \sum_{g=1}^{G} \mathrm{E}[F_{Gg}(\cdot, \gamma_1)] - G^{-1} \sum_{g=1}^{G} \mathrm{E}[F_{Gg}(\cdot, \gamma_2)] \right| \\
&\leq G^{-1} \sum_{g=1}^{G} |\mathrm{E}[F_{Gg}(\cdot, \gamma_1) - F_{Gg}(\cdot, \gamma_2)]| \\
&\leq G^{-1} \sum_{g=1}^{G} \mathrm{E}[\, |F_{Gg}(\cdot, \gamma_1) - F_{Gg}(\cdot, \gamma_2)|\,] \\
&\leq G^{-1} \sum_{g=1}^{G} \mathrm{E}[D_{Gg}]\, h(d(\gamma_1, \gamma_2)) \leq c h(d(\gamma_1, \gamma_2)),
\end{aligned}$$

where the first inequality follows from Lemma A.3, and

$$c \equiv \sup\{\mathrm{E}[D_{Gg}] : \, (g, G) \in \mathbb{G}\}$$

is finite, because $\{D_{Gg} : \, (g, G) \in \mathbb{G}\}$ is uniformly $\mathscr{L}_{1+\delta}$-bounded. The desired result therefore follows.

For (c), we have that for each $\gamma \in \Gamma$

$$G^{-1} \sum_{g=1}^{G} F_{Gg}(\cdot, \gamma) - G^{-1} \sum_{g=1}^{G} \mathrm{E}[F_{Gg}(\cdot, \gamma)] \to 0 \text{ in probability-}P \tag{10}$$

by Sen (1970, Theorem 3), given (a). It thus suffices to show that

$$\left\{ G^{-1} \sum_{g=1}^{G} F_{Gg}(\cdot, \gamma) - G^{-1} \sum_{g=1}^{G} \mathrm{E}[F_{Gg}(\cdot, \gamma)] \right\}_{G \in \mathbb{N}}$$

is stochastically equicontinuous uniformly in $\gamma \in \Gamma$ (Andrews, 1992, Theorem 1). The stochastic equicontinuity holds if (A)

$$\left\{\gamma \mapsto G^{-1}\sum_{g=1}^{G} \mathrm{E}[f_{Gg}(\cdot,\gamma)] : \Gamma \to \mathbb{R}\right\}_{G\in\mathbb{N}}$$

is equicontinuous uniformly in $\gamma \in \Gamma$, and (B)

$$G^{-1}\sum_{g=1}^{G} D_{Gg} = \mathrm{O}_P(1) \text{ as } G \to \infty,$$

in addition to (9) assumed here (Andrews, 1992, Lemma 1(a)). Because (A) has been already established in (b), and (B) immediately follows from Sen (1970, Theorem 3), the desired result follows. $\square$

Lemma A.5 *Let $\{U_{Gg} : (g,G) \in \mathbb{G}\}$ be a rowwise independent array of zero-mean $v \times 1$ random vectors.*

(a) *If $\{U_{Gg}\}$ is uniformly $\mathscr{L}_{2+\delta}$-bounded for some $\delta \geq 0$, then*

$$G^{-1/2}\sum_{g=1}^{G} U_{Gg} = \mathrm{O}_P(1).$$

(b) *If $\{U_{Gg}\}$ is uniformly $\mathscr{L}_{2+\delta}$-bounded for some $\delta > 0$, and $\{\mathrm{var}[U_{Gg}] : (g,G) \in \mathbb{G}\}$ is uniformly positive definite, then*

$$V_G^{-1/2} G^{-1/2}\sum_{g=1}^{G} U_{Gg} \overset{A}{\sim} \mathrm{N}(0, I) \text{ as } G \to \infty,$$

where

$$V_G \equiv G^{-1}\sum_{g=1}^{G} \mathrm{var}[U_{Gg}], \quad G \in \mathbb{N}.$$

Proof of Lemma A.5: The uniform $\mathscr{L}_{2+\delta}$-boundedness of $\{U_{Gg}\}$ implies the uniform $\mathscr{L}_2$-boundedness of $\{U_{Gg}\}$. Because $\{U_{Gg}\}$ is rowwise independent with zero-means and finite second moments, we have that

$$\mathrm{E}\Big[\Big|G^{-1/2}\sum_{g=1}^{G}U_{Gg}\Big|^2\Big] = \mathrm{E}\Big[\Big(G^{-1/2}\sum_{g=1}^{G}U_{Gg}\Big)'\Big(G^{-1/2}\sum_{g=1}^{G}U_{Gg}\Big)\Big]$$
$$= G^{-1}\sum_{g=1}^{G}\mathrm{E}[U_{Gg}'U_{Gg}] = G^{-1}\sum_{g=1}^{G}\|U_{Gg}\|_2^2.$$

Because $\{U_{Gg}\}$ is uniformly $\mathscr{L}_2$-bounded, it follows that

$$\sup_{G\in\mathbb{N}}\mathrm{E}\Big[\Big|G^{-1/2}\sum_{g=1}^{G}U_{Gg}\Big|^2\Big] < \infty.$$

Applying the Markov inequality (Davidson, 1994, p. 132) with this fact establishes (a).

To establish (b), by the Cramér-Wold device (Davidson, 1994, Theorem 25.5, p. 405), it suffices to show for each unit-length $\lambda \in \mathbb{R}^v$,

$$\lambda' V_G^{-1/2} G^{-1/2}\sum_{g=1}^{G}U_{Gg} \overset{A}{\sim} \mathrm{N}(0,1) \text{ as } G \to \infty. \tag{11}$$

Pick a unit-length $\lambda \in \mathbb{R}^v$ arbitrarily. Then the array

$$\{\Xi_{Gg} \equiv \lambda' V_G^{-1/2} G^{-1/2} U_{Gg} : (g,G) \in \mathbb{G}\}$$

satisfies that

$$|\Xi_{Gg}| \le |\lambda|\,|V_G^{-1/2}|\,G^{-1/2}\,|U_{Gg}| = |V_G^{-1/2}|\,G^{-1/2}\,|U_{Gg}|, \quad (g,G) \in \mathbb{G}.$$

The uniform nonsingularity of $\{\mathrm{var}[U_{Gg}]\}$ implies that the uniform nonsingularity of V_G, so that $|V_G^{-1/2}|$, which is dominated by the reciprocal of the square root of the minimum eigenvalue of V_G, is bounded uniformly in $G \in \mathbb{N}$. It follows that

$$|\Xi_{Gg}| \le c\,G^{-1/2}\,|U_{Gg}|, \quad (g,G) \in \mathbb{G},$$

where $c \equiv \sup_{G \in \mathbb{N}} |V_G^{-1/2}|$. Using this fact, we obtain that

$$\begin{aligned}\sum_{g=1}^{G} \mathrm{E}[|\Xi_{Gg}|^{2+\delta}] &\leq c^{2+\delta} G^{-(1+(\delta/2))} \sum_{g=1}^{G} \mathrm{E}[|U_{Gg}|^{2+\delta}] \\ &\leq G^{-\delta/2} c^{2+\delta} \Big(\sup\{\|U_{Gg}\|_{2+\delta}\} \Big)^{2+\delta} \to 0 \text{ as } G \to \infty,\end{aligned}$$

because $\{U_{Gg}\}$ is uniformly $\mathscr{L}_{2+\delta}$-bounded. The Liapunov condition thus holds for $\{\Xi_{Gg}\}$, and so does the Lindeberg condition (Davidson, 1994, Theorem 23.11, p. 372–373). In addition, we have that

$$\sum_{g=1}^{G} \mathrm{var}[\Xi_{Gg}] = 1, \quad G \in \mathbb{N}.$$

The desired results follows by the Lindeberg theorem (Davidson, 1994, Theorem 23.6, p. 369–371). □

APPENDIX B: PROOF OF THEOREMS

We first prove several lemmas and then apply them to prove the theorems stated in the main text. The uniform law of large numbers and the central limit theorem employed in the following proofs are provided in Appendix A, along with a few other useful results.

