Statistica Sinica **28** (2018), 1437-1458
doi:https://doi.org/10.5705/ss.202016.0093

ADJUSTMENTS FOR A CLASS OF TESTS UNDER NONSTANDARD CONDITIONS

Anna Clara Monti and Masanobu Taniguchi

University of Sannio and Waseda University

Abstract: Generally the Likelihood Ratio statistic Λ for standard hypotheses is asymptotically χ^2 distributed, and the Bartlett adjustment improves the χ^2 approximation to its asymptotic distribution in the sense of third-order asymptotics. However, if the parameter of interest is on the boundary of the parameter space, Self and Liang (1987) show that the limiting distribution of Λ is a mixture of χ^2 distributions. For such "nonstandard setting of hypotheses", the present paper develops the third-order asymptotic theory for a class $\mathcal{S}$ of test statistics, which includes the Likelihood Ratio, the Wald, and the Score statistic, in the case of observations generated from a general stochastic process, providing widely applicable results. In particular, it is shown that Λ is Bartlett adjustable despite its nonstandard asymptotic distribution. Although the other statistics are not Bartlett adjustable, a nonlinear adjustment is provided for them which greatly improves the χ^2 approximation to their distribution and allows a subsequent Bartlett-type adjustment. Numerical studies confirm the benefits of the adjustments on the accuracy and on the power of tests whose statistics belong to $\mathcal{S}$.

Key words and phrases: Bartlett adjustment, boundary parameter, high-order asymptotic theory, likelihood ratio test, nonstandard conditions, score test, Wald test.

1. Introduction

Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ be a collection of p-dimensional random vectors generated by a stochastic process and let $p_{n,\theta}(\boldsymbol{x}_n)$, with $\boldsymbol{x}_n \in \mathbb{R}^{np}$ and $\boldsymbol{\theta} = \left(\theta^1, \ldots, \theta^q\right) \in \Theta \subset \mathbb{R}^q$, denote the probability density function of $\boldsymbol{X}_n$. The interest focuses on the statistical hypothesis

$$H : \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{1.1}$$

Notice that the data $\boldsymbol{X}_n$ can be dependent and/or not identically distributed, hence the problem considered here has applications in multivariate analysis and in time series analysis.

If the statistical model is a regular one whose probability density function is smooth with respect to $\boldsymbol{\theta}$, its derivatives have finite moments, and the value $\boldsymbol{\theta}_0$

of the parameter under H is an "interior" point of the parameter space $\boldsymbol{\Theta}$, then inference is carried out under "standard conditions".

The Likelihood Ratio (LR) statistic for (1.1) is given by

$$\Lambda = 2\log\left\{l_n(\hat{\boldsymbol{\theta}}_{Ml}) - l_n(\boldsymbol{\theta}_0)\right\}, \tag{1.2}$$

where $l_n(\boldsymbol{\theta}) = \log\{p_{n,\theta}(\boldsymbol{X}_n)\}$ and $\hat{\boldsymbol{\theta}}_{ML}$ is the Maximum Likelihood Estimator (MLE). Under standard conditions, Λ is asymptotically χ^2_q distributed, where $q = \dim\boldsymbol{\theta}_0$. To enhance the χ^2_q approximation to the distribution of the test statistic, Bartlett (1937) introduces - for $i.i.d.$ data - the adjusted statistic $\Lambda^* = (1+B/n)\Lambda$, where $(1+B/n) \approx q/\mathrm{E}(\Lambda)$, n is the sample size, and B is called the Bartlett adjustment factor.

Under these conditions we have

$$\mathrm{P}_{\theta_0}(\Lambda \le x) = \mathrm{F}_{\chi^2_q}(x) + n^{-1}A_\Lambda(x) + o(n^{-1}), \tag{1.3}$$

where $\mathrm{F}_{\chi^2_q}(x)$ is the distribution function of a χ^2_q random variable (r.v.). Lawley (1956) shows that

$$\mathrm{P}_{\theta_0}\left\{\left(1+\frac{B}{n}\right)\Lambda \le x\right\} = \mathrm{F}_{\chi^2_q}(x) + o(n^{-1}), \tag{1.4}$$

hence the n^{-1}-order term in (1.3) vanishes. Henceforth, if the test statistic satisfies (1.4), we say that the test is Bartlett adjustable (B-adjustable).

The hypothesis (1.1) can also be tested through the Wald statistic. In case $q = 1$, the statistic is

$$W = n\big(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\big)^2 I(\hat{\boldsymbol{\theta}}_{ML}), \tag{1.5}$$

where $I(\boldsymbol{\theta})$ is the Fisher information. Alternatively the modified Wald statistic can be used

$$MW = n\big(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\big)^2 I(\boldsymbol{\theta}_0),$$

where $I(\boldsymbol{\theta})$ is evaluated at $\boldsymbol{\theta}_0$ instead of at $\hat{\boldsymbol{\theta}}_{ML}$.

Under standard conditions, W is asymptotically χ^2_1 distributed. Furthermore $T = W^{1/2}\mathrm{sign}\big(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0\big)$ is asymptotically distributed as a $N(0,1)$ r.v.. Hayakawa and Puri (1985), Phillips and Park (1988), Ferrari and Cribari-Neto (1993) derive the asymptotic expansion

$$\mathrm{P}_{\theta_0}(W \le x) = \mathrm{F}_{\chi^2_1}(x) + n^{-1}A_W(x) + o(n^{-1}). \tag{1.6}$$

Under standard conditions and in the context of $i.i.d.$ data, Bartlett-type adjustments for the Wald statistic have been introduced by Phillips and Park (1988) and Ferrari and Cribari-Neto (1993) to enhance the χ^2_1 approximation to the distribution of W (see also the book by Cordeiro and Cribari-Neto (2014)).

The adjusted statistic is given by $W^* = \{1 + B(W)/n\}W$, where $B(W)$ is the Bartlett-type adjustment factor. Unlike the Bartlett adjustment for Λ, in the case of the Wald test the adjustment is nonlinear and depends on W. Ferrari and Cribari-Neto (1993) show that

$$\mathrm{P}_{\theta_0}(W^* \leq x) = \mathrm{F}_{\chi^2_1}(x) + o(n^{-1}),$$

hence the Bartlett correction eliminates the n^{-1}-order term in (1.6).

The score test may also be carried out on the hypothesis (1.1). The test statistic, when $q = 1$, is

$$SC = n \left\{ \left. \frac{\partial l_n(\boldsymbol{\theta})}{\partial \theta} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\}^2 I(\boldsymbol{\theta}_0)^{-1}.$$

An expansion analogous to (1.3) and (1.6) holds also for SC (Harris (1985)), and Bartlett-type adjustments have been investigated by Kakizawa (1997) (see also Cordeiro and Cribari-Neto (2014)).

The Bartlett adjustments have been mainly developed in the context of *i.i.d.* data. Nevertheless for the case of dependent data and/or non identically distributed data, Taniguchi (1991) and Taniguchi and Kakizawa (2000) introduce a class $\mathcal{S}$ of test statistics that includes the LR statistic, the Wald statistic, and the Score statistic as special cases, and they derive third-order asymptotic expansions analogous to (1.3) and (1.6). Such expansions allow one to determine sufficient conditions for a statistic $T \in \mathcal{S}$ to be B-adjustable. Accordingly, the higher order asymptotic theory for tests has been extensively developed when $\boldsymbol{\theta}_0$ is an interior point of the parameter space $\boldsymbol{\Theta}$.

If $\boldsymbol{\theta}_0$ is on the boundary of the parameter space, we say that the conditions are "nonstandard". These conditions arise when a parameter is known to be not smaller (or not greater) than a given threshold. For example, when testing the presence of an upward tendency in financial data against absence of this tendency in a given period, implying a non-negative parameter.

When the value of the parameter $\boldsymbol{\theta}_0$ is on the boundary of $\boldsymbol{\Theta}$, Chant (1974) shows (in the *i.i.d.* case) that the asymptotic distribution of the MLE is mixed normal. This result, in turn, implies that the asymptotic distribution of W for the null hypothesis (1.1), with $\boldsymbol{\theta}_0$ on the boundary of $\boldsymbol{\Theta}$, is given by a mixtures of χ^2 distributions. For the same testing problem Self and Liang (1987) show (again for the *i.i.d.* case) that the limiting distributions of the LR statistics are mixtures of χ^2 distributions. Nevertheless, for *i.i.d.* data, DiCiccio and Monti (2017) provide empirical evidence that the Bartlett adjustment can be applied to improve the χ^2 approximation to the distribution of the LR statistic when the parameter is on the boundary.

