Testing Distributional Assumptions: A L-moment Approach

Ba Chu and Mark Salmon

Warwick Business School

Preliminary version. Comments are especially welcome

Abstract

Stein (1972, 1986) provides a flexible method for measuring the deviation of any probability distribution from a given distribution, thus effectively giving the upper bound of the approximation error which can be represented as the expectation of a Stein's operator. Hosking (1990, 1992) proposes the concept of L-moment which better summarizes the characteristics of a distribution than conventional moments (C-moments). The purpose of the paper is to propose new tests for conditional parametric distribution functions with weakly dependent and strictly stationary data generating processes (DGP) by constructing a set of the Stein equations as the L-statistics of conceptual ordered sub-samples drawn from the population sample of distribution; hereafter are referred to as the L-moment (GMLM) tests. The limiting distributions of our tests are nonstandard, depending on test criterion functions used in conditional L-statistics restrictions; the covariance kernel in the tests reflects parametric dependence specification. The GMLM tests can resolve the choice of orthogonal polynomials remaining as an identification issue in the GMM tests using the Stein approximation (Bontemps and Meddahi, 2005, 2006) because L-moments are simply the expectations of quantiles which can be linearly combined in order to characterize a distribution function. Thus, our test statistics can be represented as functions of the quantiles of the conditional distribution under the null hypothesis. In the broad context of goodness-of-fit tests based on order statistics, the methodologies developed in the paper differ from existing methods such as tests based on the (weighed) distance between empirical distribution and a parametric distribution under the null or the tests based on likelihood ratio of Zhang (2002) in two respects: 1) our tests are motivated by the L-moment theory and Stein's method; 2) offer more flexibility because we can select an optimal number of L-moments so that the sample size necessary for a test to attain a given level of power is minimal. Finally, we provide some Monte-Carlo simulations for *IID* data to examine the size, the power and the robustness of the GMLM test and compare with both existing moment-based tests and tests based on order statistics.

1 Introduction

Over recent years, there have been big advances in goodness-of-fit testing which can be classified into three groups:

1. tests based on empirical distribution function such as the Cramér-von Mises test, the Watson test, and the Anderson-Darling test (see e.g. Stephens (1974, 1978)) or tests based on the likelihood ratio (Zhang, 2002)
2. tests based on moments such as the Meddahi & Bomtemps (MB) test and the Kiefer & Salmon (KS) test
3. tests based on kernel density, e.g. Zheng (2000).

However, none of those tests are consistent against all types of alternatives to the null and not robust against outliers in a small sample.

This paper proposes an alternative approach to test for conditional distribution assumptions in dependent data by making use of L-moment theory and Stein's equation. Our simulations suggest that the GMLM tests can potentially avoid the above-mentioned problems of conventional goodness-of-fit tests by using a bootstrap procedure to estimate the covariance matrix of L-statistics.

The econometric literature on testing for valid conditional distributional assumptions for time-series data has received much recent interest. (See e.g. Bai 2003; Corradi and Swanson 2006; Li and Tkacz 2006; Zheng 2000.) The influential work was done by Andrews (1997) who proposes a conditional Kolmogorov test of model specification for a class of parametric nonlinear regression models by using a sample of independent data pairs. The test is shown to be consistent against all possible misspecification of distributions. However, the asymptotic distributions are not free from nuisance parameters, thus a parametric bootstrap procedure is used to obtain critical values for the test. Bai (2003); Bai and Chen (2006) make an important attempt to remove parameter uncertainty in the conditional Kolmogorov test by applying the probability integral transform to show that $F_t(y_t|\mathcal{F}_{t-1}, \theta_0)$ are IID uniform random variable on $[0, 1]$ and the martingale transformation of Khmaladze (1981) to transform empirical processes into martingales which then converge to the Brownian motions by the FCLT for martingales. As a result, Bai's tests have the Brownian motion limit. Zheng (2000) and Stinchcombe and White (1998) provide nonparametric tests of conditional ditributions by using the distances (e.g. Kullback-Leibler information criterion) between nonparametric estimators of distributions and the hypothesized distribution to construct valid moment conditions; then the method of consistent conditional moment test proposed by Bierens (1990) is extended to obtain efficient tests.

All of the above-mentioned methods rely on the topological distance between empirical/nonparametric estimates of distributions and the true distributions. Bontemps and Meddahi (2005, 2006) pursues a different approach by using the Stein equation (Stein, 1972, 1986) to construct moment conditions in their GMM tests. In these papers, they use the result proved by Stein: A random variable X has a standard normal distribution if and only if for any differential function f,

$$E[f^{'}(X) - Xf(X)] = 0. \tag{1.1}$$

Bontemps and Meddahi (2005) asserts that orthogonal polynomials can be used for f in the moment conditions (1.1) because any function f in the Hilbert space can be approximated by a linear combination of orthogonal polynomials. The heuristic argument behind this approach is: Supposing that the Carleman criterion (Chow and Teicher, 1997, pp. 303) the standard normal is completely characterized by its moments; thus there always exist ways to yield linear combinations of moments which are equal to zero, plausible choices for the functions f are polynomials of $\{X, X^2, \cdots, X^N\}$ in the $Span\{X, X^2, \cdots, X^N\}$ as N is sufficiently large so that the moment conditions as defined in (1.1) can completely characterize the standard normal (in this case $f^{'}(X) - Xf(X) = H(X)$, where $H(X)$ is a Hermite polynomial.) Moreover, Bontemps and Meddahi (2005) shows that the problem of parameter uncertainty can be dealt with by adding a moment condition into the GMM setting.

L-moments provide a promising alternative to summarize the shape of a distribution. Hosking (1990, 1992) has shown that L-moments have good properties as measures of distributional shape and useful for fitting distribution to data. Since the essential advantages of L-moments are 1) being more robust to the presence of outliers in the data 2) being less subject to bias in estimation and approximating their asymptotic values more closely in finite samples (Sankarasubramanian and Srinivasan, 1999); thus L-moments are useful in formulating goodness-of-fit test statistics.

The formulation of the GMLM tests follows two steps: First we assume strict stationarity for our DGP, then construct a set of moment conditions $E_{\mathbb{P}_{OS}}[\mathcal{A}\underset{(N\times 1)}{\widetilde{\boldsymbol{f}}(X)}|\boldsymbol{Z}] = 0$, where $\mathcal{A}$ is the Stein operator and $\widetilde{\boldsymbol{f}}'(X) = \{\widetilde{f}_1(X), \cdots, \widetilde{f}_r(X), \cdots, \widetilde{f}_N(X)\}$ with $\widetilde{f}_r(X)$ is the linear function of order statistics $\{X_{1:r}, \cdots, X_{r:r}\}$ for a given $\boldsymbol{Z}$ in $E_{\mathbb{P}_{OS}}[\widetilde{f}_r(X)] = \lambda_r$, ($\lambda_r$ is r-th L-moment); $\mathbb{P}_{OS}$ is the probability distribution of the order statistics. Second, we construct different test statistics associated with different criterion functions $g(\boldsymbol{Z})$: If $g(\boldsymbol{Z})$ is an exponential function, the test statistic is the classical GMM test and the limiting distribution is χ^2 with a covariance kernel dependent on serial dependence structure; if $g(\boldsymbol{Z})$ is an indicator function, the test statistics are nonstandard and the limiting distributions are mixed normal.