LEMMA B.1 *Suppose that Assumption 1(a) hold. Then it holds that for each $(g, G) \in \mathbb{G}$,*

$$\frac{2}{\bar{n}} \leq \frac{n_g G}{N_G} \leq \frac{\bar{n}}{2}.$$

Also, for each $g \in \mathbb{N}$,

$$1 \leq \frac{n_g}{n_g - 1} \leq 2.$$

Proof of LEMMA B.1: The group sample size n_g is between 2 and $\bar{n}$ by Assumption 1(a), and so is the average group sample size, N_G/G. It follows that $(n_g G)/N_G = n_g/(N_G/G)$ is between $2/\bar{n}$ and $\bar{n}/2$. This establishes the first inequality.

For the second inequality, note that $n/(n-1)$ is decreasing in n. Thus, $n_g/(n_g-1)$ attains its maximum when $n_g = 2$, and it is never smaller than $\lim_{n\to\infty} n/(n-1) = 1$. This verifies the second inequality. $\square$

LEMMA B.2 *Suppose that Assumptions 1(a)(b) and 4(b) hold. Then there exists an uniformly $\mathscr{L}_{2+2\delta}$-bounded, independently distributed sequence of random variables, $\{D_{H,g}\}_{g\in\mathbb{N}}$, such that*

$$|\nabla^2 q_{gi}(\cdot,\theta)| \le D_{H,g}, \quad \theta \in \Theta_0,\, i \in \mathbb{I}_g,\, g \in \mathbb{N} \quad \textit{and}$$

$$\begin{aligned}|\nabla^2 q_{gi}(X_{gi},\theta_2) - \nabla q_{gi}(X_{gi},\theta_1)| &\le D_{H,g} h(|\theta_2-\theta_1|),\\ &(\theta_1,\theta_2) \in \Theta_0^2,\, i \in \mathbb{I}_g,\, g \in \mathbb{N}.\end{aligned}$$

Proof of LEMMA B.2: Because h is continuous and Θ_0 is compact, we have that $\bar{h} \equiv \sup\{h(|\theta_2-\theta_1|) : (\theta_1,\theta_2) \in \Theta_0^2\} < \infty$. For each $g \in \mathbb{N}$, set

$$D_{H,g} \equiv \sum_{i=1}^{n_g} |\nabla^2 q_g(X_{gi},\theta_0)| + (1+\bar{h}) \sum_{i=1}^{n_g} d_1(X_{gi}).$$

Then, $\{D_{H,g}\}_{g\in\mathbb{N}}$ is a independent sequence, because $\{X_{gi} : i \in \mathbb{I}_g,\, g \in \mathbb{N}\}$ is an independently distributed array by Assumption 1(a). Also, we have that

$$\begin{aligned}\|D_{H,g}\|_{2+2\delta} &\le \sum_{i=1}^{n_g} \|\nabla^2 q_g(X_{gi},\theta_0)\|_{2+2\delta} + (1+\bar{h}) \sum_{i=1}^{n_g} \|d_1(X_{gi})\|_{2+2\delta}\\ &\le \bar{n}\|\nabla^2 q_g(X_{g1},\theta_0)\|_{2+2\delta} + (1+\bar{h})\bar{n}\|d_1(X_{g1})\|_{2+2\delta}, \quad g \in \mathbb{N}.\end{aligned}$$

The uniform $\mathscr{L}_{2+2\delta}$-boundedness of $\{D_{H,g}\}$ follows from this inequality, because both $\{\nabla^2 q_g(X_{g1},\theta_0)\}_{g\in\mathbb{N}}$ and $\{d_1(X_{g1})\}_{g\in\mathbb{N}}$ are uniformly $\mathscr{L}_{2+2\delta}$-bounded by Assumption 4(b). Furthermore, it follows from Assumption 4(b) that

$$\begin{aligned}|\nabla^2 q_g(X_{gi},\theta)| &\leq |\nabla^2 q_g(X_{gi},\theta_0)| + |\nabla^2 q_g(X_{gi},\theta) - \nabla^2 q_g(X_{gi},\theta_0)| \\ &\leq |\nabla^2 q_g(X_{gi},\theta_0)| + d_1(X_{gi})h(|\theta-\theta_0|) \\ &\leq |\nabla^2 q_g(X_{gi},\theta_0)| + \bar{h}d_1(X_{gi}) \leq D_{H,g}, \quad \theta \in \Theta_0,\, i \in \mathbb{I}_g,\, g \in \mathbb{N}.\end{aligned}$$

and that for each $(\theta_1,\theta_2) \in \Theta_0^2$,

$$\begin{aligned}|\nabla^2 q_g(X_{gi},\theta_2) - \nabla^2 q_g(X_{gi},\theta_1)| &\leq d_1(X_{gi})h(|\theta_2-\theta_1|) \\ &\leq D_{H,g}h(|\theta_2-\theta_1|), \quad i \in \mathbb{I}_g,\, g \in \mathbb{N}.\end{aligned}$$

This completes the proof. □

LEMMA B.3 *Suppose that Assumptions 1(a)(b) and 4(b) hold. Define* $\{\check{A}_G : \Omega \times \Theta \to \mathbb{R}^{p\times p}\}_{G\in\mathbb{N}}$ *by*

$$\check{A}_G(\cdot,\theta) \equiv \nabla^2 Q_G(\cdot,\theta) = G^{-1}\sum_{g=1}^{G} \check{a}_{Gg}(\cdot,\theta), \quad \theta \in \Theta,\, G \in \mathbb{N},$$

where $\check{a}_{Gg} : \Omega \times \Theta \to \mathbb{R}^{p\times p}$ *is defined by*

$$\check{a}_{Gg} \equiv \frac{G}{N_G}\sum_{i=1}^{n_g} \nabla^2 q_g(X_{gi},\theta), \quad \theta \in \Theta,\, G \in \mathbb{N}.$$

Then $\{A_G : \Theta \to \mathbb{R}^{p\times p}\}_{G\in\mathbb{N}}$ *is bounded and equicontinuous uniformly on* Θ_0*. Also,* $\{|\check{A}_G(\cdot,\theta) - A_G(\theta)|\}_{G\in\mathbb{N}}$ *converges to zero uniformly in* $\theta \in \Theta_0$ *in probability-*P*.*

Proof of LEMMA B.3: We apply Lemma A.4, taking $\{\check{a}_{Gg} : (g,G) \in \mathbb{G}\}$ for $\{F_{Gg}\}$. The set Θ_0 is a compact subset of the p-dimensional Euclidean space. For each $\theta \in \Theta_0$, the array $\{\check{a}_{Gg}(\cdot,\theta)\}$ is rowwise independent. By Lemma B.1 and B.2, we have that for each

$(g, G) \in \mathbb{G}$,

$$|\check{a}_{Gg}(\cdot, \theta)| \leq \frac{G}{N_G} \sum_{i=1}^{n_g} |\nabla^2 q_g(X_{gi}, \theta)| \leq \frac{n_g G}{N_G} D_{H,g} \leq \frac{\bar{n}}{2} D_{H,g}, \quad \theta \in \Theta_0,$$

$$\begin{aligned}|\check{a}_{Gg}(\cdot, \theta_2) - \check{a}_{Gg}(\cdot, \theta_1)| &\leq \frac{G}{N_G} \sum_{i=1}^{n_g} D_{H,g} h(|\theta_2 - \theta_1|) \\ &\leq \frac{n_g G}{N_G} D_{H,g} h(|\theta_2 - \theta_1|) \leq \frac{\bar{n}}{2} D_{H,g} h(|\theta_2 - \theta_1|), \quad (\theta_1, \theta_2) \in \Theta_0^2,\end{aligned}$$

where $\{D_{H,g}\}_{g \in \mathbb{N}}$ is a uniformly $\mathscr{L}_{2+2\delta}$-bounded, independent sequence of random variables. Thus, Lemma A.4 applies to $\{\check{a}_{Gg} : (g, G) \in \mathbb{G}\}$, and the desired results follow. $\square$

LEMMA B.4 *Suppose that Assumptions 1(a)(b) and 4(d) hold. Then there exists an uniformly $\mathscr{L}_{4+4\delta}$-bounded, independently distributed sequence of random variables, $\{D_{S,g}\}_{g \in \mathbb{N}}$, such that*

$$|\nabla q_{gi}(\cdot, \theta)| \leq D_{S,g}, \quad \theta \in \Theta_0,\ i \in \mathbb{I}_g,\ g \in \mathbb{N},$$

and

$$\begin{aligned}&|\nabla q_{gi}(X_{gi}, \theta_2) - \nabla q_{gi}(X_{gi}, \theta_1)| \leq D_{S,g} h(|\theta_2 - \theta_1|), \\ &\qquad\qquad\qquad\qquad (\theta_1, \theta_2) \in \Theta_0^2,\ i \in \mathbb{I}_g,\ g \in \mathbb{N}.\end{aligned}$$