The present paper develops higher order asymptotic theory for a class $\mathcal{S}$ of test statistics, that includes the LR, the Wald and the Score statistic, under nonstandard conditions in the context of non-*i.i.d.* data. Initially the case of a scalar parameter is considered. This allows one to focus on the case that the parameter of interest is a scalar function of a vector of parameters $\boldsymbol{\theta}$. In particular, a sufficient condition for the LR statistic to be B-adjustable is given. A nonlinear transformation of the other statistics is proposed which leads to a more accurate χ^2 approximation to their distributions. For the Wald and the Score test, a sufficient condition for the modified test statistics to be B-adjustable is given, though the Bartlett-type adjustment, in these cases, is a nonlinear function of $\hat{\theta}_{ML}$. Numerical studies are provided which support the theoretical results.

The paper is organized as follows. In Section 2, observations may be dependent and/or non-identically distributed, and the focus is on the testing problem

$$H : \theta = \theta_0, \qquad A : \theta > \theta_0, \tag{1.7}$$

where $\theta \in \Theta \subset \mathbb{R}^1$ is a scalar parameter and θ_0 is on the boundary of Θ. We derive the third-order asymptotic expansion of the distribution of Λ under H, and prove that its limiting distribution is a mixture of 0 and a χ^2_1 distribution. Bartlett adjustments are discussed, and a sufficient condition is given for Λ to be B-adjustable. Bartlett coefficients for concrete statistical models are provided.

Section 2 also provides the third-order asymptotic expansion of the distribution of W under H. The statistic W usually is not B-adjustable, but we provide a nonlinear adjustment after which Wald statistics are B-adjustable.

Section 3 introduces a family of curved probability distributions $p_{n,\theta(u)}(\boldsymbol{x}_n)$, where $u \in \mathcal{H} \subset \mathbb{R}$, and $\boldsymbol{\theta} = \boldsymbol{\theta}(u) \in \mathbb{R}^q$, embedded in $\mathcal{F} = \{p_{n,\theta}(\boldsymbol{x}_n)\}$. The focus is on the testing problem

$$H : u = u_0, \qquad A : u > u_0 \tag{1.8}$$

under nonstandard conditions.

This setting is general and arises any time the parameter of interest is a function of $\boldsymbol{\theta}$. A motivating example for this testing problem is given by the optimal portfolio problem. Let $\{\boldsymbol{X}_t; t = 1, \ldots, n\}$ be a p-dimensional asset return process with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{V}$. Let $\boldsymbol{w} = (w^1, \ldots, w^p)'$ be the portfolio coefficient on the p assets. The portfolio return mean and variance are given, respectively, by

$$\mu(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\mu}, \qquad \eta^2(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{V}\boldsymbol{w}. \tag{1.9}$$

Suppose that a risk-free asset exists, whose return is denoted by R_0 and whose amount by w_0. The mean-variance optimal portfolio is determined by

$$\max_{w_0,\boldsymbol{w}} \left\{\mu(\boldsymbol{w}) + R_0 w_0 - \beta\eta^2(\boldsymbol{w})\right\} \qquad \text{subject to} \quad \sum_{j=0}^{p} w_j = 1, \tag{1.10}$$

where β is a given positive number. The solution for $\boldsymbol{w}$ is

$$\boldsymbol{w}_{opt} = \frac{1}{2\beta}\boldsymbol{V}^{-1}\left(\boldsymbol{\mu} - R_0\boldsymbol{e}\right) \tag{1.11}$$

(e.g., Taniguchi, Hirukawa and Tamaki (2008, p.278)), where $\boldsymbol{e} = (1, \ldots, 1)'$. When the interest focuses on the optimal portfolio coefficient on one asset, say the first one, w^1_{opt}, then $u = w^1_{opt}$ can be set. Let $\boldsymbol{\theta} = (\boldsymbol{\mu}', \mathrm{vech}(\boldsymbol{V})')'$; by (1.11) we get

$$u = u(\boldsymbol{\theta}) = u\left\{\boldsymbol{\mu}, \mathrm{vech}(\boldsymbol{V})\right\}. \tag{1.12}$$

Consequently hypotheses on the elements of (1.11) are in the framework of (1.8). This example, and similar ones, highlight the need for investigating procedures to handle the statistical problem introduced by (1.8).

For (1.8), Section 3 introduces a class $\mathcal{S}$ of test statistics which includes the LR, the Wald, and the Score statistic as special cases. Under the assumption that the third central moment of the score function K vanishes, the third-order asymptotic expansion of the distribution of a test statistic $T \in \mathcal{S}$ is derived under nonstandard conditions. This allows one to derive a sufficient condition for T to be adjustable up to third-order when $K = 0$. This assumption holds in many situations of interest, though there are some constraints to its application, as discussed in Section 3.

Section 4 investigates the benefits of the proposed adjustments, including the Bartlett adjustment, through various numerical studies. The results highlight the enhancement in the approximation to the asymptotic distributions of the test statistics which can be achieved by the adjustments, and their impact on the significance level and on the power of the test.

Proofs are in Section 5.

2. Higher Order Asymptotic Theory

The current section considers the case when $p_{n,\theta}(\boldsymbol{x}_n)$, the probability density function of the collection $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ of p-dimensional random vectors generated by a stochastic process, depends on an unknown scalar parameter $\theta \in \Theta$, and $\Theta = [\theta_0, b)$ (or $(b, \theta_0]$), where b is a finite constant.

We need some assumptions.

Assumption 1. *$p_{n,\theta} = p_{n,\theta}(\boldsymbol{x}_n)$ is continuously five times differentiable with respect to $\theta \in \Theta$. At $\theta = \theta_0$, the derivative $\partial/\partial\theta$ is taken from the right.*

Assumption 2. *The derivative $\partial/\partial\theta$ and the expectation E_θ with respect to $p_{n,\theta}$ are interchangeable.*

Assumption 3. *If $l_n(\theta) = \log\{p_{n,\theta}(\boldsymbol{X}_n)\}$ and*

$$Z_i = n^{-1/2}\left\{\frac{\partial^i l_n(\theta)}{\partial\theta^i} - E_\theta\left[\frac{\partial^i l_n(\theta)}{\partial\theta^i}\right]\right\}, \quad (i = 1, 2, 3),$$

the cumulants of Z_i have asymptotic expansions of the form

$$\mathrm{cum}_\theta\{Z_i, Z_j\} = \kappa_{ij}^{(1)}(\theta) + n^{-1}\kappa_{ij}^{(2)}(\theta) + o(n^{-1}), \tag{2.1}$$

$$\mathrm{cum}_\theta\{Z_i, Z_j, Z_k\} = n^{-1/2}\kappa_{ijk}^{(1)}(\theta) + o(n^{-1}), \tag{2.2}$$

$$\mathrm{cum}_\theta\{Z_i, Z_j, Z_k, Z_m\} = n^{-1}\kappa_{ijkm}^{(1)}(\theta) + o(n^{-1}), \tag{2.3}$$

$i, j, k, m = 1, 2, 3$, and the Jth-order ($J \geq 5$) cumulants satisfy

$$\mathrm{cum}_\theta^{(J)}\{Z_{i_1}, \ldots, Z_{i_J}\} = O\left(n^{-J/2+1}\right), \tag{2.4}$$

where $i_1, \ldots, i_J \in \{1, 2, 3\}$.

These assumptions are mild, e.g. Gaussian piecewise smooth time series models (see Taniguchi and Kakizawa (2000)) satisfy them.

Henceforth we write $I = \kappa_{11}^{(1)}(\theta)$, $J = \kappa_{12}^{(1)}(\theta)$, $K = \kappa_{111}^{(1)}(\theta)$, $M = \kappa_{22}^{(1)}(\theta)$, $N = \kappa_{112}^{(1)}(\theta)$, and $H = \kappa_{1111}^{(1)}(\theta)$. The Z_i and these quantities are functions of θ, but - when no ambiguity occurs - this argument is dropped for simplicity.