Although the theory developed in the paper generally holds for any strictly stationary DGP with a certain mixing level, we implement Monte Carlo simulations only for the *IID* cases because computing the estimate of the covariance matrix for serially dependent data is potentially time-consuming. Designing an efficient computation algorithm to do this could be a topic for future research. However, we think that this is sufficient to justify our purpose of constructing test statistics which are more robust against outliers and subject to less biases in a small sample. Simulation results suggest that our tests have reasonably good power and are robust against big outliers in rather small samples.

The remainder of the paper is organized as follows; in Section 2 we bring together derivations of L-moment conditions and formulation of the GMLM statistics. In Section 3, we discuss the limiting distribution of test statistics and the issues of choosing optimal number of L-moments. Some Monte Carlo simulations results are provided in Section 4.

2 Test Functions as Linear Combinations of Order Statistics

2.1 Stein's Equation

Let $\mathcal{M}(f)$ denotes a set of continuous or piecewise continuously differentiable test functions $f(\bullet)$ such that $E_{\mathbb{P}_0}[|f'(\bullet)|] < \infty$. Stein (1972, 1986) shows that for any real function $h(\bullet)$, there exist a function $f \in \mathcal{M}(f)$ and the following relation holds

$$h(X) - E_{\mathbb{P}_0}[h(X)] = \mathcal{A}f(X), \tag{2.1}$$

where $\mathcal{A}$ is the Stein operator for the probability distribution $\mathbb{P}_0$-this can be constructed, e.g. Barbour's generator approach (Barbour and Chen, 2005). Lemma 1 below states the functional form of $\mathcal{A}$ for the conditional distribution of the random variable X for a given random event Z, namely $p_\Theta(X|Z)$.

We first make some assumptions.

Assumption 1.

$$\sup_{X \in U_{X_0}(\delta)} \int_{U_{Z_0}(\gamma)} \sup_{\theta \in U_\theta(\epsilon)} |f(X)p_\theta(X|Z)|\lambda(dZ) < \infty,$$

$$\sup_{X^* \in U_{X_0}(\delta)} \int_{U_{Z_0}(\gamma)} \sup_{\theta \in U_\theta(\epsilon)} \left| \frac{\partial}{\partial X}\{f(X)p_\theta(X|Z)\} \right|_{X=X^*} \lambda(dZ) < \infty,$$

and there exists a constant M such that

$$\sup_{X \in U_{X_0}(\delta)} \int_{U_{Z_0}(\gamma)} \sup_{h \in U_0(\eta)} \frac{\Delta_h\{f(X)p_\theta(X|Z)\}}{h} \mathbf{1}\left\{Z : \frac{\Delta_h\{f(X)p_\theta(X|Z)\}}{h} > M\right\} \lambda(dZ) < \epsilon \ ,$$

where $U_X(\epsilon)$ is a neighborhood of X; ϵ, δ and η are arbitrarily small generic constants. $\Delta_h\{f(X)p_\Theta(X|Z)\} = f(X+h)p_\Theta(X+h|Z) - f(X)p_\Theta(X|Z)$.

Assumption 2. $\lim_{X \longrightarrow \partial\Omega_X} f(X)p_\theta(X|Z) = 0$ *for all $Z \in \Omega_Z$.*

Lemma 1. *Let X and Z denote random variables in the compact and separable spaces $(\Omega_X, \mathcal{F}_X, \mathbb{P})$*

and $(\Omega_Z, \mathcal{F}_Z, \lambda)$ *with bounded supports* $\text{supp}(X)$ *and* $\text{supp}(Z)$ *respectively; the parameter vector* θ *belongs to a bounded set* $\boldsymbol{\Theta}$*.*

Supposing that the functions $f(X) \in \mathcal{M}(f)$ *and* $p_\theta(X|Z)$ *belongs to the exponential family. Moreover,* $p_\theta(X|Z)$ *is measurable with respect to* (X, Z) *and satisfies Assumptions 1 and 2, then* $X \sim p_\theta(X|Z)$ *is equivalent to*

$$E_{p_\theta(X|Z)}[\mathcal{A}f(X)] = 0, \ \forall \ f \in \mathcal{M}(f), \tag{2.2}$$

where $\mathcal{A} = \frac{\partial}{\partial X} + \frac{\partial}{\partial X} \log p_\theta(X|Z)$*.*

Let $p_\theta^*(X|Z)$ *denotes a conjugate probability distribution of* $p_\theta(X|Z)$ *such that*

$$\frac{p_\theta^*(X|Z)}{p_\theta(X|Z)} = \kappa_\theta(X|Z),$$

where $\kappa_\theta(X|Z)$ *is a polynomial of* X*. Then, we have*

$$E_{p_\theta^*(X|Z)}[\mathcal{A}^* f(X)] = 0, \ \forall \ f \in \mathcal{M}(f), \tag{2.3}$$

where $\mathcal{A}^* = \frac{\partial}{\partial X} + \frac{\partial}{\partial X} \log p_\theta^*(X|Z)$*.*

Proof. In the first part, we show that $X \sim p_{\theta^*}(X|Z)$ implies that $E_{p_{\theta^*}(X|Z)}[\mathcal{A}f(X)] = 0$. Stein (1972) asserts that for a given measurable continuous or piecewise continuous function $h(X): \ \Omega_X \Longrightarrow \mathbb{R}$, we obtain

$$E_{p(Z)}\big[E_{p_\theta(X|Z)}[h(X)]\big] - E_{p(Z)}\big[E_{p_{\theta^*}(X|Z)}[h(X)]\big] = E_{p_\theta(X,Z)}[\mathcal{A}f(X)],$$

for a function $f \in \mathcal{M}(f)$. Thus $p_\theta(X|Z) \equiv p_{\theta^*}(X|Z)$ implies that $E_{p_\theta(X,Z)}[\mathcal{A}f(X)] = 0$. On the other hand, since $\frac{\partial}{\partial X} E_{p_\theta(X,Z)}[f(X)] = 0$ we have

$$E_{\lambda(dZ)}[E_{p_\theta(X|Z)}[\mathcal{A}f(X)]] = \frac{\partial}{\partial X} E_{\lambda(dZ)}[E_{p_\theta(X|Z)}[f(X)]], \tag{2.4}$$

where $\mathcal{A}$ can be found by the derivative in the RHS of (2.3), i.e.

$$E_{p_\Theta(X,Z)}[\mathcal{A}f(X)] := \frac{\partial}{\partial X}\int_{\text{supp}(Z)}\left\{\int_{\text{supp}(X)} f(X)p_\theta(X|Z)dX\right\}\lambda(dZ). \tag{2.5}$$