Proof of LEMMA B.4: The result can be obtained by repeating the proof of Lemma B.2, replacing $\nabla^2 q_g$ with ∇q_g and $2 + 2\delta$ with $4 + 4\delta$. $\square$

LEMMA B.5 *Suppose that Assumptions 1(a)(b) and 4(d) hold. For each $(g, G) \in \mathbb{G}$, define $\check{r}_{Gg} : \Omega \times \Theta \to \mathbb{R}^p$ by*

$$\check{r}_{Gg}(\cdot, \theta) \equiv (G/N_G) \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta), \quad \theta \in \Theta,\ (g, G) \in \mathbb{G}.$$

Then there exists an $\mathscr{L}_{4+4\delta}$-bounded, independently distributed sequence of random variable, $\{D_{r,g}\}_{g\in\mathbb{N}}$, such that

$$|\check{r}_{Gg}(\cdot,\theta)| \le D_{r,g}, \quad \theta \in \Theta_0,\ (g,G)\in\mathbb{G},$$

and for each $(\theta_1,\theta_2)\in\Theta_0^2$,

$$|\check{r}_{Gg}(\cdot,\theta_2) - \check{r}_{Gg}(\cdot,\theta_1)| \le D_{r,g} h(|\theta_2-\theta_1|), \quad (g,G)\in\mathbb{G}.$$

Proof of Lemma B.5: For each $g\in\mathbb{N}$, set $D_{r,g} \equiv (\bar{n}/2)D_{S,g}$, where $D_{S,g}$ is as in Lemma B.4. Because $\{D_{S,g}\}_{g\in\mathbb{N}}$ is a $\mathscr{L}_{4+4\delta}$-bounded, independent sequence, so is $\{D_{r,g}\}_{g\in\mathbb{N}}$. The array $\{D_{r,g}\}$ satisfies that

$$|\check{r}_{Gg}(\cdot,\theta)| \le \frac{n_g G}{N_G} n_g^{-1}\sum_{i=1}^{n_g} |\nabla q_g(X_{gi,\theta})| \le \frac{\bar{n}}{2} D_{S,g} = D_{r,g}, \quad (g,G)\in\mathbb{G},$$

and for each $(\theta_1,\theta_2)\in\Theta_0^2$,

$$\begin{aligned}|\check{r}_{Gg}(\cdot,\theta_2) - \check{r}_{Gg}(\cdot,\theta_2)| &\le \frac{n_g G}{N_G} n_g^{-1}\sum_{i=1}^{n_g} |\nabla q_g(X_{gi,\theta_2}) - \nabla q_g(X_{gi,\theta_1})| \\ &\le \frac{\bar{n}}{2} D_{S,g} h(|\theta_2-\theta_1|) = D_{S,g} h(|\theta_2-\theta_1|), \quad (g,G)\in\mathbb{G}.\end{aligned}$$

The desired result therefore follows. $\square$

Lemma B.6 *Suppose that Assumptions 1(a)(b)(d) and 4(d) hold. Then*

$$G^{1/2}\nabla Q_G(\cdot,\theta_G^*) = G^{-1/2}\sum_{g=1}^{G} r_{Gg}^* = \mathrm{O}_P(1),$$

where for each $(g,G)\in\mathbb{G}$, $r_{Gg}^* \equiv \check{r}_{Gg}(\cdot,\theta_G^*)$, and $\{\check{r}_{Gg} : (g,G)\in\mathbb{G}\}$ is as in Lemma B.5.

Proof of Lemma B.6: The first equality immediately follows from the definitions of $\{Q_G\}_{G\in\mathbb{N}}$ and $\{\check{r}_{Gg} : (g,G)\in\mathbb{G}\}$. Because $\{r_{Gg}^* : (g,G)\in\mathbb{G}\}$ is an $\mathscr{L}_{4+4\delta}$-bounded, rowwise independent array by Lemma B.5, the second equality follows from Lemma A.5(a). $\square$

Lemma B.7 *Suppose that Assumptions 1(a)(b)(d), 3(a), and 4 hold. Then*

$$\begin{aligned} G^{1/2}(\hat{\theta}_G - \theta^*_G) &= -A^{*}_G{}^{-1} G^{-1/2} \sum_{g=1}^{G} \frac{G}{N_G} \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta^*_G) + \mathrm{o}_P(1) \\ &= \mathrm{O}_P(1). \end{aligned}$$

Proof of Lemma B.7: Given Lemmas B.3 and B.6, the desired results can be established by using the standard linearization technique often employed in derivation of the asymptotic normality of M-estimators (see, e.g., the proof of White (1994, Theorem 6.10)). $\square$

Lemma B.8 *Suppose that Assumptions 1(a)(b) and 4(b)(d) hold. Define* $\{\check{L}_G : \Omega \times \Theta \times \mathbb{S}^m \to \mathbb{R}^{m\times p}\}_{G\in\mathbb{N}}$ *by*

$$\begin{aligned} \check{L}_G(\cdot, \theta, W) &\equiv \nabla_\theta \check{\alpha}(\cdot, \theta, W) \\ &= G^{-1} \sum_{g=1}^{G} \check{l}_{Gg}(\cdot, \theta, W), \quad \theta \in \Theta,\, W \in \mathbb{S}^m,\, G \in \mathbb{N}, \end{aligned}$$

where $\{\check{l}_{Gg} : \Omega \times \Theta \times \mathbb{S}^m \to \mathbb{R}^p :\ (g, G) \in \mathbb{G}\}$ *is defined by*

$$\begin{aligned} \check{l}_{Gg}(\cdot, \theta, W) \equiv \frac{2G}{N_G}\Big(&-\frac{1}{n_g - 1} \sum_{i=1}^{n_g} \nabla s_{gi}(\cdot, \theta) W s_{gi}(\cdot, \theta) \\ &+ \frac{n_g^2}{n_g - 1} \nabla \tilde{s}_{gi}(\cdot, \theta) W \tilde{s}_{gi}(\cdot, \theta)\Big), \quad \theta \in \Theta,\, W \in \mathbb{R}^{m\times m},\, G \in \mathbb{N}. \end{aligned}$$

Then for each bounded sequence $\{\bar{W}_G \in \mathbb{S}^m\}_{G\in\mathbb{N}}$*, it holds that* $\{L_G(\cdot, \bar{W}_G) : \Theta \to \mathbb{R}^{m\times p}\}_{G\in\mathbb{N}}$ *is bounded and equicontinuous uniformly on* Θ_0 *and that* $\{\check{L}_G(\cdot, \theta, \bar{W}_G) - L_G(\theta, \bar{W}_G)\}_{G\in\mathbb{N}}$ *converges to zero uniformly in* $\theta \in \Theta_0$ *in probability-*P*.*

Proof of Lemma B.8: We apply Lemma A.4(b)(c), taking for $\{\check{l}_{Gg} :\ (g, G) \in \mathbb{G}\}$ for $\{F_{Gg}\}$. The set Θ_0 is a compact subset of the p-dimensional Euclidean space. For each $\theta \in \Theta_0$, the array $\{\check{a}_{Gg}(\cdot,\theta)\}$ is rowwise independent. Let $\{D_{H,g}\}_{g\in\mathbb{N}}$ be as in Lemma B.2 and $\{D_{S,g}\}_{g\in\mathbb{N}}$ as in Lemma B.4. Also, let $c \equiv \sup\{|\bar{W}_G| : G \in \mathbb{N}\}$. Then, by Lemma A.2, it follows from B.2 and B.4 that

$$\begin{aligned}
&\left|\frac{1}{n_g-1} n_g^{-1} \sum_{i=1}^{n_g} \nabla s_{gi}(\cdot,\theta)\bar{W}_G s_{gi}(\cdot,\theta)\right| \\
&\qquad \le \frac{1}{n_g-1} n_g^{-1} \sum_{i=1}^{n_g} |\nabla s_{gi}(\cdot,\theta)\bar{W}_G s_{gi}(\cdot,\theta)| \le c D_{H,g} D_{S,g}, \\
&\qquad \theta \in \Theta_0,\ (g,G)\in\mathbb{G} \quad \text{and}
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\left|\frac{1}{n_g-1} n_g^{-1} \sum_{i=1}^{n_g} \nabla s_{gi}(\cdot,\theta_2)\bar{W}_G s_{gi}(\cdot,\theta_2)\right. \\
&\qquad \left. - \frac{1}{n_g-1} n_g^{-1} \sum_{i=1}^{n_g} \nabla s_{gi}(\cdot,\theta_1)\bar{W}_G s_{gi}(\cdot,\theta_1)\right| \\
&\le \frac{1}{n_g-1} n_g^{-1} \sum_{i=1}^{n_g} |\nabla s_{gi}(\cdot,\theta_2)\bar{W}_G s_{gi}(\cdot,\theta_2) - \nabla s_{gi}(\cdot,\theta_1)\bar{W}_G s_{gi}(\cdot,\theta_1)| \\
&\le 2c(D_{H,g}+D_{S,g})h(|\theta_2-\theta_1|), \quad (\theta_1,\theta_2)\in\Theta_0^2,\ (g,G)\in\mathbb{G}.
\end{aligned} \tag{13}$$