For the LR statistic for the hypotheses in (1.7), the Bartlett adjustment factor is

$$B = B(\theta_0) \equiv \frac{J^2}{4I^3} + \frac{-M + 2N + H}{4I^2},$$

which yields the adjusted statistic

$$\Lambda^* \equiv \left(1 + \frac{B}{n}\right)\Lambda.$$

Theorem 1. *If $K = 0$, we have*

$$\mathrm{P}_{n,\theta_0}(\Lambda \leq x) = \begin{cases} \dfrac{1}{2} + O(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + O(n^{-1}), & \text{if } x > 0; \end{cases} \tag{2.5}$$

$$\mathrm{P}_{n,\theta_0}(\Lambda^* \leq x) = \begin{cases} \dfrac{1}{2} + o(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + o(n^{-1}), & \text{if } x > 0. \end{cases} \tag{2.6}$$

We turn then to the asymptotic distributions of W and MW under H.

Theorem 2. *If $K = 0$, we have*

$$\mathrm{P}_{n,\theta_0}(W \leq x) = \begin{cases} \dfrac{1}{2} + O(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + O(n^{-1}), & \text{if } x > 0; \end{cases} \tag{2.7}$$

$$\mathrm{P}_{n,\theta_0}(MW \leq x) = \begin{cases} \dfrac{1}{2} + O(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + O(n^{-1}), & \text{if } x > 0. \end{cases} \tag{2.8}$$

In the general non-*i.i.d.* case, then, the distribution of the Wald statistics - under nonstandard conditions - is the same as in the *i.i.d.* context. Since the hypothesis (1.7) concerns a scalar parameter, the Wald test can also be carried out through T or $MT = MW^{1/2}$.

To use a linear Bartlett correction factor of the form $1 + B/n$, where B is a constant, it is necessary to apply a nonlinear correction to the Wald statistics (Taniguchi and Kakizawa (2000)). Let $h_W(\theta)$ be a function with derivatives

$$\frac{\partial h_W(\theta)}{\partial \theta} = -\frac{3J + K}{3I},$$
$$\frac{\partial^2 h_W(\theta)}{\partial \theta^2} = -\frac{12M + 18N + 8L + 3H}{6I} + \frac{27J^2 + 20JK + 4K^2}{6I^2};$$

and let $h_{MW}(\theta)$ be a function whose derivatives satisfy

$$\frac{\partial h_{MW}(\theta)}{\partial \theta} = \frac{3J + 2K}{3I},$$
$$\frac{\partial^2 h_{MW}(\theta)}{\partial \theta^2} = \frac{12N + 4L + 3H}{6I} + \frac{J^2}{2I^2}.$$

The corrected statistics are $\widetilde{W} = h_W(\hat{\theta}_{ML})W$ and $\widetilde{MW} = h_{MW}(\hat{\theta}_{ML})MW$. The Bartlett adjustment factor

$$B = B(\theta_0) \equiv \frac{\Delta}{I} + \frac{N - JKI^{-1}}{I^2} + \frac{H}{4I^2} - \frac{5K^2}{12I^3} \tag{2.9}$$

yields the Bartlett-adjusted statistics

$$\widetilde{W}^* \equiv \left(1 + \frac{B}{n}\right)\widetilde{W}, \qquad \widetilde{MW}^* \equiv \left(1 + \frac{B}{n}\right)\widetilde{MW}.$$

Theorem 3. *If $K = 0$, we have*

$$\mathrm{P}_{n,\theta_0}(\widetilde{W} \le x) = \begin{cases} \dfrac{1}{2} + O_{BA}(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + O_{BA}(n^{-1}), & \text{if } x > 0; \end{cases}$$

$$\mathrm{P}_{n,\theta_0}(\widetilde{MW} \le x) = \begin{cases} \dfrac{1}{2} + O_{BA}(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + O_{BA}(n^{-1}), & \text{if } x > 0. \end{cases}$$

The terms $O_{BA}(n^{-1})$ in Theorem 3 are such that they reduce to $o(n^{-1})$ after the Bartlett adjustment.

Theorem 4. *If $K = 0$, we have*

$$\mathrm{P}_{n,\theta_0}(\widetilde{W}^* \le x) = \begin{cases} \dfrac{1}{2} + o(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + o(n^{-1}), & \text{if } x > 0; \end{cases}$$

$$\mathrm{P}_{n,\theta_0}(\widetilde{MW}^* \le x) = \begin{cases} \dfrac{1}{2} + o(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + o(n^{-1}), & \text{if } x > 0. \end{cases}$$

The Bartlett-adjusted version of T and MT are

$$\widetilde{T}^* = (\widetilde{W}^*)^{1/2} \equiv \left(1 + \frac{\rho}{n}\right)^{1/2} h_W(\hat{\theta}_{ML})^{1/2} T,$$

$$\widetilde{MT}^* = (\widetilde{MW}^*)^{1/2} \equiv \left(1 + \frac{\rho}{n}\right)^{1/2} h_{MW}(\hat{\theta}_{ML})^{1/2} MT.$$

3. General Asymptotic Theory

The current section develops the general asymptotic theory under nonstandard conditions. Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ be a collection of p-dimensional random vectors generated by a stochastic process. Let $p_{n,\theta}(\boldsymbol{x}_n)$, with $\boldsymbol{x}_n \in \mathbb{R}^{np}$ and $\boldsymbol{\theta} = \left(\theta^1, \ldots, \theta^q\right) \in \boldsymbol{\Theta} \subset \mathbb{R}^q$, denote the probability density function of $\boldsymbol{X}_n$. Interest focuses on a family of curved probability densities $\mathcal{M} = \{p_{n,\theta(u)}(\boldsymbol{x}_n); u \in \Omega = [u_0, b), b < +\infty\}$ (b is a finite constant). The assumptions of Section 2 are adapted to the new context as follows.

Assumption 4.

(i) $p_{n,\theta} = p_{n,\theta}(\boldsymbol{x}_n)$ *is continuously five times differentiable with respect to* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. *At* $\boldsymbol{\theta} = \boldsymbol{\theta}(u_0)$, *the derivative* $\partial/\partial\boldsymbol{\theta}$ *is taken from the right.*

(ii) The embedding map $\boldsymbol{\theta} = \boldsymbol{\theta}(u)$ is continuously five times differentiable with respect to $u \in [u_0, b)$. At $u = u_0$, the derivative $\partial/\partial u$ is taken from the right.

Assumption 5. *The derivative $\partial/\partial\theta^i$ and the expectation E_θ are interchangeable.*

Let $l_n(\theta) = \log\{p_{n,\theta}(\boldsymbol{X}_n)\}$ and take

$$\begin{aligned}
Z_i &= n^{-1/2}\frac{\partial l_n(\theta)}{\partial\theta^i},\\
Z_{ij} &= n^{-1/2}\left\{\frac{\partial^2 l_n(\theta)}{\partial\theta^i\partial\theta^j} - E_\theta\left[\frac{\partial^2 l_n(\theta)}{\partial\theta^i\partial\theta^j}\right]\right\},\\
Z_{ijk} &= n^{-1/2}\left\{\frac{\partial^3 l_n(\theta)}{\partial\theta^i\partial\theta^j\partial\theta^k} - E_\theta\left[\frac{\partial^3 l_n(\theta)}{\partial\theta^i\partial\theta^j\partial\theta^k}\right]\right\},
\end{aligned}$$

where $i, j, k = 1, \ldots, q$, and at $\theta = \theta(u_0)$, the derivative $\partial/\partial\theta$ is taken from the right.

Assumption 6. *The moments and cumulants of Z_i, Z_{ij}, and Z_{ijk} satisfy*

$$\begin{aligned}
&E(Z_iZ_j) = I_{ij} + O(n^{-1}),\\
&E(Z_iZ_{jk}) = J_{ijk} + O(n^{-1}),\\
&E(Z_iZ_jZ_k) = n^{-1/2}K_{ijk} + O(n^{-3/2}),\\
&E(Z_iZ_{jkm}) = L_{ijkm} + O(n^{-1}),\\
&\mathrm{cum}(Z_{ij}, Z_{km}) = M_{ijkm} + O(n^{-1}),\\
&E(Z_iZ_jZ_{km}) = n^{-1/2}N_{ijkm} + O(n^{-3/2}),\\
&\mathrm{cum}(Z_i, Z_j, Z_k, Z_m) = n^{-1}H_{ijkm} + O(n^{-2}),
\end{aligned}$$

and the Jth-order cumulants of Z_i, Z_{ij} and Z_{ijk} are $O(n^{-J/2+1})$ for $J \geq 3$.