By virtue of the compactness and the separateness of the sample spaces Ω_X, Ω_Z with bounded supports, and the boundedness of the parameter space $\mathbf{\Theta}$, let $\{U_{X_{\bar{j}}}(\delta) : \bar{j} = 1, \cdots, \bar{J}_X\}$, $\{U_{Z_{\bar{j}}}(\gamma) : \bar{j} = 1, \cdots, \bar{J}_Z\}$, and $\{U_{\theta_{\bar{j}}}(\eta) : \bar{j} = 1, \cdots, \bar{J}_\Theta\}$ be finite covers of the bounded subsets $A \in \Omega_X$, $B \in \Omega_Z$, and $\Theta \in \mathbf{\Theta}$ so that $A \subseteq \bigcup_{\bar{j}=1}^{\bar{J}_X} U_{X_{\bar{j}}}(\delta)$, $B \subseteq \bigcup_{\bar{j}=1}^{\bar{J}_Z} U_{Z_{\bar{j}}}(\gamma)$, and $\Theta \subseteq \bigcup_{\bar{j}=1}^{\bar{J}_\Theta} U_{\theta_{\bar{j}}}(\gamma)$ respectively. In view of Assumption 1 we can easily show that

$$\begin{aligned}\sup_{X\in A}\int_B \sup_{\theta\in\Theta}|f(X)p_\theta(X|Z)|\lambda(dZ) &\leq \sup_{1\leq\bar{j}\leq\bar{J}_X}\sup_{X\in U_{X_{\bar{j}}}(\delta)}\sum_{\bar{j}=1}^{\bar{J}_Z}\int_{U_{Z_{\bar{j}}}(\gamma)}\sup_{1\leq\bar{j}\leq\bar{J}_\Theta}\sup_{\theta\in U_{\theta_{\bar{j}}}(\eta)}|f(X)p_\theta(X|Z)|\lambda(dZ)\\ &< \infty\end{aligned}$$

$$\begin{aligned}&\sup_{X^*\in A}\int_B \sup_{\theta\in\Theta}\left|\frac{\partial}{\partial X}\{f(X)p_\theta(X|Z)\}\right|_{X=X^*}\lambda(dZ)\\ &\leq \sup_{1\leq\bar{j}\leq\bar{J}_X}\sup_{X\in U_{X_{\bar{j}}}(\delta)}\sum_{\bar{j}=1}^{\bar{J}_Z}\int_{U_{Z_{\bar{j}}}(\gamma)}\sup_{1\leq\bar{j}\leq\bar{J}_\Theta}\sup_{\theta\in U_{\theta_{\bar{j}}}(\eta)}\left|\frac{\partial}{\partial X}\{f(X)p_\theta(X|Z)\}\right|_{X=X^*}\lambda(dZ)\\ &< \infty\end{aligned}$$

$$\begin{aligned}&\sup_{X\in A}\int_B \sup_{h\in U_0(\eta)}\frac{\Delta_h\{f(X)p_\theta(X|Z)\}}{h}\mathbf{1}\left\{Z : \frac{\Delta_h\{f(X)p_\theta(X|Z)\}}{h} > M\right\}\lambda(dZ)\\ &\leq \sup_{1\leq\bar{j}\leq\bar{J}_X}\sup_{X\in U_{X_{\bar{j}}}(\delta)}\sum_{\bar{j}=1}^{\bar{J}_Z}\int_{U_{Z_{\bar{j}}}(\gamma)}\sup_{1\leq\bar{j}\leq\bar{J}_\Theta}\sup_{\theta\in U_{\theta_{\bar{j}}}(\eta)}\sup_{h\in U_0(\eta)}\frac{\Delta_h\{f(X)p_\theta(X|Z)\}}{h}\\ &\quad\mathbf{1}\left\{Z : \frac{\Delta_h\{f(X)p_\theta(X|Z)\}}{h} > M\right\}\lambda(dZ) \leq \bar{J}_\Theta\epsilon,\end{aligned}$$

where $\bar{J}_\Theta\epsilon$ can be made small by making the generic constant M large enough. Therefore, an application of Theorem A.7 in Dudley (1999, pp. 391) enables us to take derivatives inside the first integral in (2.5). In view of Assumption 2 and the integration by part formula, it is straightforward to obtain $\mathcal{A}f = \frac{\partial}{\partial X} f(X) + \frac{\partial}{\partial x} \log p_\theta(X|Z)$.

To complete the proof, we need to show that $E_{p_\theta(X|Z)}[\mathcal{A}f(X)] = 0$, $\forall\ f \in \mathcal{M}(f)$ implies $X \sim p_\theta(X|Z)$. Supposing that $p_\theta(X|Z)$ is a probability density in the exponential family (Lehmann, 2000, pp. 56) as given by

$$p_\theta(X|Z) = C(\theta) \exp\left\{\sum_{j=1}^{K} Q_j(\theta) T_j(X, Z)\right\} \ell(X), \tag{2.6}$$

where $T_j(X, Z)$ are polynomials and $\ell(X)$ are polynomials which are nonnegative on the support of X. Particular cases of (2.6) are the distribution of a sample from a binomial, Poisson, or normal distribution. Now, by letting $f(X)$ be a polynomial (or its L_2-approximation by orthogonal polynomials) we can show that $\mathcal{A}f = f'(X) + f \log' p_\theta(X|Z)$ is the fraction of polynomials, say

$$\mathcal{A}f(X) = \frac{\sum_{i=0}^{m} g_i(Z, Q_1(\theta), \cdots, Q_K(\theta)) X^{m-i}}{\ell(X)}. \tag{2.7}$$

Therefore, the function $h(X)$ in the Stein equation $h(X) - E_{p_{\theta^*}(X|Z)}[h(X)] = \mathcal{A}f(X)$ is a polynomial such that $h(X)\ell(X)$ has the highest order m; and the coefficients of $h(X)\ell(X)$ can be found by matching with those in the numerator of (2.7). Since $E_{p_\theta(X|Z)}[\mathcal{A}f(X)] = 0$ implies that

$$E_{p_\theta(X|Z)}[h(X)] = E_{p_{\theta^*}(X|Z)}[h(X)]. \tag{2.8}$$

Supposing that the moments of the distribution (2.6) satisfy the Carleman condition in (Chow and Teicher, 1997, pp. 303), the conditional density $p_\theta(X|Z)$ is fully characterized by its moments. Hence, Equation (2.8) implies that $p_\theta(X|Z) = p_{\theta^*}(X|Z)$, $\forall\ f \in \mathcal{M}(f)$. Thus $X \sim p_\theta(X|Z)$. Equation (2.2) follows.

Since $\kappa_\theta(X|Z)$ is a polynomial and $p_\theta(X|Z)$ belongs to the exponential family, $p^*_\theta(X|Z)$ also belongs to the exponential family. Equation (2.3) follows. $\square$

2.2 L-moments and Stein's Equation

A probability distribution is conventionally summarized by its moments or cumulants. The sample estimators of moments are shown to suffer from biases due to the presence of outliers in the data or even the small size of samples. Hosking (1990, 1992) proposed a robust approach based on quantiles which are called L-moments. Instead of using the expectations of polynomials of a random variable X to summarize a distribution, Hosking's approach utilizes linear combinations of the expectations of the order statistics $X_{1:r} \leq X_{2:r} \leq \cdots \leq X_{r:r}$ for $1 \leq r \leq \infty$ which are constructed from random samples drawn from the distribution of X. Define the L-moments of X as

$$\begin{aligned} \lambda_r &= E_{\mathbb{P}_\theta^{(OS)}}[\widetilde{\lambda}_r] \\ &= r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E_{\mathbb{P}^{(OS)}_{\theta, r-k:r}}[X], \end{aligned} \tag{2.9}$$

where $\widetilde{\lambda}_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{r-k:r}$ and $\mathbb{P}^{(OS)}_{\theta, r-k:r}$ is the probability distribution of the order statistics $X_{r-k:r}$.