Analogously, we can show that

$$\begin{aligned}
&\left|\frac{n_g}{n_g-1}\nabla\tilde{s}_g(\cdot,\theta)\bar{W}_G\tilde{s}_g(\cdot,\theta)\right| \le \frac{n_g}{n_g-1}|\nabla\tilde{s}_g(\cdot,\theta)\bar{W}_G\tilde{s}_g(\cdot,\theta)| \\
&\qquad \le 2cD_{H,g}D_{S,g}, \qquad \theta\in\Theta_0,\ (g,G)\in\mathbb{G} \quad \text{and}
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\left|\frac{n_g}{n_g-1}\nabla\tilde{s}_g(\cdot,\theta_2)\bar{W}_G\tilde{s}_g(\cdot,\theta_2) - \frac{n_g}{n_g-1}\nabla\tilde{s}_g(\cdot,\theta_1)\bar{W}_G\tilde{s}_g(\cdot,\theta_1)\right| \\
&\qquad \le \frac{n_g}{n_g-1}|\nabla\tilde{s}_g(\cdot,\theta_2)\bar{W}_G\tilde{s}_g(\cdot,\theta_2) - \nabla\tilde{s}_g(\cdot,\theta_1)\bar{W}_G\tilde{s}_g(\cdot,\theta_1)| \\
&\qquad \le 4c(D_{H,g}+D_{S,g})h(|\theta_2-\theta_1|), \quad (\theta_1,\theta_2)\in\Theta_0^2,\ (g,G)\in\mathbb{G}.
\end{aligned} \tag{15}$$

For each $g \in \mathbb{N}$, let $D_{l,Gg} \equiv \bar{n}c(3D_{H,g}D_{S,g} + 6D_{H,g} + 6D_{S,g})$. Because

$$\check{l}_{Gg}(\cdot,\theta,W) = \frac{2n_g G}{N_G}\bigg(- \frac{1}{n_g - 1} n_g^{-1} \sum_{i=1}^{n_g} \nabla s_{gi}(\cdot,\theta)\bar{W}_G s_{gi}(\cdot,\theta) + \frac{n_g}{n_g - 1} \nabla \tilde{s}_{gi}(\cdot,\theta)\bar{W}_G \tilde{s}_{gi}(\cdot,\theta)\bigg),$$

it follows from Lemma B.1 and (12)–(15) that for each $(g,G) \in \mathbb{G}$,

$$|\check{l}_{Gg}(\cdot,\theta,\bar{W}_G)| \le D_{l,Gg}, \quad \theta \in \Theta_0 \quad \text{and}$$

$$|\check{l}_{Gg}(\cdot,\theta_2,\bar{W}_G) - \check{l}_{Gg}(\cdot,\theta_1,\bar{W}_G)| \le D_{l,Gg}\, h(|\theta_2 - \theta_1|), \quad (\theta_1,\theta_2) \in \Theta_0^2.$$

Because both $D_{H,g}$ and $D_{S,g}$ are functions of $X_{g1}, \ldots, X_{gn_g}$, the process $\{(D_{H,g}, D_{S,g})\}_{g\in\mathbb{N}}$ is independent, so is $\{D_{l,Gg}\}_{g\in\mathbb{N}}$. Further, by Lemma A.1, we have that

$$\|D_{l,Gg}\|_{1+\delta} \le \bar{n}c(3\|D_{H,g}\|_{2+2\delta}\,\|D_{S,g}\|_{2+2\delta} + 6\|D_{H,g}\|_{1+\delta} + 6\|D_{S,g}\|_{1+\delta}).$$

Since $\{D_{H,g}\}$ is uniformly $\mathscr{L}_{2+2\delta}$-bounded, and $\{D_{S,g}\}$ is uniformly $\mathscr{L}_{4+4\delta}$-bounded, the right-hand side of the above inequality is bounded over $g \in \mathbb{N}$. It follows that $\{D_{l,Gg}\}$ is uniformly $\mathscr{L}_{1+\delta}$-bounded.

Thus, Lemma A.4 applies to $\{\check{l}_{Gg},\, (g,G) \in \mathbb{G}\}$. Because for each $\theta \in \Theta$ and each $G \in \mathbb{N}$,

$$L_G(\theta,\bar{W}_G) = \mathrm{E}[\check{L}_G(\cdot,\theta,\bar{W}_G)] = L(\theta,\bar{W}_G),$$

the desired results follow. $\square$

Lemma B.9 *Suppose that Assumptions 1(a)(b) and 4(d) hold. Define an array* $\{\check{\zeta}_{Gg} : \Omega \times$

$\Theta \times \mathbb{S}^m \times \mathbb{R}^{p\times p} \times \mathbb{R}^m \to \mathbb{R} : (g, G) \in \mathbb{G}\}$ *by*

$$\begin{aligned}\check{\zeta}_{Gg}(\cdot, \theta, W, B, L) \equiv \frac{G}{N_G}\Bigg(&\sum_{i=1}^{n_g} s_{gi}(\cdot, \theta)' W s_{gi}(\cdot, \theta) \\ &- \frac{n_g}{n_g - 1} \sum_{i=1}^{n_g} (s_{gi}(\cdot, \theta) - \tilde{s}_g(\cdot, \theta))' W (s_{gi}(\cdot, \theta) - \tilde{s}_g(\cdot, \theta)) \\ &- L' B \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta) \Bigg),\end{aligned}$$

$(\theta, W, B, L) \in \Theta \times \mathbb{S}^m \times \mathbb{R}^{p\times p} \times \mathbb{R}^m$, $(g, G) \in \mathbb{G}$. *Let* $\mathbb{W}$, $\mathbb{B}$, *and* $\mathbb{L}$ *be compact subsets of* $\mathbb{S}^m$, $\mathbb{R}^{p\times p}$, *and* $\mathbb{R}^m$, *respectively. Then there exists an* $\mathscr{L}_{2+2\delta}$*-bounded, independently distributed sequence of random variables* $\{D_{\zeta,g}\}_{g\in\mathbb{N}}$ *and a continuous function* $h^\zeta : \mathbb{R} \to \mathbb{R}$ *such that* $h^\zeta(y) \downarrow 0$ *as* $y \downarrow 0$ *that satisfy that*

$$|\check{\zeta}_{Gg}(\cdot, \theta, W, B, L)| \leq D_{\zeta,g}, \quad (\theta, W, B, L) \in \Theta_0 \times \mathbb{W} \times \mathbb{B} \times \mathbb{L}, \ (g, G) \in \mathbb{G} \tag{16}$$

and for each pair $(\theta_1, W_1, B_1, L_1)$ *and* $(\theta_2, W_2, B_2, L_2)$ *in* $\Theta_0 \times \mathbb{W} \times \mathbb{B} \times \mathbb{L}$

$$\begin{aligned}&|\check{\zeta}_{Gg}(\cdot, \theta_2, W_2, B_2, L_2) - \check{\zeta}_{Gg}(\cdot, \theta_1, W_1, B_1, L_1)| \\ &\qquad \leq D_{\zeta,g}\, h^\zeta(|\theta_2 - \theta_1| + |W_2 - W_1| + |B_2 - B_1| + |L_2 - L_1|), \\ &\qquad (g, G) \in \mathbb{G}.\end{aligned} \tag{17}$$

Proof of Lemma B.9: Note that $\check{\zeta}_{Gg}$ can be written as

$$\begin{aligned}\check{\zeta}_{Gg}(\cdot, \theta, W, B, L) \equiv \frac{n_g G}{N_G}\Bigg(&- \frac{1}{n_g - 1} n_g^{-1} \sum_{i=1}^{n_g} s_{gi}(\cdot, \theta)' W s_{gi}(\cdot, \theta) \\ &+ \frac{n_g}{n_g - 1} \tilde{s}_g(\cdot, \theta)' W \tilde{s}_g(\cdot, \theta) - L' B n_g^{-1} \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta) \Bigg), \\ &(\theta, W, B, L) \in \Theta \times \mathbb{S}^m \times \mathbb{R}^{p\times p} \times \mathbb{R}^m, \ (g, G) \in \mathbb{G}.\end{aligned} \tag{18}$$

By Lemma B.1, $\{n_g G/N_G : (g, G) \in \mathbb{G}\}$ is bounded. We first examine each of the three terms in the parentheses on the right-hand side in (18) and then combine the results to establish the desired result.