We estimate $u \in [u_0, b)$ initially by estimating $\boldsymbol{\theta}$ in the ambient large class $\mathcal{F} = \{p_{n,\theta}; \boldsymbol{\theta} \in \Theta\}$ by the *MLE* $\hat{\boldsymbol{\theta}}_{ML}$, which is known to be asymptotically sufficient. Estimation of u in $\mathcal{M}$ requires solving $\hat{\boldsymbol{\theta}}_{ML} = \boldsymbol{\theta}(\hat{u})$ with respect to $\hat{u}$; this problem cannot be solved because $\dim\boldsymbol{\theta} = q > \dim u = 1$. New extra coordinates $\boldsymbol{v} = (v^1, \ldots, v^{q-1})$ are introduced so that $\boldsymbol{w} = (w^1, \ldots, w^q) = (u, \boldsymbol{v}) = (u, v^1, \ldots, v^{q-1})$ becomes a coordinate system in $\mathcal{F}$. Then the equation

$$\hat{\boldsymbol{\theta}}_{ML} = \boldsymbol{\theta}(\hat{u}, \hat{\boldsymbol{v}}) \tag{3.1}$$

can be uniquely solved with respect to $\hat{u}$ and $\hat{\boldsymbol{v}}$. It is assumed that $\boldsymbol{\theta}(u, 0) = \boldsymbol{\theta}(u)$. By fixing u, we locally define the ancillary space $A(u) = \{(u, \boldsymbol{v}) \mid (u, \boldsymbol{v}) \in \mathcal{F}\}$ so that the family $\{A(u)\}$ defines a local foliation of $\mathcal{F}$. The determination of the estimator $\hat{u}$ of u is a one-to-one correspondence of the local foliation of $\{A(u)\}$, which is called the ancillary family associated with the estimator $\hat{u}$ (for *i.i.d.* curved exponential families, see Amari (1985))

Assumption 7. *The map $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{w})$ is continuously five times differentiable with respect to $\boldsymbol{w}$. At $\boldsymbol{w} = (u_0, 0)$, the derivatives are taken from the right.*

Let $\hat{u}_{ML}$ be the MLE of u obtained by (3.1). Then by Taniguchi and Watanabe (1994), we obtain the higher-order stochastic expansion and asymptotic expansion for the distribution of $\hat{u}_{ML}$ in terms of Z_i, Z_{ij}, Z_{ijk}, I_{ij}, J_{ijk}, K_{ijk}, ..., and $B^i_\alpha = \partial\theta^i/\partial w^\alpha$ and $\bar{B}^\alpha_i = \partial w^\alpha/\partial\theta^i$.

In this case $\hat{u}_{ML}$ is a function of the q-dimensional MLE $\hat{\boldsymbol{\theta}}_{ML}$. If the distribution of $\boldsymbol{X}_n$ is specified by u as itself, $p_{n,u}(\boldsymbol{X}_n)$, then the MLE of u is given by

$$\bar{u}_{ML} = \arg\max_u \left[\log\left\{p_{n,u}(\boldsymbol{X}_n)\right\}\right]. \tag{3.2}$$

While $\hat{u}_{ML}$ and $\bar{u}_{ML}$ differ, if the curvature of the larger model vanishes,

$$M_{iji'j'} - J_{kij}J_{k'i'j'}I^{kk'} = 0 \tag{3.3}$$

for all $i, j, i', j' = 1, \ldots, q$, we have a result essentially due to Taniguchi and Watanabe (1994).

Theorem 5. *Under Assumptions* 4-7*, if* (3.3) *holds, the Edgeworth expansions of the distribution of $\hat{u}_{ML}$ and that of $\bar{u}_{ML}$ are the same up to the term of order $n^{-3/2}$.*

We return to the general model $p_{n,\theta(u)}(\cdot)$, where u is a function of $\boldsymbol{\theta} = \left(\theta^1, \ldots, \theta^q\right)$, $u = u\left(\theta^1, \ldots, \theta^q\right)$, and $u \in [u_0, b)$, with the purpose of handling the testing problem (1.8) under nonstandard conditions.

This setting is general and optimal portfolio choice problems provide a relevant motivating example. In this context we usually use test statistics based on $\hat{u}_{ML} = u\big\{\hat{\boldsymbol{\mu}}_{ML}, \mathrm{vech}(\hat{\boldsymbol{V}}_{ML})\big\}$. However, if the return process $\{\boldsymbol{X}_t\}$ is *i.i.d.* Gaussian, it can be verified that (3.3) holds, and, by Theorem 5, we can use $\hat{u}_{ML}$ instead of $\bar{u}_{ML}$. Therefore the theoretical results of Section 2 can be applied for the general testing problem (1.8).

On the basis of Theorem 5, we assume that the distribution of $\boldsymbol{X}_n$ depends on a scalar u. We write $Z_1 = Z_1(\boldsymbol{\theta})$, $Z_2 = Z_2(\boldsymbol{\theta}), \ldots, I = I(\boldsymbol{\theta})$, $J = J(\boldsymbol{\theta}), \ldots$, etc. as $Z_1 = Z_1(u)$, $Z_2 = Z_2(u), \ldots, I = I(u)$, $J = J(u), \ldots$, etc. Take $W_1 = Z_1/I^{1/2}$, $W_2 = Z_2 - JI^{-1}Z_1$, and $W_3 = Z_3 - LI^{-1}Z_1$. For (1.8) we introduce a class of test statistics $\mathcal{S} = \{T\}$ such that, conditionally on $\mathcal{X}\{W_1 > 0\}$, T has the stochastic expansion

$$
\begin{aligned}
T = {} & W_1^2 + n^{-1/2}\left(a_1 W_1^2 W_2 + a_2 W_1^3\right) \\
& + n^{-1}\left(b_1 W_1^2 + b_2 W_1^2 W_2^2 + b_3 W_1^4 + b_4 W_1^3 W_2 + b_5 W_1^3 W_3\right) \\
& + o_p\left(n^{-1}\right),
\end{aligned} \tag{3.4}
$$

where a_i $(i = 1, 2)$ and b_i $(i = 1, \ldots, 5)$ are nonrandom constants.

Let $l_n(u) = \log\{p_{n,u}(\boldsymbol{X}_n)\}$ and define

$$
\begin{aligned}
&\Lambda(\bar{u}_{ML}) = 2\{l_n(\bar{u}_{ML}) - l_n(u_0)\}, \\
&W(\bar{u}_{ML}) = n(\bar{u}_{ML} - u_0)^2 I(\bar{u}_{ML}), \\
&MW(\bar{u}_{ML}) = n(\bar{u}_{ML} - u_0)^2 I(u_0), \\
&SC = W_1^2 \mathcal{X}(W_1 > 0).
\end{aligned}
$$

Theorem 6. *If $K = 0$ then*

(i) the test statistics $\Lambda(\bar{u}_{ML})$, $W(\bar{u}_{ML})$, $MW(\bar{u}_{ML})$, and SC belong to $\mathcal{S}$;

(ii) for $T \in \mathcal{S}$,

$$
\mathrm{P}_{n,u_0}(T \le x) = \begin{cases} \dfrac{1}{2} + O(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + O(n^{-1}), & \text{if } x > 0. \end{cases} \tag{3.5}
$$

Theorem 7.

(i) Suppose $h = h(u)$ is continuously three times differentiable with respect to u and $h(u_0) = 1$. If $K = 0$ and the derivatives $h' = h'(u)$ and $h'' = h''(u)$ satisfy

$$
h' = -I^{1/2} a_2 \tag{3.6}
$$

$$
h'' = -\frac{I}{2}\left(M - \frac{J^2}{I}\right) a_1^2 - N a_1 - 2 I b_3 + \frac{a_2}{I^{1/2}}\left(2 I^{3/2} a_2 - J\right) - \frac{H}{6I}, \tag{3.7}
$$

then the modified test statistic $\tilde{T} = h\{\bar{u}_{ML}\} T$ is B-adjustable.

(ii) If B is the Bartlett adjustment factor, for $\tilde{T}^* = (1 + B/n)\tilde{T}$ we have

$$
\mathrm{P}_{n,u_0}(\tilde{T}^* \le x) = \begin{cases} \dfrac{1}{2} + o(n^{-1}), & \text{if } x = 0, \\ \dfrac{1}{2}\left\{1 + F_{\chi_1^2}(x)\right\} + o(n^{-1}), & \text{if } x > 0. \end{cases} \tag{3.8}
$$

Remark 1. If (3.3) holds, then by applying Theorem 5 it can be shown that $\Lambda(\hat{u}_{ML})$, $W(\hat{u}_{ML})$, $MW(\hat{u}_{ML})$, and SC belong to $\mathcal{S}$. Thus the results of Theorems 6 and 7 hold, and $\Lambda(\hat{u}_{ML})$, $W(\hat{u}_{ML})$, $MW(\hat{u}_{ML})$, and SC can be applied to test the hypothesis (1.8) as required.