Moreover, if X_t is non-IID, for instance X_t has the DGP:

$$X_t = f(X_{t-1}, \xi_t),$$

where $f(X)$ is an analytical function and ξ_t is an IID noise. Note that the sample paths $\{X_t\}_{t=1}^T$ are initialized by X_0. Assume that X_0 is a realization of a stochastic process Z_s, $\mathbb{P}^{(OS)}_{\theta, r-k:r}$ is the function of the conditional distributions $P_\theta(X_t|\mathcal{F}_{t-1})$ where $\mathcal{F}_{t-1} := \{X_{t-1}, X_{t-2}, \cdots, X_1, Z, \xi_0, \xi_1, \xi_2, \cdots, \xi_t\}$ and the unconditional distribution $P(Z_s)$. The L-moment conditions for the sample $(X_{t_1}, \cdots, X_{t_r})|\mathcal{F}_0 \sim p_\theta(X_1, \cdots, X_r|\mathcal{F}_0)$ are given in Lemma 2.

Lemma 2. *There exists the nonlinear operators $\mathcal{A}_{r-k,r}$ such that*

$$E_{\mathbb{P}_\theta^{(OS)}}[\widetilde{\lambda}_r|\boldsymbol{Z}] - E_{\mathbb{P}_{\theta^*}^{(OS)}}[\widetilde{\lambda}_r|\boldsymbol{Z}] = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E_{\mathbb{P}^{(OS)}_{\theta, r-k:r}}[\mathcal{A}_{r-k,r} X|\boldsymbol{Z}].$$

Therefore, $X_t \sim P_{\theta^*}(X_t|\mathcal{F}_{t-1})$ *if and only if*

$$r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E_{\mathbb{P}^{(OS)}_{\theta^*,r-k:r}}[\mathcal{A}_{r-k,r}X|\boldsymbol{Z}] = 0,\ \forall\ r \geq 1,$$

where $\boldsymbol{Z}$ *is* $\mathcal{F}_0$*-measurable.*

Proof. The proof follows from Lemma 1. □

Now we make assumptions about the DGP of X_t and Z_s:

Assumption 3. *$\{X_t\}$ and $\{Z_s\}$ are 1) strictly stationary and uniform mixing (ϕ-mixing) of size $\frac{-r}{r-2}$ for some $r > 2$ and 2) their mixing coefficients satisfy*

$$\sum_{m=1}^{\infty} m^2\phi^{1/2}(m) < \infty.$$

In addition, let $\{X_{j:r}\}_{j=1}^{r}$ denotes an ordered random sample taken from the DGP of X_t. The conditional densities of the order statistics $X_{j:r}$ are given in the following lemma;

Lemma 3. *Supposing that the first part of Assumption 3 holds, then we obtain*

$$\begin{aligned} P\{X_{j:r} \in dx|\mathcal{F}_0\} &:= \frac{r!}{(j-1)!(r-j)!}F^{j-1}(x|\mathcal{F}_0)[1-F_1(x|\mathcal{F}_0)]^{r-j} \\ &\quad dF_1(x|\mathcal{F}_0) \\ P\{X_{j_1:r} \in dx_1, X_{j_2:r} \in dx_2, \cdots, X_{j_k:r} \in dx_k|\mathcal{F}_0\} &:= \frac{r!}{(j_1-1)!(j_2-j_1-1)!\dots(r-j_k)!}F_1^{j_1-1}(x_1|\mathcal{F}_0) \\ &\quad \{F_1(x_2|\mathcal{F}_0)-F_1(x_1|\mathcal{F}_0)\}^{j_2-j_1-1}\dots \\ &\quad \cdots\{1-F_1(x_k|\mathcal{F}_0)\}^{r-j_k}.\prod_{i=1}^{k} dF_1(x_i|\mathcal{F}_0), \end{aligned}$$

where $F_1(x|\mathcal{F}_0) = F(X_1 = x|\mathcal{F}_0)$.

Proof. We start with a lemma given in Bai and Chen (2006), who noted that a similar idea can be traced back to Rosenbaltt (1952).

Lemma 4. *If the random variable X_t conditional on $\mathcal{F}_{t-1}$ has the continuous cumulative distribution function (cdf) $F_1(X_t|\mathcal{F}_{t-1})$. Then the random variables $U_t = F_1(X_t|\mathcal{F}_{t-1})$ are IID uniform on $(0,1)$.*

We shall derive the bivariate conditional probability density for the ordered random sample $\{X_{j:r}\}_{j=1}^{r}$. The multivariate case can be shown similarly. First, define the bivariate random set

$$A(\omega) := \lim_{\substack{\delta_x \longrightarrow 0 \\ \delta_y \longrightarrow 0}} \Big\{\omega \in \Omega : \ \{x \leq X_{m:r}(\omega) \leq x + \delta_x\} \bigcap \{y \leq X_{\ell:r}(\omega) \leq y + \delta_y\}\Big\}$$

for $1 \leq m \leq \ell \leq r$ and $x \leq y$. We obtain

$$\begin{aligned} P\{A(\omega)|\mathcal{F}_0\} &= \lim_{\substack{\delta_x \longrightarrow 0 \\ \delta_y \longrightarrow 0}} \text{Prob}\Big\{\{X_{m-1:r} \leq x\} \bigcap \{x \leq X_{m:r} \leq x + \delta_x\} \bigcap \{x \leq X_{m+1:r} \leq \cdots \leq X_{\ell-1:r} \leq y\} \\ &\quad \bigcap \{y \leq X_{\ell:r} \leq y + \delta_y\} \bigcap \{y + \delta_y \leq X_{\ell+1:r} \leq X_{\ell+2:r} \leq \cdots \leq X_{r:r}\}\Big\} \\ &= \text{Prob}\left\{ \underbrace{\bigcap_{j=1}^{m-1} \{F(X_{\widetilde{t}_j}|\mathcal{F}_{\widetilde{t}_j-1}) \leq F_{\widetilde{t}_j}(x|\mathcal{F}_{\widetilde{t}_j-1})\}}_{\text{exact } r-1 \text{ observations of } \{X_{t_i}\}_{i=1}^{r} \text{ are less than x}} \bigcap \{F(X_{\widetilde{t}_m}|\mathcal{F}_{\widetilde{t}_m-1}) \in dF_{\widetilde{t}_m}(x|\mathcal{F}_{\widetilde{t}_m-1})\} \right. \\ &\quad \underbrace{\bigcap_{j=m+1}^{\ell-1} \{F_{\widetilde{t}_j}(x|\mathcal{F}_{\widetilde{t}_j-1}) \leq F(X_{\widetilde{t}_j}|\mathcal{F}_{\widetilde{t}_j-1}) \leq F_{\widetilde{t}_j}(y|\mathcal{F}_{\widetilde{t}_j-1})\}}_{\ell-m-1 \text{ observations of } \{X_{t_i}\}_{i=1}^{r} \text{ are in } [x,y]} \bigcap \{F(X_{\widetilde{t}_\ell}|\mathcal{F}_{\widetilde{t}_\ell-1}) \in dF_{\widetilde{t}_\ell}(y|\mathcal{F}_{\widetilde{t}_\ell-1})\} \\ &\quad \left. \underbrace{\bigcap_{j=\ell+1}^{r} \{F(X_{\widetilde{t}_{\ell+1}}|\mathcal{F}_{\widetilde{t}_{\ell+1}-1}) \geq F(y|\mathcal{F}_{\widetilde{t}_{\ell+1}-1})\}}_{\text{exact } r-\ell \text{ observations are greater or equal to } y} \right| \mathcal{F}_0 \Bigg\}, \end{aligned}$$