By Lemma B.4, we have that for each $(\theta, W) \in \Theta_0 \times \mathbb{W}$ and each $g \in \mathbb{N}$,

$$
\begin{aligned}
&\left| -\frac{1}{n_g - 1} n_g^{-1} \sum_{i=1}^{n_g} s_{gi}(\cdot, \theta)' W s_{gi}(\cdot, \theta) \right| \\
&\qquad \leq \frac{1}{n_g - 1} \, |W| \, n_g^{-1} \sum_{i=1}^{n_g} |s_{gi}(\cdot, \theta)|^2 \leq c_{\mathbb{W}} \, D_{S,g}^2, \qquad (19)
\end{aligned}
$$

where $c_{\mathbb{W}} \equiv \sup\{|W| : W \in \mathbb{W}\} < \infty$. We also have that for each $i \in \mathbb{I}_g$, each $g \in \mathbb{N}$, and each pair (θ_1, W_1) and (θ_2, W_2) in $\Theta_0 \times \mathbb{W}$,

$$
\begin{aligned}
&|s_{gi}(\cdot, \theta_2)' W_2 s_{gi}(\cdot, \theta_2) - s_{gi}(\cdot, \theta_1)' W_1 s_{gi}(\cdot, \theta_1)| \\
&\quad \leq |s_{gi}(\cdot, \theta_2)' W_2 s_{gi}(\cdot, \theta_2) - s_{gi}(\cdot, \theta_1)' W_2 s_{gi}(\cdot, \theta_1)| \\
&\quad\quad + |s_{gi}(\cdot, \theta_1)' W_2 s_{gi}(\cdot, \theta_1) - s_{gi}(\cdot, \theta_1)' W_1 s_{gi}(\cdot, \theta_1)| \\
&\quad \leq |s_{gi}(\cdot, \theta_2)' W_2 s_{gi}(\cdot, \theta_2) - s_{gi}(\cdot, \theta_1)' W_2 s_{gi}(\cdot, \theta_1)| \\
&\quad\quad + |s_{gi}(\cdot, \theta_1)' (W_2 - W_1) s_{gi}(\cdot, \theta_1)| \\
&\quad \leq 4 c_{\mathbb{W}} D_{S,g} \, h(|\theta_2 - \theta_1|) + D_{S,g}^2 \, |W_2 - W_1| \\
&\quad \leq (4 c_{\mathbb{W}} + 1)(D_{S,g}^2 + D_{S,g})(h(|\theta_2 - \theta_1|) + |W_2 - W_1|), \\
&\quad i \in \mathbb{I}_g, \; g \in \mathbb{N},
\end{aligned}
$$

where the second last inequality follows by Lemmas A.2 and B.4. It follows that for each pair (θ_1, W_1) and (θ_2, W_2) in $\Theta_0 \times \mathbb{W}$,

$$
\begin{aligned}
&\Bigg|\Big(-\frac{1}{n_g-1}n_g^{-1}\sum_{i=1}^{n_g} s_{gi}(\cdot,\theta_2)'W_2 s_{gi}(\cdot,\theta_2)\Big) \\
&\qquad - \Big(-\frac{1}{n_g-1}n_g^{-1}\sum_{i=1}^{n_g} s_{gi}(\cdot,\theta_1)'W_1 s_{gi}(\cdot,\theta_1)\Big)\Bigg| \\
&\leq \frac{1}{n_g-1}n_g^{-1}\sum_{i=1}^{n_g} |s_{gi}(\cdot,\theta_2)'W_2 s_{gi}(\cdot,\theta_2) - s_{gi}(\cdot,\theta_1)'W_1 s_{gi}(\cdot,\theta_1)| \\
&\leq (4c_{\mathbb{W}}+1)(D_{S,g}^2 + D_{S,g})(h(|\theta_2-\theta_1|) + |W_2 - W_1|), \quad g \in \mathbb{N}.
\end{aligned} \tag{20}
$$

For the second term, we can analogously derive that

$$
\left|-\frac{n_g}{n_g-1}\tilde{s}_g(\cdot,\theta)'W\tilde{s}_g(\cdot,\theta)\right| \leq 2c_{\mathbb{W}}\, D_{S,g}^2, \quad (\theta, W) \in \Theta_0 \times \mathbb{W},\, g \in \mathbb{N}, \tag{21}
$$

and that for each pair (θ_1, W_1) and (θ_2, W_2) in $\Theta_0 \times \mathbb{W}$,

$$
\begin{aligned}
&\Bigg|\Big(-\frac{n_g}{n_g-1}\tilde{s}_g(\cdot,\theta_2)'W_2\tilde{s}_g(\cdot,\theta_2)\Big) \\
&\qquad - \Big(-\frac{n_g}{n_g-1}\tilde{s}_g(\cdot,\theta_1)'W_1\tilde{s}_g(\cdot,\theta_1)\Big)\Bigg| \\
&\leq 2(4c_{\mathbb{W}}+1)(D_{S,g}^2 + D_{S,g})(h(|\theta_2-\theta_1|) + |W_2 - W_1|), \quad g \in \mathbb{N}.
\end{aligned} \tag{22}
$$

For the third term, let $c_{\mathbb{L}} \equiv \sup\{|L| : L \in \mathbb{L}\}$, $c_{\mathbb{B}} \equiv \sup\{|B| : B \in \mathbb{B}\}$, and $c_{\mathbb{LB}} \equiv c_{\mathbb{L}} + c_{\mathbb{B}} + c_{\mathbb{L}}c_{\mathbb{B}}$. Then for each $(L, B) \in \mathbb{L} \times \mathbb{B}$, $|LB| \leq c_{\mathbb{LB}}$, and for each pair (L_1, B_1) and (L_2, B_2) in $\mathbb{L} \times \mathbb{B}$,

$$
|L_2B_2 - L_1B_1| \leq c_{\mathbb{LB}}(|L_2 - L_1| + |B_2 - B_1|).
$$

It follows that for each $(\theta, B, L) \in \Theta_0 \times \mathbb{R}^{p\times p} \times \mathbb{R}^m$,

$$
\begin{aligned}
|L' B n_g^{-1} \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta)| &\leq |L'B|\, n_g^{-1} \sum_{i=1}^{n_g} |\nabla q_g(X_{gi}, \theta)| \\
&\leq c_{\mathbb{LB}} D_{S,g}, \quad g \in \mathbb{N}.
\end{aligned} \tag{23}
$$

Application of Lemma A.2 also yields that for each pair (θ_1, B_1, L_1) and (θ_2, B_2, L_2) in $\Theta_0 \times \mathbb{R}^{p\times p} \times \mathbb{R}^m$,

$$
\begin{aligned}
&|L_2 B_2 n_g^{-1} \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta_2) - L_1 B_1 n_g^{-1} \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta_1)| \\
&\quad \leq (c_{\mathbb{LB}} + D_{S,g})(h(|\theta_2 - \theta_1|) + |L_2 - L_1| + |B_2 - B_1|), \quad g \in \mathbb{N}.
\end{aligned} \tag{24}
$$