Remark 2. There are contexts in which $K = 0$ does not hold. A noteworthy case is the test on the variance in random/mixed effect models. In other cases when $K(u) \neq 0$, an alternative parameterization $\theta(u')$ can be introduced such that

$$I(u') \equiv \frac{d\theta^i}{du'}\frac{d\theta^j}{du'} I_{ij}(\boldsymbol{\theta}) \neq 0,$$
$$K(u') \equiv \frac{d\theta^i}{du'}\frac{d\theta^j}{du'}\frac{d\theta^k}{du'} K_{ijk}(\boldsymbol{\theta}) = 0,$$

for $q \geq 2$ (c.f., Taniguchi and Kakizawa (2000, p.222)). These functional equations might be not easy to solve, and the transformation $u \to u'$ changes the meaning of the parameter. Extensions of Theorems 6 and 7 to the case of non-vanishing K are still under investigation.

4. Numerical Analysis

Let $X_1, \cdots, X_n$ be generated from the $AR(1)$ process

$$X_t = \theta X_{t-1} + u_t, \qquad (X_0 \equiv 0), \tag{4.1}$$

where the parameter θ is known to be nonnegative

$$\theta \in \Theta = [0, 1), \tag{4.2}$$

and the u_t's are *i.i.d.* $N(0, \sigma^2)$ random variables. Interest focuses on testing the hypotheses

$$H : \theta_0 = 0, \qquad A : 0 < \theta_0 < 1, \tag{4.3}$$

which are in the framework of (1.7) if σ^2 is known, and in the framework of (1.8) in the general case of unknown innovation variance.

Under (4.1) and (4.2) the MLE of θ is given approximately by

$$\widetilde{\theta}_{ML} = \begin{cases} (1 - n^{-1})(\sum_{t=1}^{n-1} X_t X_{t+1})(\sum_{t=2}^{n-1} X_t^2)^{-1} & \text{if } \sum_{t=1}^{n-1} X_t X_{t+1} > 0, \\ 0 & \text{if } \sum_{t=1}^{n-1} X_t X_{t+1} \leq 0. \end{cases}$$

where (e.g., Fujikoshi and Ochi (1984)) $\sqrt{n}(\widetilde{\theta}_{ML} - \theta_0) = \sqrt{n}(\widehat{\theta}_{ML} - \theta_0) + o_p(n^{-1})$ under H, and $\widehat{\theta}_{ML}$ is the exact MLE.

Let $\hat{u}_t = X_t - \widetilde{\theta}_{ML} X_{t-1}$ be the ML residuals for $t = 1, \ldots n$, where $\hat{u}_1 = X_1$ since $X_0 \equiv 0$. For the hypotheses in (4.3), the LR statistic when σ^2 is known is

$$\Lambda = \frac{1}{\sigma^2}\left(\sum_{t=1}^{n} X_t^2 - \sum_{t=1}^{n} \hat{u}_t^2\right).$$

Obviously $\Lambda = 0$ when $\widetilde{\theta}_{ML} = 0$.

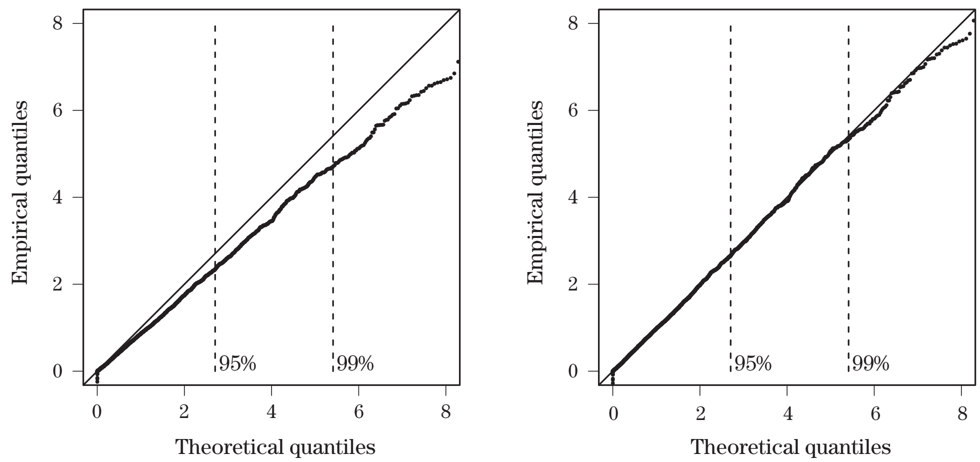

Figure 1. QQ-plot of the percentiles of Λ (left panel) and of Λ^* (right panel) versus the percentiles of the asymptotic distribution ($n = 15$, 10,000 simulations).

In this case the Bartlett adjustment factor is $B(\theta_0) = 2$ (c.f. Taniguchi and Kakizawa (2000, p.257)), hence the Bartlett adjusted statistic is

$$\Lambda^* = \left(1 + \frac{2}{n}\right)\Lambda.$$

By Theorem 1, the asymptotic distributions of Λ and Λ^* are given by the same mixture of 0 and a χ^2_1 r.v., though with a different rate of convergence. Figure 1 shows the QQ-plot of the percentiles of Λ (left panel) and of Λ^* (right panel) versus the percentiles of the asymptotic distribution when $n = 15$ and $\sigma^2 = 1$, based on 10,000 simulations. It can be appreciated how the Bartlett adjustment greatly enhances the approximation by the asymptotic distribution even for such a small sample size, and even far out in the tail of the distribution.

Remark 3. The $AR(1)$ model with known innovation variance provides some insight on the relevance of Theorem 5. If $X_t \in \mathcal{F}$ is generated by an $AR(1)$ model with autoregressive parameter θ and unknown variance σ^2, (3.3) is satisfied (Taniguchi and Kakizawa (2000, p.232)). The assumption that σ^2 is known turns the model to a curved $AR(1)$ model $\mathcal{M}$. Let $\hat{\theta}_{ML}$ be the MLE for θ in $\mathcal{F}$ and let $\widetilde{\theta}_{ML}$ be the MLE in $\mathcal{M}$; since (3.3) holds for $\mathcal{F}$, Theorem 5 implies that the third-order Edgeworth expansions of $\hat{\theta}_{ML}$ and $\bar{u}_{ML}$ are identical.

The LR statistic when σ^2 is unknown is

$$\Lambda_p = n \log\left(\frac{\sum_{t=1}^n X_t^2}{\sum_{t=1}^n \hat{u}_t^2}\right),$$

Table 1. Simulated significance level of the test based on Λ_p and Λ_p^* (100,000 simulations).

	Λ			Λ^*		
n	10%	5%	1%	10%	5%	1%
10	8.90	4.31	0.75	10.92	5.83	1.32
15	9.25	4.48	0.83	10.67	5.49	1.20
20	9.46	4.56	0.82	10.55	5.39	1.10
30	9.62	4.63	0.90	10.36	5.15	1.07
50	9.55	4.64	0.90	9.99	4.95	1.03

Table 2. Simulated Power of the test based on Λ and Λ^* (5% significance level, 100,000 simulations).

	$\theta = 0.05$		$\theta = 0.10$		$\theta = 0.15$		$\theta = 0.25$	
n	Λ_p	Λ_p^*	Λ_p	Λ_p^*	Λ_p	Λ_p^*	Λ_p	Λ_p^*
10	5.72	7.56	7.46	9.78	9.65	12.50	15.72	19.63
15	6.51	7.94	9.10	10.81	12.38	14.59	21.80	24.93
20	7.07	8.18	10.37	11.87	14.79	16.70	27.28	29.89
30	8.00	8.81	12.65	13.78	19.07	20.53	37.23	39.24
50	9.23	9.76	16.42	17.29	26.70	27.77	53.63	54.86

and the Bartlett adjusted statistic is

$$\Lambda_p^* = \left(1 + \frac{2}{n}\right)\Lambda_p.$$

By virtue of Theorems 6 and 7, the asymptotic distributions of Λ_p and Λ_p^* are still given by a mixture of 0 and a χ_1^2 r.v. with an error which is $O\left(n^{-1}\right)$ for the former statistic and $o\left(n^{-1}\right)$ for the latter.