where $\widetilde{t}_j$ is the true time of the order statistics $X_{j:r}$ in the random sample. Lemma 4 and the strict stationarity of X_t (i.e. $F(X_t|\mathcal{F}_{t-1}) = F(X_{t+\tau}|\mathcal{F}_{t+\tau-1})$) yield

$$\begin{aligned} P\{A(\omega)|\mathcal{F}_0\} &= \text{Prob}\Bigg\{\bigcap_{j=1}^{m-1} \{U_{\widetilde{t}_j} \leq F_1(x|\mathcal{F}_0)\} \bigcap \{U_{\widetilde{t}_m} \in dF_1(x|\mathcal{F}_0)\} \bigcap_{j=m+1}^{\ell-1} \{F_1(x|\mathcal{F}_0) \leq U_{\widetilde{t}_j} \leq F_1(y|\mathcal{F}_0)\} \\ &\quad \bigcap \{U_{\widetilde{t}_\ell} \in dF_1(y|\mathcal{F}_0)\} \bigcap_{j=\ell+1}^{r} \{U_{\widetilde{t}_j} \geq F_1(y|\mathcal{F}_0)\}\Bigg\}. \end{aligned}$$

Since $\{U_{\widetilde{t}_j}\}_{j=1}^r$ are IID, the conventional argument for deriving bivariate probability densities for IID order statistics can be applied. Therefore, we obtain the main results. $\square$

Lemma 5. *Suppose that under the null hypothesis (H_0), X_t has the cdf $F_1(X_t|\mathcal{F}_{t-1})$ and the pdf $f_1(X_t|\mathcal{F}_{t-1})$ then the operator $\mathcal{A}_{j,r}X$ as given in Lemma 2 can be written as*

$$\mathcal{A}_{j,r}X := 1 + Xf_1(X|\mathcal{F}_0)\left\{\frac{j-1}{F_1(X|\mathcal{F}_0)} - \frac{r-j}{1-F_1(X|\mathcal{F}_0)}\right\} + X\frac{f_1^{'}(X|\mathcal{F}_0)}{f_1(X|\mathcal{F}_0)}.$$

Hence, the L-moment conditions can be written as

$$r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E_{P(\boldsymbol{Z})}\left[g(\boldsymbol{Z}_t).E_{\mathbb{P}^{(OS)}_{\theta^*,r-k:r}}[\mathcal{A}_{r-k,r}X|\boldsymbol{Z}]\right] = 0 \;\forall\; r \geq 1,$$

where $g(\bullet)$ is any function in the totally revealing space as defined in Stinchcombe and White (1998).

Remark 2.1. *According to Stinchcombe and White (1998), the revealing functions $g(\bullet)$ for constructing consistent specification tests should be chosen in the class $\mathcal{H}_G := \{h_\tau : h_\tau(x) = G(\widetilde{x}^{'}\tau),\ \tau \in \mathbb{R}^{k+1}\}$, where $\widetilde{x} = (1, x^{'})^{'}$ and G is analytic. The exponential function $\exp\{\tau^{'}\widetilde{x}\}$ as proposed by Bierens (1990) is a special case.*

By the same argument as Hosking (1992), in view of Lemma 5 and $r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k} = 0$, the L-moment conditions given in Lemma 2 can be represented as a probability weighed function (PWF);

$$\int_{\mathbb{R}^k} g(\boldsymbol{Z})\int_0^1\left\{X(F)\{f_1(X(F)|\boldsymbol{Z})\left[P^{*}_{r-1}(F) - P^{**}_{r-1}(F)\right] + \frac{f_1^{'}(X(F)|\boldsymbol{Z})}{f_1(X(F)|\boldsymbol{Z})}P^{***}_{r-1}(F)\}\right\}dFdP_{\boldsymbol{Z}} = 0, \tag{2.10}$$

where

$$\begin{aligned} P_r^* &= \sum_{k=0}^{r} p_{r,k}^* F^{k-1} \\ P_r^{**} &= \sum_{k=0}^{r} p_{r,k}^{**} \frac{F^k}{1-F} \\ P_r^{***} &= \sum_{k=0}^{r} p_{r,k}^{***} F^k, \end{aligned}$$

with

$$\begin{aligned} p_{r,k}^* &= (-1)^{r-k}(r-k)\binom{r}{k}\binom{r+k}{k} \\ p_{r,k}^{**} &= (-1)^{r-k}k\binom{r}{k}\binom{r+k}{k} \\ p_{r,k}^{***} &= (-1)^{r-k}\binom{r}{k}\binom{r+k}{k}. \end{aligned}$$

Supposing that we have samples of $(X_t^s, \boldsymbol{Z}_s)$ for $s = 1, \ldots, T$ and $t = 1, \ldots, T$, in view of Hosking (1992) the sample analogue of the L-moment condition in Lemma 2 is the standard U-statistics as given by

$$\mathfrak{M}_{T,r} = T^{-1} \sum_{s=1}^{T} \binom{T}{r}^{-1} \sum_{1 \le t_1 \le t_2 \le \cdots \le t_r \le T} \cdots \sum r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} g(\boldsymbol{Z}_s) \mathcal{A}_{r-k,r}(X_{t_{r-k}:T}^s | \boldsymbol{Z}_s),$$

where $\mathcal{A}_{r-k,r}$ is defined in Lemma 5; the inner sum is the sample estimate of L-moments for an initial value $\boldsymbol{Z}_0$; and the outer sum is the sample estimate of the moment of $g(\boldsymbol{Z}_0)$.