Now, set

$$
D_{\zeta,g} \equiv \frac{\bar{n}}{2}(3(4c_{\mathbb{W}} + 1)(D_{S,g}^2 + D_{S,g}) + c_{\mathbb{LB}} D_{S,g} + c_{\mathbb{LB}} + D_{S,g}), \quad g \in \mathbb{N}.
$$

and

$$
h^\zeta(y) \equiv \begin{cases} 3 \sup\{h(z) :\, z \in [0, y]\} + 3y & \text{if } y \geq 0, \\ 0 & \text{otherwise}, \quad y \in \mathbb{R}. \end{cases}
$$

Then $\{D_{\zeta,g}\}_{g\in\mathbb{N}}$ is an uniformly $\mathscr{L}_{2+2\delta}$-bounded, independently distributed sequence, and h^ζ satisfies that $h^\zeta(y) \downarrow 0$ as $y \downarrow 0$. Also, by using (19)–(24), we can verify that $\{D_{\zeta,g}\}_{g\in\mathbb{N}}$ and h^ζ satisfy (16) and (17). The result therefore follows. $\square$

Lemma B.10 *Suppose that Assumptions 1, 3(a), and 4 hold. Suppose that* $\{\bar{W}_G \in \mathbb{S}^m\}_{G\in\mathbb{N}}$

is bounded. Then

$$\begin{aligned}
&\check{\alpha}_G(\cdot,\hat{\theta}_G,\bar{W}_G)-\alpha_G(\theta^*_G,\bar{W}_G)\\
&=G^{-1}\sum_{g=1}^{G}\check{\zeta}_{Gg}(\cdot,\theta^*_G,\bar{W}_G,A^{*\,-1}_G,L^*_G)-\alpha_G(\theta^*_G,\bar{W}_G)+\mathrm{o}_P(G^{-1/2}) &(25)\\
&=\mathrm{O}_P(G^{-1/2}), &(26)
\end{aligned}$$

where $\{\check{\zeta}_{Gg}:(g,G)\in\mathbb{G}\}$ *is as in Lemma B.9.*

Proof of Lemma B.10: By definition of $\{\check{\zeta}_{Gg}:(g,G)\in\mathbb{G}\}$, we have that for each *Grange*,

$$\begin{aligned}
&G^{-1}\sum_{g=1}^{G}\check{\zeta}_{Gg}(\cdot,\theta^*_G,\bar{W}_G,A^{*\,-1}_G,L^*_G)\\
&\quad=\check{\alpha}_G(\cdot,\theta^*_G,\bar{W}_G)-L^{*\prime}_G A^{*\,-1}_G G^{-1}\sum_{g=1}^{G}\frac{G}{N_G}\sum_{i=1}^{n_g}\nabla q_g(X_{gi},\theta^*_G).
\end{aligned}$$

By subtracting $\alpha_G(\theta^*_G,\bar{W}_G)$ from both sides of this equality and rearranging the resulting equality yields that for each $G\in\mathbb{N}$,

$$\begin{aligned}
\check{\alpha}_G(\cdot,\theta^*_G,\bar{W}_G)-\alpha_G(\theta^*_G,\bar{W}_G)&=G^{-1}\sum_{g=1}^{G}\check{\zeta}_{Gg}(\cdot,\theta^*_G,\bar{W}_G,A^{*\,-1}_G,L^*_G)\\
&\quad-\alpha_G(\theta^*_G,\bar{W}_G)+L^{*\prime}_G A^{*\,-1}_G G^{-1}\sum_{g=1}^{G}\frac{G}{N_G}\sum_{i=1}^{n_g}\nabla q_g(X_{gi},\theta^*_G).
\end{aligned}$$

It follows that for each $G \in \mathbb{N}$,

$$\begin{aligned}
&\breve{\alpha}_G(\cdot, \hat{\theta}_G, \bar{W}_G) - \alpha_G(\theta_G^*, \bar{W}_G) \\
&= (\breve{\alpha}_G(\cdot, \hat{\theta}_G, \bar{W}_G) - \breve{\alpha}_G(\cdot, \theta^*, \bar{W}_G)) + (\breve{\alpha}_G(\cdot, \theta^*, \bar{W}_G) - \alpha_G(\theta_G^*, \bar{W}_G)) \\
&= \left(G^{-1} \sum_{g=1}^{G} \breve{\zeta}_{Gg}(\cdot, \theta_G^*, \bar{W}_G) - \alpha_G(\theta_G^*, \bar{W}_G) \right) \\
&\quad + (\breve{\alpha}_G(\cdot, \hat{\theta}_G, \bar{W}_G) - \breve{\alpha}_G(\cdot, \theta^*, \bar{W}_G)) \\
&\quad + L_G^{*\prime} A_G^{*-1} G^{-1} \sum_{g=1}^{G} \frac{G}{N_G} \sum_{i=1}^{n_g} \nabla q_g(X_{gi}, \theta_G^*). \qquad (27)
\end{aligned}$$

To prove the equality in (25), it thus suffices to prove that in the above equality, the sum of the second and third terms on the right-hand side is $\mathrm{o}_P(G^{-1/2})$. Because $\{\theta_G^*\}$ is uniformly interior to Θ_0, there exists a real number $\epsilon > 0$ such that the open ball with radius ϵ centered at θ_G^* is contained in $\operatorname{int} \Theta_0$ for each $G \in \mathbb{N}$. By the mean value theorem for random functions (Jennrich, 1969, Lemma 3), there exists a sequence of random vectors $\{\ddot{\theta}_G : \Omega \to \Theta\}_{G \in \mathbb{N}}$ such that for each $G \in \mathbb{N}$, $\ddot{\theta}_G$ is on the line segment connecting $\hat{\theta}_G$ and θ_G^*, and

$$\breve{\alpha}_G(\cdot, \hat{\theta}_G, \bar{W}_G) - \breve{\alpha}_G(\cdot, \theta_G^*, \bar{W}_G) = \check{L}_G(\cdot, \ddot{\theta}_G, \bar{W}_G)'(\hat{\theta}_G - \theta_G^*)$$

whenever $|\hat{\theta}_G - \theta_G^*| < \epsilon$ (where ϵ is as described above). By Lemma B.8, $\{\check{L}_G(\cdot, \theta, \bar{W}_G) - L_G(\theta, \bar{W}_G)\}_{G \in \mathbb{N}}$ converges to zero uniformly in $\theta \in \Theta_0$ in probability-P, and $\{L_G(, \bar{W}_G) : \Theta \to \mathbb{R}^{m \times p}\}_{G \in \mathbb{N}}$ is equicontinuous uniformly on Θ_0. Also, we have that

$$|\ddot{\theta}_G - \theta_G^*| \le |\hat{\theta}_G - \theta_G^*| = \mathrm{O}_P(G^{-1/2})$$

by Lemma B.7. It follows that $|\check{L}_G(\cdot, \ddot{\theta}_G, \bar{W}_G) - L_G^*| = \mathrm{o}_P(1)$, and

$$\breve{\alpha}_G(\cdot, \hat{\theta}_G, \bar{W}_G) - \breve{\alpha}_G(\cdot, \theta_G^*, \bar{W}_G) = L_G^*(\hat{\theta}_G - \theta_G^*) + \mathrm{o}_P(G^{-1/2}).$$

Applying the first equality of Lemma B.7 in this equality establishes that the sum of the second and third terms on the right-hand side of (27) is $o_P(G^{-1/2})$.

To prove the second equality, note that

$$\mathrm{E}\left[G^{-1}\sum_{g=1}^{G}\check{\zeta}_{Gg}(\cdot,\theta_G^*,\bar{W}_G,A_G^{*-1},L_G^*)-\alpha_G(\theta_G^*,\bar{W}_G)\right]=0.$$

By Lemma B.9,

$$\{\check{\zeta}_{Gg}(\cdot,\theta_G^*,\bar{W}_G,A_G^{*-1},L_G^*)-\alpha_G(\theta_G^*,\bar{W}_G):\ (g,G)\in\mathbb{G}\}$$

is an uniformly $\mathscr{L}_{2+\delta}$-bounded, independently distributed array. Applying Lemma A.5(a) to it proves that the first term in (25) is $\mathrm{O}_P(G^{-1/2})$. The equality in (26) therefore follows. □

Lemma B.11 *Suppose that Assumptions 1, 3(a), and 4 hold. Then*

$$\hat{S}_G-\hat{\Sigma}_G-N_G^{-1}\sum_{g=1}^{G}n_g\mathrm{E}[s_{Gg1}^*]\mathrm{E}[s_{Gg1}^*]'=\mathrm{O}_P(G^{-1/2}),$$

and

$$\hat{S}_G-\hat{\Sigma}_G=\mathrm{O}_P(1).$$

Proof of Lemma B.11: Application of Lemma B.10 taking for $\bar{W}_G$ the $m\times m$ matrix such that its (i,j)- and (j,i)-elements are ones, and all other elements are zeros for each $(i,j)\in\{1,2,\ldots,m\}^2$ establishes the first equality.