To investigate the impact of the Bartlett adjustment, a simulation experiment was carried out by generating 100,000 samples from (4.1), under H, for various samples sizes ($n = 10, 15, 20, 30, 50$). Table 1 shows the simulated significance level of the test based on Λ_p and on Λ_p^*, when the nominal level is 10%, 5%, and 1%. The Bartlett adjustment enhances the accuracy of the test.

Table 2 shows the power (simulated again on 100,000 samples) of the two tests when $\theta = 0.05, 0.10, 0.15, 0.25$, the nominal level is 5% and the sample sizes are those of Table 1. The results show that the Bartlett adjustment produces considerable benefits in terms of power.

For the testing problem (4.3), the Wald statistic and the modified Wald statistic are

$$W = n\widetilde{\theta}_{ML}^2\left(1 - \widetilde{\theta}_{ML}^2\right)^{-1}, \qquad MW = n\widetilde{\theta}_{ML}^2$$

(whether σ^2 is known or unknown). The nonlinear corrections, which make the

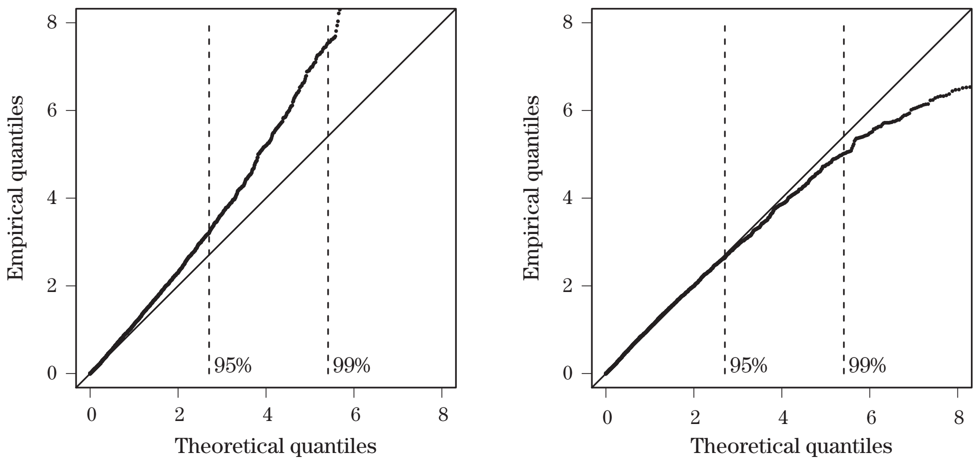

Figure 2. QQ-plot of the percentiles of W (left panel) and of MW (right panel) versus the percentiles of the asymptotic distribution ($n = 15$, 10,000 simulations).

statistics B-adjustable, are

$$h_W(\widetilde{\theta}_{ML}) = 1 - \frac{\widetilde{\theta}^2_{ML}}{2}, \qquad h_{MW}(\widetilde{\theta}_{ML}) = 1 + \frac{\widetilde{\theta}^2_{ML}}{2}.$$

Consequently the Bartlett-adjusted statistics are

$$\widetilde{W} = h_W(\widetilde{\theta}_{ML})W = n\widetilde{\theta}^2_{ML}\left(1 - \frac{\widetilde{\theta}^2_{ML}}{2}\right)(1 - \widetilde{\theta}^2_{ML})^{-1},$$

$$\widetilde{MW} = h_{MW}(\widetilde{\theta}_{ML})\widetilde{MW} = n\left(1 + \frac{\widetilde{\theta}^2_{ML}}{2}\right)\widetilde{\theta}^2_{ML}.$$

In this case, (2.9) yields the Bartlett adjustment factor $B(\theta_0) = -1/2$ (c.f. Taniguchi and Kakizawa (2000, p.257)), hence the Bartlett adjusted statistics are

$$\widetilde{W}^* = \left(1 - \frac{1}{2n}\right)\widetilde{W}, \qquad \widetilde{MW}^* = \left(1 - \frac{1}{2n}\right)\widetilde{MW}.$$

Figure 2 shows the QQ-plot of the percentiles of W (left panel) and MW (right panel) versus the percentiles of the asymptotic distribution when $n = 15$, obtained from 10,000 simulations. For the same samples, Figure 3 shows the QQ-plot for $\widetilde{W}$ and $\widetilde{MW}$ and Figure 4 shows the QQ-plot for $\widetilde{W}^*$ and $\widetilde{MW}^*$. The correction that leads from W and MW to $\widetilde{W}$ and $\widetilde{MW}$ largely enhances the approximation by the asymptotic distribution, and a further sensible improvement is achieved by applying the Bartlett adjustment.

In order to investigate the impact of the correction of the test statistics and of the Bartlett adjustment on the significance level, a simulation experiment was

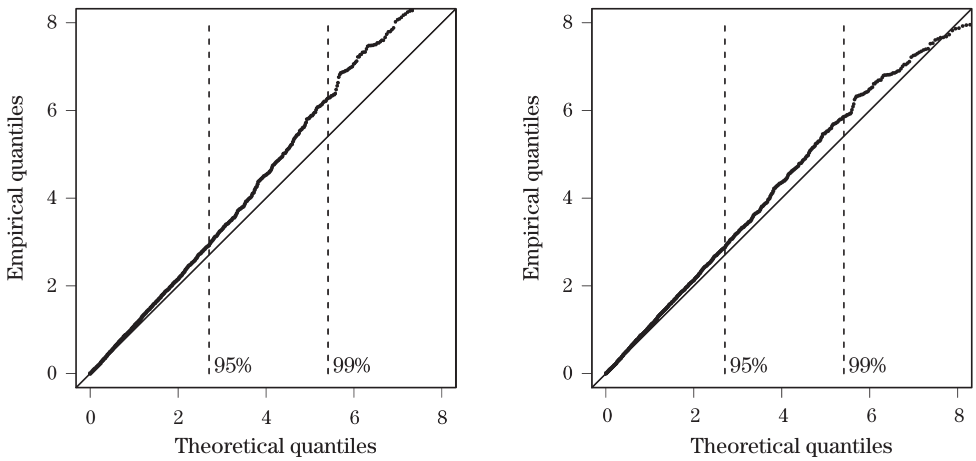

Figure 3. QQ-plot of the percentiles of $\widetilde{W}$ (left panel) and of $\widetilde{MW}$ (right panel) versus the percentiles of the asymptotic distribution ($n = 15$, 10,000 simulations).

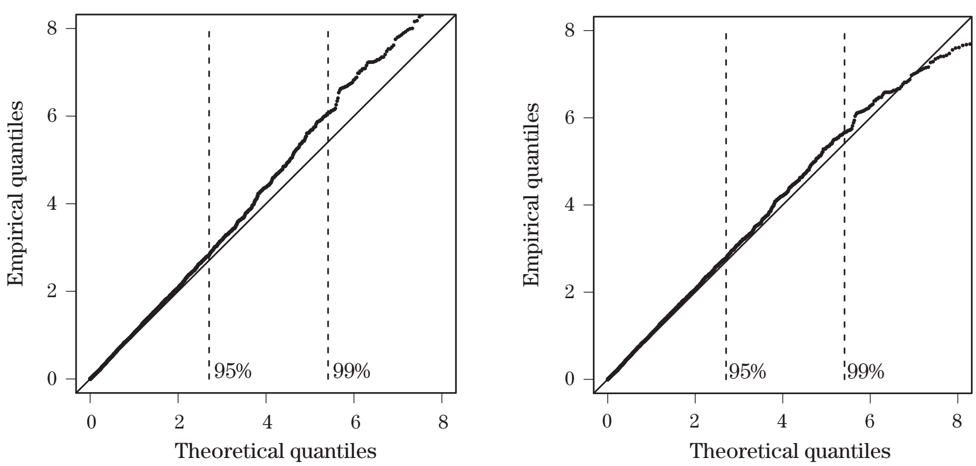

Figure 4. QQ-plot of the percentiles of $\widetilde{W}^*$ (left panel) and of $\widetilde{MW^*}$ (right panel) versus the percentiles of the asymptotic distribution ($n = 15$, 10,000 simulations).

carried out by generating 100,000 samples from (4.1) with $\theta = 0$, for various samples sizes ($n = 10, 15, 20, 30$). Table 3 shows the simulated level of the tests based on W and related enhanced statistics, whereas Table 4 shows the significance level of the tests (performed on the same samples) based on MW and related statistics, when the nominal level is 10%, 5%, and 1%. The accuracy of the Bartlett-adjusted test is remarkable.