Moreover, the sample estimate of the L-moment condition based on the PWF is

$$\mathfrak{N}_{T,r} = T^{-1} \sum_{s=1}^{T} g(\boldsymbol{Z}_s) T^{-1} \sum_{t=1}^{T} \left\{ X_{t:T}^s \{ f_1(X_{t:T}^s | \boldsymbol{Z}_z) \left[P_{r-1}^*(t/T) - P_{r-1}^{**}(t/T) \right] + \frac{f_1'(X_{t:T}^s | \boldsymbol{Z}_s)}{f_1(X_{t:T}^s | \boldsymbol{Z}_s)} P_{r-1}^{***}(t/T) \} \right\}.$$

Since it is straightforward to prove that $|\mathfrak{M}_{T,r} - \mathfrak{N}_{T,r}| = o_p(T^{-1})$, the asymptotic distribution of the statistics $\mathfrak{M}_{T,r}$ is similar to that of $\mathfrak{N}_{T,r}$ which can be derived by using the central limit theorem for the weighted sum of functionals of the order statistics of weakly dependent data. Before deriving the asymptotic distribution for $\mathfrak{M}_{T,r}$, we are going to state Theorem 1.

Let $\mathfrak{T}_T := T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{K} P_{T,j}(t/T)h_j(X_{t:T})$ denotes the weighted sum of functionals of the order statistics $X_{t:T}$. Let $(a,b)^-$ and $(a,b)^+$ denote the minimum and the maximum of (a,b) respectively. We begin with some further assumptions

Assumption 4. *The functions h_j, $\forall\ j = 1,\cdots,K$ are absolutely continuous on $(F^{-1}(\epsilon), F^{-1}(1-\epsilon))$ for all $\epsilon > 0$, $h_j^{'}$ $\forall\ j = 1,\cdots,K$ exist a.s. and $|h_j^{'}| \leq R_j$ a.s., where R_j $\forall\ j = 1,\cdots,K$ are reproducing U-shaped (or increasing) function. For a large T, $|P_{T,j}| \leq \phi_j$ with $\int_0^1 q(F)R_j(X(F))\phi_j(F)dF < \infty$, where $q(t,\theta) = K[t(1-t)]^{1/2-\theta}$ $\forall\ 0 < \theta \leq 1/2$.*

Assumption 5. *$|h_j| \leq M_j[(1-dF)]^{-\alpha_j}$ for some generic constants M_j and $\{|P_{T,j}(0)|, |P_{T,j}(1)\}^+ = o(T^{1/2-\alpha_j})$.*

$$P_{T,j} \longrightarrow P_j \ a.s. \quad (2.11)$$

$$T^{1/2}\int_0^1 \{P_{T,j}(F)\mathbf{1}(X \in (F^{-1}(1/T), F^{-1}(1-1/T)) - P_j\}h_j(X(F))dF \longrightarrow 0. \quad (2.12)$$

Assumption 6. *$E[|g(\boldsymbol{Z})|^{2(r+\delta)} < \infty$ for some $r \geq 2$ and $\delta > 0$.*

Theorem 1. *Supposing that Assumptions 3-5 hold, then*

$$T^{1/2}(\mathfrak{T}_T - \mathfrak{T}) \Longrightarrow N(0,\sigma^2),$$

where $\mathfrak{T} = \int_{\mathbb{R}}\sum_{j=1}^{K} P_j(F)h_j(X(F))dF$; and

$$\sigma^2 = \int_{\mathbb{R}^2}\sigma(F(X),F(Y))\sum_{j=1}^{K}\sum_{k=1}^{K}\{P_j(F(X))P_k(F(Y))h_j^{'}(X)h_k^{'}(Y)\}dF(X)dF(Y)$$

with $\sigma(F,G) := [(F(X),G(Y))^- - F(X).G(Y)] + \sum_{j=2}^{\infty}(F_{1,j}(X,Y) - F(X).G(Y)) + \sum_{j=2}^{\infty}(F_{1,j}(Y,X) - F(X).G(Y))$, *where $F_{1,j}(X,Y) = Prob\{X_j = Y|X_1 = X\}$.*

Proof. See Appendix. □

Now let's define

$$\begin{aligned}\mathfrak{N}^{\circ}_{T,r,s} &= T^{-1}\sum_{t=1}^{T} X^s_{t:T}\Big\{f_1(X^s_{t:T}|\boldsymbol{Z}_s)\left[P^{*}_{r-1}(t/T)-P^{**}_{r-1}(t/T)\right]+\frac{f_1'(X^s_{t:T}|\boldsymbol{Z}_s)}{f_1(X^s_{t:T}|\boldsymbol{Z}_s)}P^{***}_{r-1}(t/T)\Big\},\\ h_{Z,1}(X) &= Xf_1(X|Z),\ h_{Z,2}(X)=Xf_1(X|Z),\ \text{and } h_{Z,3}=X\frac{f_1'(X|Z)}{f_1(X|Z)}.\end{aligned}$$

An application of Theorem 1 yields

Corollary 2. *If the functions $h_{Z,i}$ (i=1,2, and 3), P^{*}_{r-1}, P^{**}_{r-1} and P^{***}_{r-1} satisfy Assumptions 3-5, then*

$$T^{1/2}\mathfrak{N}^{\circ}_{T,r,s}\Longrightarrow \mathfrak{W}_s,$$

where $\mathfrak{W}_s\sim N(0,\sigma^2(\boldsymbol{Z_s}))$ and

$$\begin{aligned}\sigma^2(\boldsymbol{Z}_s) &= \int_{\mathbb{R}^2}\sigma(F_1(X|\boldsymbol{Z}_s),F_1(Y|\boldsymbol{Z}_s))\Big\{h_{Z,1}(X)P^{*}_{r-1}(F_1(X|\boldsymbol{Z}_s))+h_{Z,2}P^{**}_{r-1}(F_1(X|\boldsymbol{Z}_s))\\ &+ h_{Z,3}P^{***}_{r-1}(F_1(X|\boldsymbol{Z}_s))\Big\}\Big\{h_{Z,1}(Y)P^{*}_{r-1}(F_1(Y|\boldsymbol{Z}_s))+h_{Z,2}P^{**}_{r-1}(F_1(Y|\boldsymbol{Z}_s))\\ &+ h_{Z,3}P^{***}_{r-1}(F_1(Y|\boldsymbol{Z}_s))\Big\}dF_1(X|\boldsymbol{Z}_s)dF_1(Y|\boldsymbol{Z}_s).\end{aligned}$$

Furthermore, supposing that $E|\sigma(\boldsymbol{Z})|^{2(r+\delta)}<\infty$ for some $r>2$ and $\delta>0$ and Assumption 6 holds we have

$$T^{1/2}\mathfrak{N}_{T,r} \Longrightarrow N(0,\sigma^{*2}) \tag{2.13}$$

$$T\mathfrak{N}_{T,r} \Longrightarrow N_1(0,1)N_2(0,\sigma^{**2}) \tag{2.14}$$

where $N_1(0,1)$ and $N_2(0,\sigma^{**2})$ are independent normal random variables; $\sigma^{*2}=E^2_{P(\boldsymbol{Z})}\big[g(\boldsymbol{Z})\sigma(\boldsymbol{Z})\big]$; and

$$\sigma^{**2}=E[g^2(\boldsymbol{Z}_1)\sigma^2(\boldsymbol{Z}_1)]+2\sum_{j=2}^{\infty}E[g(\boldsymbol{Z}_1)\sigma(\boldsymbol{Z}_1)g(\boldsymbol{Z}_j)\sigma(\boldsymbol{Z}_j)].$$