For the second equality, note that

$$\begin{aligned}\left|N_G^{-1}\sum_{g=1}^{G}n_g\mathrm{E}[s_{Gg1}^*]\mathrm{E}[s_{Gg1}^*]'\right| &\le G^{-1}\sum_{g=1}^{G}\frac{Gn_g}{N_G}|\mathrm{E}[s_{Gg1}^*]|^2\\ &\le G^{-1}\sum_{g=1}^{G}\frac{\bar{n}}{2}\mathrm{E}[|s_{Gg1}^*|]^2=\frac{\bar{n}}{2}G^{-1}\sum_{g=1}^{G}\|s_{Gg1}^*\|_2^2,\quad G\in\mathbb{N}.\end{aligned}$$

As $\{s_{Gg1} : \theta \in \Theta_0,\, (g, G) \in \mathbb{G}\}$ is uniformly $\mathscr{L}_{4+\delta}$-bounded by Lemma B.4, it holds that

$$\sup_{G\in\mathbb{N}} \left| N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[s^*_{Gg1}]\mathrm{E}[s^*_{Gg1}]' \right| < \infty.$$

The desired result therefore follows. □

Lemma B.12 *Suppose that Assumptions 1 and 2(b) hold. Then*

$$\left| N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[s^*_{Gg1}]\mathrm{E}[s^*_{Gg1}]' \right| = O(\alpha^*_G).$$

Proof of Lemma B.12: Let $\bar{\lambda}$ denote the infimum of the minimum eigenvalue of W_G taken over $G \in \mathbb{N}$. Then $\bar{\lambda}$ is positive, and we have that

$$\alpha^*_G = N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[s^*_{Gg1}]' W_G \mathrm{E}[s^*_{Gg1}] \geq \bar{\lambda} N_G^{-1} \sum_{g=1}^{G} n_g |\mathrm{E}[s^*_{Gg1}]|^2, \quad G \in \mathbb{N}.$$

It follows that

$$\begin{aligned} \left| N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[s^*_{Gg1}]\mathrm{E}[s^*_{Gg1}]' \right| &\leq N_G^{-1} \sum_{g=1}^{G} n_g |\, \mathrm{E}[s^*_{Gg1}]\mathrm{E}[s^*_{Gg1}]' \,| \\ &\leq N_G^{-1} \sum_{g=1}^{G} n_g |\, \mathrm{E}[s^*_{Gg1}] \,|^2 \leq \bar{\lambda}^{-1} \alpha^*_G. \end{aligned}$$

This establishes the desired result. □

Lemma B.13 *Suppose that Assumptions 1, 2(a), 3, and 4 hold and that* $\alpha^*_G = O(G^{-1/2})$. *Then*

$$G^{1/2}(\hat{\alpha}_G - \check{\alpha}_G(\cdot, \hat{\theta}_G, W_G)) = \mathrm{o}_P(1).$$

Proof of Lemma B.13: Because $N_G^{-1} \sum_{g=1}^{G} n_g \mathrm{E}[s^*_{Gg1}]\mathrm{E}[s^*_{Gg1}]' = O(\alpha^*_G) = O(G^{-1/2})$ under the current assumption by Lemma B.12, it follows from Lemma B.11 that $|\hat{S}_G - \hat{\Sigma}_G| =$

$\mathrm{O}_P(G^{-1/2})$. Given this, we obtain that

$$|G^{1/2}(\hat{\alpha}_G - \check{\alpha}_G(\cdot, \hat{\theta}_G, W_G))| = \left|\mathrm{tr}\Big((\hat{W}_G - W_G)G^{1/2}(\hat{S}_G - \hat{\Sigma}_G)\Big)\right|$$
$$\leq m\,|\hat{W}_G - W_G|\,G^{1/2}|\hat{S}_G - \hat{\Sigma}_G| = \mathrm{o}_P(1).$$

The result therefore follows. $\square$

Proof of THEOREM 3.1: The difference between $\hat{\alpha}_G$ and α_G^* can be decomposed as

$$\hat{\alpha}_G - \alpha_G^* = (\hat{\alpha}_G - \check{\alpha}_G(\cdot, \hat{\theta}_G, W_G)) + (\check{\alpha}_G(\cdot, \hat{\theta}_G, W_G) - \alpha_G^*)$$
$$= \mathrm{tr}((\hat{W}_G - W_G)(\hat{S}_G - \hat{\Sigma}_G))$$
$$- \,\mathrm{tr}\left(W_G\left(\hat{S}_G - \hat{\Sigma}_G - N_G^{-1}\sum_{g=1}^{G} n_g \mathrm{E}[s_{Gg1}^*]\mathrm{E}[s_{Gg1}^*]'\right)\right), \quad G \in \mathbb{N}.$$

It follows from Assumption 3(b) and Lemma B.11 that the first term on the right-hand side in this equality converges to zero in probability-P. Also, it follows from Assumption 2(a) and Lemma B.11 that the second term, being $\mathrm{O}_P(G^{-1/2})$, is $\mathrm{o}_P(1)$. Claim (a) therefore follows.

To prove claim (b), note that

$$G^{1/2}(\hat{\alpha}_G - \alpha_G^*) = G^{1/2}(\hat{\alpha}_G - \check{\alpha}_G(\cdot, \hat{\theta}_G, W_G))$$
$$+ \, G^{1/2}(\check{\alpha}_G(\cdot, \hat{\theta}_G, W_G) - \alpha_G^*), \quad G \in \mathbb{N}.$$

In this equality, application of Lemma B.13 shows that the first term on the right-hand side is $\mathrm{o}_P(1)$, given that $\alpha_G^* = O(G^{-1/2})$. Applying Lemma B.10, setting $\bar{W}_G = W_G$, in the second term on the right-hand side of the above equality then establishes (6), from which (7) follows by Lemma A.5 and the asymptotic equivalence lemma. $\square$

Proof of THEOREM 3.2: Note that for each $(g, G) \in \mathbb{G}$,

$$\hat{\xi}_{Gg} = \check{\zeta}_{Gg}(\cdot, \hat{\theta}_G, \hat{W}_G, \hat{A}_G^+, \hat{L}_G).$$

By Assumption 2(a), there exists a compact subset $\mathbb{W}$ of $\mathbb{S}^m$, to which $\{W_G\}_{G\in\mathbb{N}}$ is uniformly interior. By Lemma B.3 and Assumption 4(c), there also exists a compact subset $\mathbb{B}$ of $\mathbb{R}^{p\times p}$, to which $\{A_G^{*-1}\}_{G\in\mathbb{N}}$ is uniformly interior. Further, by Lemma B.8, there exists a compact subset $\mathbb{L}$ of $\mathbb{R}^p$, to which $\{L_G^*\}_{G\in\mathbb{N}}$ is uniformly interior. By Lemma B.9 and A.2, there exist a $\mathscr{L}_{2+\delta}$-bounded, independently distributed sequence $\{D_{\zeta,g}\}_{g\in\mathbb{N}}$ and a continuous function $h^\zeta : \mathbb{R} \to \mathbb{R}$ such that $h^\zeta(y) \downarrow 0$ as $y \downarrow 0$ that satisfy that

$$\check{\zeta}_{Gg}(\cdot,\theta,W,B,L)^2 \le D_{\zeta,g}^2, \quad (\theta,W,B,L) \in \Theta_0 \times \mathbb{W} \times \mathbb{B} \times \mathbb{L}, (g,G) \in \mathbb{G},$$

and for each pair (θ_1,W_1,B_1,L_1) and (θ_2,W_2,B_2,L_2) in $\Theta_0 \times \mathbb{W} \times \mathbb{B} \times \mathbb{L}$,

$$\begin{aligned}&|\check{\zeta}_{Gg}(\cdot,\theta_2,W_2,B_2,L_2) - \check{\zeta}_{Gg}(\cdot,\theta_1,W_1,B_1,L_1)|\\ &\quad \le D_{\zeta,g}^2 h^\zeta(|\theta_2-\theta_1| + |W_2-W_1| + |B_2-B_1| + |L_2-L_1|),\end{aligned}$$