We return to the optimal portfolio problem (1.10). Suppose $\boldsymbol{X}(1)$, $\boldsymbol{X}(2), \ldots,$

Table 3. Simulated significance level of the test based on W, $\widetilde{W}$ and $\widetilde{W}^*$ (100,000 simulations).

	W			$\widetilde{W}$			$\widetilde{W}^*$		
n	10%	5%	1%	10%	5%	1%	10%	5%	1%
10	12.17	7.43	2.84	11.38	6.39	2.01	10.85	5.96	1.84
15	11.19	6.35	1.98	10.64	5.63	1.39	10.25	5.33	1.28
20	10.80	5.85	1.58	10.33	5.30	1.15	10.04	5.07	1.07
30	10.40	5.40	1.31	10.11	5.04	1.03	9.91	4.88	0.99

Table 4. Simulated significance level of the test based on MW, $\widetilde{MW}$ and $\widetilde{MW}^*$ (100,000 simulations).

	MW			$\widetilde{MW}$			$\widetilde{MW}^*$		
n	10%	5%	1%	10%	5%	1%	10%	5%	1%
10	10.47	5.10	0.92	11.35	6.20	1.65	10.80	5.76	1.44
15	9.94	4.75	0.73	10.57	5.49	1.21	10.19	5.18	1.09
20	9.86	4.62	0.70	10.29	5.23	1.04	10.01	4.99	0.96
30	9.78	4.61	0.76	10.10	5.00	1.00	9.89	4.85	0.95

$\boldsymbol{X}(t)$, ... are generated by a 2-dimensional *i.i.d.* $N(\boldsymbol{\mu}, \boldsymbol{V})$ return process where $\boldsymbol{\mu} = (\mu_1, 0)$ and $\boldsymbol{V} = \{v_{ij}, i, j = 1, 2\}$. We assume that $\mu_1 \geq 0$. In this problem, for simplicity, we set $R_0 = 0$, $w_0 = 0$, and $\beta = 1/2$. If we are interested in the first portfolio coefficient w_1, then its optimal value is given by

$$u \equiv w_1^{opt} = v^{11}\mu_1, \tag{4.4}$$

where v^{11} is the $(1,1)$-element of $\boldsymbol{V}^{-1}$. Consider the testing problem

$$H: \ u = 0, \qquad A: \ u > 0, \tag{4.5}$$

which is analogous to (1.8) with $u_0 = 0$. Let $\{X_1(1), X_2(1)\}', \ldots, \{X_1(n), X_2(n)\}'$ be the observed stretch. The estimators of the elements of $\boldsymbol{\mu}$ and $\boldsymbol{V}$ are given by

$$\hat{\mu}_i = \frac{1}{n}\sum_{t=1}^{n} X_i(t) \qquad i = 1, 2;$$

$$\hat{v}_{ij} = \frac{1}{n}\sum_{t=1}^{n} \{X_i(t) - \hat{\mu}_i\}\{X_j(t) - \hat{\mu}_j\} \qquad i, j = 1, 2.$$

The Wald statistic for (4.5) (by neglecting the $o_p(n^{-1})$ term) is $W = n\tilde{\mu}_1^2/\hat{v}_{11}$ where $\tilde{\mu}_1 = \hat{\mu}_1 \mathcal{X}\{\hat{\mu}_1 > 0\}$. It is easily proved that W is B-adjustable with $B = -3$, so $W^* \equiv (1 - 3/n)W$ is the Bartlett-adjusted statistic.

Figure 5 shows the QQ-plot of the percentiles of W (left panel) and W^* (right panel) versus the percentiles of the asymptotic distribution when $n = 30$,

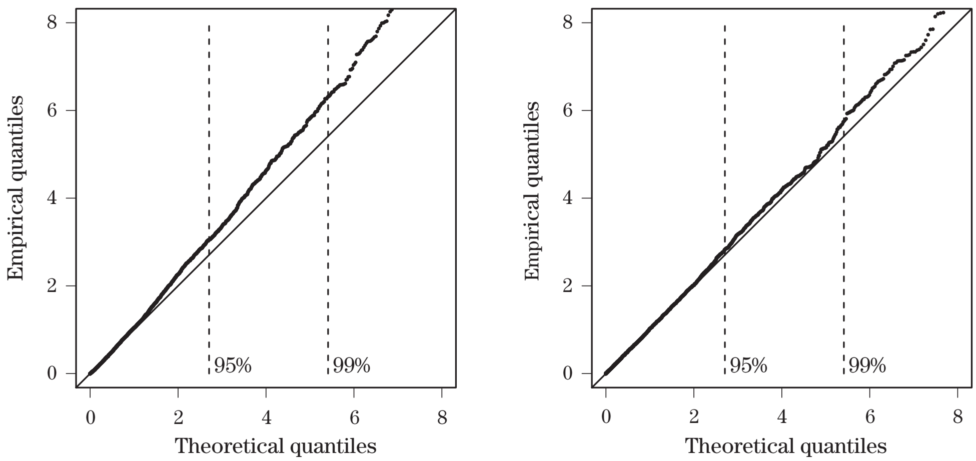

Figure 5. QQ-plot of the percentiles of W (left panel) and of W^* (right panel) versus the percentiles of the asymptotic distribution in the portfolio problem (n = 30, 10,000 simulations).

obtained from 10,000 simulations when the correlation coefficient between $X_1(t)$ and $X_2(t)$ was $\rho = 0.7$. Although the asymptotic distribution provides already a fairly good approximation to the actual distribution of the test statistics, the Bartlett adjustment yields a large improvement.

Table 5 shows the performance of the Wald test evaluated through a simulation experiment on 100,000 samples of sizes $n = 30, 50, 100$, generated under H of (4.5). The elements of $\boldsymbol{V}$ are $v_{11} = v_{22} = 1$, while $v_{12} = \rho = 0.10$, 0.30, 0.50, 0.70, 0.90. The simulated level of the tests based on W and W^* are compared for the nominal levels 10%, 5%, and 1%. The results provide further evidence on the substantial increase in accuracy of the test that can be achieved by the Bartlett adjustment, for the correlation scenarios, even for fairly small samples.

5. Proofs

Let $L = \kappa_{13}^{(1)}(\theta)$, $\Delta = \kappa_{11}^{(2)}(\theta)$, and denote by $W_1 = Z_1/\sqrt{I}$ the standardized score function. Let $W_2 = Z_2 - J \cdot I^{-1} Z_1$, and $W_3 = Z_3 - L \cdot I^{-1} Z_1$.

Proposition 1. *If $K = 0$, we have*

$$\mathrm{P}_{n,\theta}(W_1 \geq 0) = \frac{1}{2} + o(n^{-1}). \tag{5.1}$$

Proof. From (4.1.3) of Taniguchi and Kakizawa (2000, p.170) and (2.1)-(2.4), the asymptotic expansion of the distribution of W_1 is given by

Table 5. Simulated significance level of the test based on W and W^* for the portfolio problem - $n = 30, 50, 100$ (100,000 simulations).

	W			W^*		
ρ	10%	5%	1%	10%	5%	1%
			$n = 30$			
0.10	10.96	5.91	1.44	9.77	5.03	1.08
0.30	11.04	5.92	1.52	9.84	5.04	1.12
0.50	10.87	5.76	1.49	9.69	4.87	1.14
0.70	10.89	5.79	1.46	9.73	4.87	1.07
0.90	10.82	5.86	1.49	9.67	4.94	1.10
			$n = 50$			
0.10	10.63	5.57	1.26	9.93	5.03	1.05
0.30	10.62	5.53	1.27	9.94	5.01	1.07
0.50	10.60	5.42	1.27	9.89	4.89	1.06
0.70	10.43	5.47	1.27	9.78	4.95	1.06
0.90	10.60	5.41	1.24	9.87	4.90	1.03
			$n = 100$			
0.10	10.46	5.39	1.19	10.13	5.10	1.08
0.30	10.30	5.20	1.15	9.95	4.95	1.06
0.50	10.18	5.19	1.15	9.87	4.92	1.06
0.70	10.33	5.24	1.14	9.98	4.97	1.04
0.90	10.28	5.34	1.15	9.95	5.07	1.05

$$\mathrm{P}_{n,\theta}(W_1 \le y) = \Phi(y) - \phi(y)\Big\{ \frac{1}{2n}\triangle \cdot y + \frac{K}{6n^{1/2}}(y^2-1) \\ + \frac{H}{24n}(y^3-3y) + \frac{K^2}{72n}(y^5 - 10y^3 + 15y) \Big\} + o(n^{-1}), \quad (5.2)$$

where $\phi(y)$ and $\Phi(y)$ are the standard normal density and distribution function. If $K = 0$, from (5.2) we obtain (5.1).