Proof. The first part follows from Theorem 1. For the second part of the theorem, we first deal with (2.13), i.e. the asymptotic distribution of $T^{1/2}\mathfrak{N}_{T,r} = T^{-1}\sum_{s=1}^{T} g(\boldsymbol{Z}_s)T^{-1/2}\mathfrak{N}^{\circ}_{T,r,s}$. Hence, by the first part, we have

$$|T^{1/2}\mathfrak{N}_{T,r} - T^{-1}\sum_{s=1}^{T} g(\boldsymbol{Z}_s)\mathfrak{W}_s| = o_p(1). \tag{2.15}$$

Now, we need to derive the asymptotic distribution of $T^{-1}\sum_{s=1}^{T} g(\boldsymbol{Z}_s)\mathfrak{W}_s$. Since $T^{1/2}\mathfrak{N}^{\circ}_{T,r,s}$ is conditioned on the initialization process $\boldsymbol{Z}_s$ and the variance of the limiting random variable $\mathfrak{W}_s$ depends on $\boldsymbol{Z}_s$, we have

$$T^{-1}\sum_{s=1}^{T} g(\boldsymbol{Z}_s)\mathfrak{W}(0,\sigma^2(\boldsymbol{Z}_s)) \Longrightarrow \int_{\text{supp}(\boldsymbol{Z})} g(\boldsymbol{Z})\mathfrak{W}(0,\sigma^2(\boldsymbol{Z}))dF(\boldsymbol{Z}),$$

which is a mixed normal distribution.

Moreover, Lemma 7 asserts that $\mathfrak{W}$ is independent of $\boldsymbol{Z}_s$, thus $\int_{\text{supp}(\boldsymbol{Z})} g(\boldsymbol{Z})\mathfrak{W}(0,\sigma^2(\boldsymbol{Z}))dF(\boldsymbol{Z}) \overset{d}{=} N(0,1)\int_{\text{supp}(\boldsymbol{Z})} g(\boldsymbol{Z})\sigma(\boldsymbol{Z})dF(\boldsymbol{Z})$ and Equation (2.13) follows. To prove Equation (2.14), note that again by Lemma 7

$$T^{-1/2}\sum_{s=1}^{T} g(\boldsymbol{Z}_s)\mathfrak{W}(0,\sigma^2(\boldsymbol{Z}_s)) \overset{d}{=} N(0,1)T^{-1/2}\sum_{s=1}^{T} g(\boldsymbol{Z}_s)\sigma(\boldsymbol{Z}_s).$$

An application of the central limit theorem for stationary mixing process in Lin and Lu (1989, pp. 51) yields

$$T^{-1/2}\sum_{t=1}^{T} g(\boldsymbol{Z}_t)\sigma(\boldsymbol{Z}_t) \Longrightarrow N_2(0,\sigma^{**2})$$

□

By the same argument as Corollary 2, we obtain the asymptotic multivariate distribution of the multivariate L-statistics $\boldsymbol{\mathfrak{N}}'_{T,N} = \{\mathfrak{N}_{T,1},\cdots,\mathfrak{N}_{T,r},\cdots,\mathfrak{N}_{T,N}\}$;

Corollary 3.

$$T^{1/2}\boldsymbol{\mathfrak{N}}'_{T,N} \Longrightarrow \boldsymbol{N}(0,\boldsymbol{\Sigma}^*), \tag{2.16}$$

where $\mathbf{\Sigma}^* = \{\Sigma^*_{1,1}, \cdots, \Sigma^*_{r,q}, \cdots, \Sigma^*_{N,N}\}$, *and*

$$\begin{aligned}
\Sigma^*_{r,q} &:= E^2_{P(\boldsymbol{Z})}\big[g(\boldsymbol{Z})\Sigma_{r,q}(\boldsymbol{Z})\big],\\
\Sigma_{r,q}(\boldsymbol{Z}) &= \int_{\mathbb{R}^2} \sigma(F_1(X|\boldsymbol{Z}), F_1(Y|\boldsymbol{Z}))\Big\{h_{Z,1}(X)P^*_{r-1}(F_1(X|\boldsymbol{Z})) + h_{Z,2}P^{**}_{r-1}(F_1(X|\boldsymbol{Z}))\\
&+ h_{Z,3}P^{***}_{r-1}(F_1(X|\boldsymbol{Z}))\Big\}\Big\{h_{Z,1}(Y)P^*_{q-1}(F_1(Y|\boldsymbol{Z})) + h_{Z,2}P^{**}_{q-1}(F_1(Y|\boldsymbol{Z}))\\
&+ h_{Z,3}P^{***}_{q-1}(F_1(Y|\boldsymbol{Z}))\Big\}dF_1(X|\boldsymbol{Z})dF_1(Y|\boldsymbol{Z}).
\end{aligned}$$

2.3 Criterion Test Functions

2.3.1 Exponential revealing functions

In the light of Corollary 3, we obtain the GMLM test;

$$\mathcal{M}_g := T\mathfrak{N}_T'\mathbf{\Sigma}^{*-1}\mathfrak{N}_T \Longrightarrow \chi^2(N).$$

This suggests a standard GMM test as long as we can consistently estimate the dependence kernel $\sigma(F_1(X|\boldsymbol{Z}), F_1(Y|\boldsymbol{Z}))$ of the covariance matrix $\mathbf{\Sigma}^*$. We will come back to this issue in the next section.

However, the GMM test is not feasible unless we know the functional form of the the revealing function $g(\boldsymbol{Z})$. A candidate function is the exponential function as suggested by the following lemma;

Lemma 6 (Bierens, 1990, Lemma 1)**.** *Let ν be a random variable or vector satisfying $E|\nu| < \infty$ and let x be a bounded random vector in $\mathbb{R}^k$ such that $P\{E[\nu|x] = 0\} < 1$. Then the set $S = \{t \in \mathbb{R}^k : E[\nu.\exp\{t'x\}] = 0\}$ has Lebesgue measure zero.*

The issue of choosing t is discussed in Theorem 4 of Bierens (1990). That is, given a point $t_0 \in \mathbb{R}^k$, real numbers $\gamma > 0$ and $\rho \in (0, 1)$. Let $\widehat{t} = \mathrm{argmax}_{t\in\mathbb{R}^k}\widehat{\mathcal{M}}(t)$, where $\widehat{\mathcal{M}}$ signifies the substitution of $\mathbf{\Sigma}^*$ in $\mathcal{M}$ with an estimate. Then $\widetilde{t}$ can be chosen such that

$$\widetilde{t} = t_0 \text{ if } \widehat{\mathcal{M}}(\widehat{t}) - \widehat{\mathcal{M}}(t_0) \leq \gamma T^{\rho};\ \widetilde{t} = \widehat{t} \text{ if } \widehat{\mathcal{M}}(\widehat{t}) - \widehat{\mathcal{M}}(t_0) > \gamma T^{\rho}.$$