$(g,G) \in \mathbb{G}$. It follows that Lemma A.4 applies to $\{\check{\zeta}_{Gg}^2 : (g,G) \in \mathbb{G}\}$. Thus,

$$\Big\{(\theta,W,B,L) \mapsto G^{-1}\sum_{g=1}^{G} \mathrm{E}[\zeta_{Gg}(\cdot,\theta,W,B,L)^2] : \\ \Theta \times \mathbb{S}^m \times \mathbb{R}^{p\times p} \times \mathbb{R}^p \to \mathbb{R}\Big\}_{G\in\mathbb{N}}$$

is equicontinuous uniformly on $\Theta_0 \times \mathbb{W} \times \mathbb{B} \times \mathbb{L}$, and

$$\Big\{G^{-1}\sum_{g=1}^{G} \zeta_{Gg}(\cdot,\theta,W,B,L)^2 - G^{-1}\sum_{g=1}^{G} \mathrm{E}[\zeta_{Gg}(\cdot,\theta,W,B,L)^2]\Big\}_{G\in\mathbb{N}}$$

converges to zero uniformly in $(\theta,W,B,L) \in \Theta_0 \times \mathbb{W} \times \mathbb{B} \times \mathbb{L}$ in probability-P. Further, $\{\hat{\theta}_G - \theta_G^*\}_{G\in\mathbb{N}}$, $\{\hat{A}_G^+ - A_G^{*-1}\}_{G\in\mathbb{N}}$, $\{\hat{W}_G - W_G\}_{G\in\mathbb{N}}$, and $\{\hat{L}_G - L_G^*\}_{G\in\mathbb{N}}$ converge to zero in probability-P by Assumption 3 and Lemmas B.3, B.8. It follows by Lemma A.4 that

$$\Big\{\hat{V}_G - \bar{V}_G = G^{-1}\sum_{g=1}^{G}\zeta_{Gg}(\cdot,\hat{\theta},\hat{W}_G,\hat{A}_G^{-1},\hat{L}_G)^2 - G^{-1}\sum_{g=1}^{G}\mathrm{E}[\zeta_{Gg}(\cdot,\theta_G^*,W_G,A_G^{*\,-1},L_G^*)^2]\Big\}_{G\in\mathbb{N}}$$

converges to zero in probability-P, and $\{\bar{V}_G\}_{G\in\mathbb{N}}$ is bounded. Claim (a) therefore follows.

For (b), note that

$$|\bar{V}_G - V_G| = \left|G^{-1}\sum_{g=1}^{G}(Gn_g/N_G)^2(\mathrm{E}[s_{Gg1}]'W_G\mathrm{E}[s_{Gg1}])^2\right|.$$

Recall that if a sequence of random variables converges in the mean to zero, it also converges in probability; and, further, if the squared sequence is uniformly integrable (in particular, if the original sequence is uniformly bounded), the squared sequence converges in the mean to zero. In our current problem, α_G^*, which is the average $(Gn_g/N_G)\mathrm{E}[s_{Gg1}^*]'W_G\mathrm{E}[s_{Gg1}^*]$ over the first G groups (the mean taken in terms of the probability measure assigning probability $1/G$ to each g), converges to zero as $G\to\infty$, and $\{(Gn_g/N_G)\mathrm{E}[s_{Gg1}^*]'W_G\mathrm{E}[s_{Gg1}^*] : (g,G)\in\mathbb{G}\}$ is uniformly bounded, the average of $(Gn_g/N_G)^2(\mathrm{E}[s_{Gg1}^*]'W_G\mathrm{E}[s_{Gg1}^*])^2$ over the first G groups also converges to zero. Thus, $\{|\bar{V}_G - V_G|\}_{G\in\mathbb{N}}$ converges to zero. It follows by (a) of the current theorem that

$$|\hat{V}_G - V_G| \le |\hat{V}_G - \bar{V}_G| + |\bar{V}_G - V_G| \to 0 \text{ in probability-}P,$$

so that $|\hat{V}_G - V_G| \to 0$ in probability-P. Given this first result, we further have that

$$\mathcal{T}_G - \frac{G^{1/2}\hat{\alpha}_G}{V_G^{1/2}} = (\hat{V}_G^{-1/2} - V_G^{-1/2})G^{1/2}\hat{\alpha}_G = \mathrm{o}_P(1),$$

because $G^{1/2}\hat{\alpha}_G = \mathrm{O}_P(1)$ by Theorem 3.1(b). It follows that

$$\mathcal{T}_G - \frac{G^{1/2}\alpha_G^*}{V_G^{1/2}} = V_G^{-1/2} G^{1/2}(\hat{\alpha}_G - \alpha_G^*) + \mathrm{o}_P(1).$$

The second result follows from this equality by the asymptotic equivalence lemma and Theorem 3.1(b).

To prove (c), note that

$$G^{-1/2}\mathcal{T}_G - \bar{V}_G^{1/2}\alpha_G^* \to 0 \text{ in probability-}P. \tag{28}$$

by Theorems 3.1(a) and 3.2(a). Let c be an arbitrary real number. Then we have that for each $G \in \mathbb{N}$,

$$\begin{aligned} P[\mathcal{T}_G > c] &= P[G^{-1/2}\mathcal{T}_G - \bar{V}_G^{-1/2}\alpha_G^* > G^{-1/2}c - \bar{V}_G^{-1/2}\alpha_G^*] \\ &\geq P[G^{-1/2}\mathcal{T}_G - \bar{V}_G^{-1/2}\alpha_G^* > G^{-1/2}c - \tau], \end{aligned}$$

where $\tau \equiv \inf\{\bar{V}_G^{-1/2}\alpha_G^* :\ G \in \mathbb{N}\} > 0$ because $\{\alpha_G^*\}_{G\in\mathbb{N}}$ is assumed to be uniformly bounded, and $\bar{V}_G > V_G$ is positive uniformly in $G \in \mathbb{N}$ by Assumption 4(e). Because $G^{-1/2}c < \tau/2$ for almost all $G \in \mathbb{N}$, we have that for almost all $G \in \mathbb{N}$,

$$\begin{aligned} P[\mathcal{T}_G > c] &\geq P[G^{-1/2}\mathcal{T}_G - \bar{V}_G^{-1/2}\alpha_G^* > -\tau/2] \\ &\geq P[|G^{-1/2}\mathcal{T}_G - \bar{V}_G^{-1/2}\alpha_G^*| < \tau/2]. \end{aligned}$$

Because the right-hand side of this equality converges to one by (28), the desired result follows. $\square$

REFERENCES

Andrews, D. W. K. (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

Chow, G. C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28.

Davidson, J. (1994), *Stochastic Limit Theory—An Introduction for Econometricians*, Advanced Textbooks in Econometrics, New York: Oxford University Press.

Dufour, J.-M. (1982), "Generalized Chow Tests for Structural Change: A Coordinate-Free Approach," *International Economic Review*, 23, 565–575.

Fisher, F. M. (1970), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," *Econometrica*, 38, 361–366.

Gouriéroux, C., Monfort, A., and Trognon, A. (1984), "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52, 681–700.

Jennrich, R. I. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

Koenker, R. W. and Bassett, Jr., G. W. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Kullback, S. and Rosenblatt, H. M. (1957), "On the Analysis of Multiple Regression in k Categories," *Biometrika*, 44, 67–83.

Luenberger, D. G. (1969), *Optimization by Vector Space Methods*, Series in Decision and Control, New York, NY: Wiley.

Pesaran, M. H. and Yamagata, T. (2008), "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics*, 142, 50–93.

Sen, P. K. (1970), "On Some Convergence Properties of One-Sample Rank Order Statistics," *Annals of Mathematical Statistics*, 41, 2140–2143.

Swamy, P. A. V. B. (1970), "Efficient Inference in a Random Coefficient Regression Model," *Econometrica*, 38, 311–323.

Toyoda, T. (1974), "Use Of Chow Test under Heteroscedasticity," *Econometrica*, 42, 601–608.

Weerahandi, S. (1987), "Testing Regression Equality with Unequal Variances," *Econometrica*, 55, 1211–1215.

White, H. (1994), *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.