Proof of Theorem 1. For all $\theta\,(\theta \ge \theta_0)$ such that $\theta - \theta_0 = O_p(n^{-1/2})$, we obtain

$$2\{\log p_{n,\theta}(\boldsymbol{X}_n) - \log p_{n,\theta_0}(\boldsymbol{X}_n)\} = -\{Z_1(\theta_0) - n^{1/2}I(\theta_0)(\theta-\theta_0)\}^2 I(\theta_0)^{-1} \\ + Z_1(\theta_0)^2 I(\theta_0)^{-1} + O_p(n)|\theta - \theta_0|^3, \quad (5.3)$$

(see Chernoff (1954) and Self and Liang (1987)). Let

$$\widetilde{W}_1 \equiv W_1 \cdot \mathcal{X}\{W_1 > 0\}. \quad (5.4)$$

From (5.3) it follows that

$$n^{1/2}(\widehat{\theta}_{ML} - \theta_0) = \frac{1}{\sqrt{I}}\widetilde{W}_1 + O_p(n^{-1/2}). \quad (5.5)$$

Hence, conditionally on $\mathcal{X}\{W_1 > 0\} = 1$,

$$n^{1/2}(\widehat{\theta}_{ML} - \theta_0) = \frac{1}{\sqrt{I}} W_1 + O_p(n^{-1/2}). \tag{5.6}$$

Since

$$0 = \frac{\partial}{\partial \theta} \log p_{n,\widehat{\theta}_{ML}}(\boldsymbol{X}_n), \quad \text{and} \quad \widehat{\theta}_{ML} \geq \theta_0, \tag{5.7}$$

by expanding (5.7) at θ_0 from the right, we have

$$0 = \frac{\partial}{\partial \theta} \log p_{n,\theta_0}(\boldsymbol{X}_n) + \left\{ \frac{\partial^2}{\partial \theta^2} \log p_{n,\theta_0}(\boldsymbol{X}_n) \right\} (\widehat{\theta}_{ML} - \theta_0) + \cdots + o_p(n^{-3/2}). \tag{5.8}$$

By rewriting (5.8) and making the replacement $U_n = n^{1/2}(\widehat{\theta}_{ML} - \theta_0)$, conditionally on $W_1 \geq 0$ we obtain

$$\begin{aligned} n^{1/2}(\widehat{\theta}_{ML} - \theta_0) = \frac{1}{\sqrt{I}} W_1 &+ \frac{1}{n^{1/2}} \{\text{polynominal of } W_1 \text{ and } W_2\} \\ &+ \frac{1}{n} \{\text{polynominal of } W_1, W_2 \text{ and } W_3\} \\ &+ o_p(n^{-1}) \end{aligned} \tag{5.9}$$

(see (4.2.71) of Taniguchi and Kakizawa (2000, pp.180-205)).

By expanding Λ at θ_0, and making the replacement (5.9), conditionally on $\mathcal{X}\{W_1 > 0\} = 1$, one has

$$\begin{aligned} \Lambda = &W_1^2 + n^{-1/2}(a_1 W_1^2 W_2) \\ &+ n^{-1}(b_1 W_1^2 + b_2 W_1^2 W_2^2 + b_3 W_1^4 + b_4 W_1^3 W_2 + b_5 W_1^4 W_3) \\ &+ o_p(n^{-1}). \end{aligned} \tag{5.10}$$

By applying Theorem 4.5.3 of Taniguchi and Kakizawa (2000, p.256) to (5.10) conditionally on $\mathcal{X}\{W_1 > 0\} = 1$, we see that

$$\text{(1) } \mathrm{P}_{n,\theta_0}(\Lambda \leq x | W_1) = \begin{cases} F_{\chi_1^2}(x) + O(n^{-1}), & \text{if } W_1 \geq 0, \\ F_{\{0\}}(x) + O(n^{-1}), & \text{if } W_1 < 0, \end{cases} \tag{5.11}$$

$$\text{(2) } \mathrm{P}_{n,\theta_0}(\Lambda^* \leq x | W_1) = \begin{cases} F_{\chi_1^2}(x) + o(n^{-1}), & \text{if } W_1 \geq 0, \\ F_{\{0\}}(x) + o(n^{-1}), & \text{if } W_1 < 0, \end{cases} \tag{5.12}$$

where

$$F_{\{0\}}(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0, \end{cases}$$

which leads to (2.5) and (2.6).

Because the proofs of Theorems 2 - 4 are essentially included in Theorems 6 and 7, they are omitted.

Proof of Theorem 5. Conditionally on $W_1 > 0$, we can apply Theorems 5, 8, and 10 of Taniguchi and Watanabe (1994) to our setting. Then, multiplying the Edgeworth expansions by $\mathrm{P}(W_1 > 0)$ leads to the conclusion.

Proof of Theorem 6. By recalling (5.9) for $\bar{u}_{ML}$, conditionally on $W_1 > 0$, we have

$$\begin{aligned} n^{1/2}(\bar{u}_{ML} - u_0) = & \frac{1}{\sqrt{I}} W_1 + \frac{1}{n^{1/2}} \{\text{polynominal of } W_1 \text{ and } W_2\} \\ & + \frac{1}{n} \{\text{polynominal of } W_1, W_2 \text{ and } W_3\} \\ & + o_p(n^{-1}). \end{aligned} \tag{5.13}$$

Substitution of (5.13) in the four statistics yields the result. The proof of (3.8) is analogous to that of (5.11) and (5.12).

Acknowledgment

Research by the first author was partly supported vy the SHAPE project within the frame of Programme STAR (CUP E68C13000020003) at University of Naples Federico II, financially supported by UniNA and Compagnia di San Paolo. Research by the second author was supported by Japanese JSPS Grant-in-Aid: Kiban(A) (15H02061) and Houga (26540015), and was done at the Research Institute for Science & Engineering, Waseda University.

References

Amari, S. (1985). Differential geometrical methods in statistics. *Lecture Notes in Statistics* **28**, Springer Series in Statistics, Springer-Verlag, New York/Berlin.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. Lond. A* **160**, 268–282.

Chant, D. (1974). On asymptotic test of composite hypotheses in nonstandard conditions. *Biometrika* **61**, 291–298.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573–578.

Cordeiro, G. M. and Cribari-Neto, F. (2014) *An Introduction to Bartlett Correction and Bias Reduction*, Springer Series in Statistics, Springer-Verlag, New York.

DiCiccio, T. J. and Monti, A. C. (2017). Testing for sub-models of the skew t-distribution. *Statistical Methodo & Applications.* `https://doi.org/10.1007/s10260-017-0387-x`.

Ferrari, S. L. P. and Cribari-Neto, F. (1993). On the corrections to the Wald test of non-linear restrictions. *Economics Letters* **42**, 321–326.

Fujikoshi, Y. and Ochi, Y. (1984). Asymptotic properties of the maximum likelihood estimate in the first order autoregressive process. *Ann. Inst. Statist. Math.* **36**, 119–128.

Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika* **72**, 653–659 (Erratum in volume 74, page 667).

Hayakawa, T. and Puri, M. L. (1985). Asymptotic expansions of the distributions of some test statistics. *Ann. Inst. Stat. Math. A* **37**, 95–108.

Kakizawa, Y. (1997). Higher order Bartlett-type adjustment. *J. Statist. Plann. Inference* **65**, 269–280.

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295–303.

Phillips, P. C. B. and Park, J. Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica* **56**, 1065–1083.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82**, 605–610.

Taniguchi, M. (1991). Third-order asymptotic properties of a class of test statistics under a local alternative. *J. Multivariate Anal.* **37**, 223–238.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer Series in Statistics, Springer-Verlag, New York.

Taniguchi, M., Hirukawa, J. and Tamaki, K. (2008). *Optima Statistical Inference in Financial Engineering*. Chapman & Hall/CRC, New York.

Taniguchi, M. and Watanabe, Y. (1994). Statistical analysis of curved probability densities. *J. Multivariate Anal.* **48**, 228–248.

Department of Law, Economics, Management and Quantitative Methods, University of Sannio, 82100 Benevento, Italy.

E-mail: acmonti@unisannio.it

Research Institute for Science & Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan.

E-mail: taniguchi@waseda.jp

(Received February 2016; accepted February 2017)