2.3.2 Indicator revealing functions

The exponential revealing function may not contain full information about $\boldsymbol{Z}$ which is $\mathcal{F}_0$-measurable in the conditional moment equation $E[v|\boldsymbol{Z}] = 0$ since any function f rather than g can contain some information about $E[v|\boldsymbol{Z}] = 0$. The idea of using an indicator function to capture all information about $\boldsymbol{Z}$ in the conditional moment equation is motivated by Billingsley (1995)'s result

$$E[v|\boldsymbol{Z}] = 0 \equiv E[v\mathbf{1}(\boldsymbol{Z} \leq \boldsymbol{x})] = 0,\ \forall\ \boldsymbol{x} \in \text{supp}(\boldsymbol{Z}).$$

It is straightforward to see that the conditional expectation in Lemma 5 can also be written as

$$r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E_{P(\tau)}\left[\mathbf{1}(\tau(\boldsymbol{Z}) \leq \underline{\boldsymbol{\tau}})E_{\mathbb{P}^{(OS)}_{\theta^*,r-k:r}}[\mathcal{A}_{r-k:r}X|\boldsymbol{Z}]\right] = 0,$$

where $P(\tau)$ is the distribution of $\tau(\boldsymbol{Z})$ with the density ∂P_τ and $\tau(\boldsymbol{Z})$ is a monotonic transformation of $\boldsymbol{Z}$ into $[0,1]^{\dim(\boldsymbol{Z})}$ ($\dim(\boldsymbol{Z})$ is the dimension of $\boldsymbol{Z}$). Let's define the statistics

$$\mathfrak{H}(\underline{\boldsymbol{\tau}}) = T^{-1/2}\sum_{s=1}^{T}\underset{N\times 1}{\partial P_{\boldsymbol{\tau}}^{-1/2}} \odot \boldsymbol{\mathcal{K}}[\mathbf{1}(\tau(\underset{N\times 1}{\boldsymbol{Z}_s}) \leq \underline{\boldsymbol{\tau}}) - P_{\boldsymbol{\tau}}(\underline{\boldsymbol{\tau}})] \odot \underset{N\times N}{\mathbf{\Sigma}^{-1/2}}.T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\mathcal{A}}^*(\underset{N\times 1}{X^s_{t:T}}, \boldsymbol{Z}_s), \tag{2.17}$$

where $\boldsymbol{\mathcal{A}}^*(X^s_{t:T}, \boldsymbol{Z}_s) = \{\mathcal{A}^*_1(X^s_{t:T}, \boldsymbol{Z}_s), \cdots, \mathcal{A}^*_r(X^s_{t:T}, \boldsymbol{Z}_s), \cdots, \mathcal{A}^*_N(X^s_{t:T}, \boldsymbol{Z}_s)\}$,

$$\mathcal{A}^*_r(X^s_{t:T}, \boldsymbol{Z}_s) = X^s_{t:T}\left\{f_1(X^s_{t:T}|\boldsymbol{Z}_s)\left[P^*_{r-1}(t/T) - P^{**}_{r-1}(t/T)\right] + \frac{f_1'(X^s_{t:T}|\boldsymbol{Z}_s)}{f_1(X^s_{t:T}|\boldsymbol{Z}_s)}P^{***}_{r-1}(t/T)\right\}$$

and

$$\boldsymbol{\mathcal{K}}[\mathbf{1}(\tau(\boldsymbol{Z}) \leq \underline{\boldsymbol{\tau}}) - P_\tau(\underline{\boldsymbol{\tau}})] = \mathbf{1}(\tau(\boldsymbol{Z}) \leq \underline{\boldsymbol{\tau}}) - P_\tau(\underline{\boldsymbol{\tau}}) + \int_0^{P_\tau(\underline{\boldsymbol{\tau}})}\frac{\mathbf{1}(\tau(\boldsymbol{Z}) \leq s) - P_\tau(\underline{\boldsymbol{\tau}})}{1-s}ds$$

is the martingale transformation proposed by Khmaladze (1981).

Since by Corollary 3, the term behind $\odot$ in (2.17) converges to the multivariate normality. In addition, an application of the central limit theorem for empirical processes yields

$$T^{-1/2}\sum_{s=1}^{T}\partial P_{\tau}^{-1/2}\odot\boldsymbol{\mathcal{K}}[\mathbf{1}(\tau(\boldsymbol{Z}_s)\leq\underline{\boldsymbol{\tau}})-P_{\tau}(\underline{\boldsymbol{\tau}})]\Longrightarrow\boldsymbol{B}(\underline{\boldsymbol{\tau}}).$$

Thus, in view of Lemma 7, we obtain

$$\mathfrak{H}(\underline{\boldsymbol{\tau}})\Longrightarrow\boldsymbol{B}(\underline{\boldsymbol{\tau}})\odot N(0,\boldsymbol{I}),$$

where $\boldsymbol{I}$ is a diagonal unity matrix; $\boldsymbol{B}(\underline{\boldsymbol{\tau}})\sim\boldsymbol{N}(0,(\underline{\boldsymbol{\tau}}^{'},\underline{\boldsymbol{\tau}}^{'})^{-})$ is the Brownian sheet, and $(\boldsymbol{a},\boldsymbol{b})^{-}$ denotes the minimum operations between all possible pairs of $\boldsymbol{a}$ and $\boldsymbol{b}$. Hence, we can define the following statistics:

$$\begin{aligned}\mathcal{M}^{*} &= \sup_{\underline{\boldsymbol{\tau}}\in[0,1]^{\dim(\boldsymbol{Z})}}\widehat{\mathfrak{H}}^{'}(\underline{\boldsymbol{\tau}})\widehat{\mathfrak{H}}(\underline{\boldsymbol{\tau}})\\ \mathcal{M}^{**} &= \int_{[0,1]^{\dim(\boldsymbol{Z})}}\widehat{\mathfrak{H}}^{'}(\underline{\boldsymbol{\tau}})\widehat{\mathfrak{H}}(\underline{\boldsymbol{\tau}})d\underline{\boldsymbol{\tau}}.\end{aligned}$$

The main results can be summarized in the following theorem:

Theorem 4.

$$\begin{aligned}\mathcal{M}^{*} &\Longrightarrow \sup_{\underline{\boldsymbol{\tau}}\in[0,1]^{N}}[N^{'}(0,\boldsymbol{I})\odot\boldsymbol{B}^{'}(\underline{\boldsymbol{\tau}})\boldsymbol{B}(\underline{\boldsymbol{\tau}})\odot N(0,\boldsymbol{I})].\\ \mathcal{M}^{**} &\Longrightarrow \int_{[0,1]^{N}}[N^{'}(0,\boldsymbol{I})\odot\boldsymbol{B}^{'}(\underline{\boldsymbol{\tau}})\boldsymbol{B}(\underline{\boldsymbol{\tau}})\odot N(0,\boldsymbol{I})]d\underline{\boldsymbol{\tau}}.\end{aligned}$$

2.3.3 Estimation of covariance matrices

As we can see, the tests are feasible if ∂P_{τ} and $\boldsymbol{\Sigma}(\boldsymbol{Z})$ can be estimated consistently. Since it is trivial that ∂P_{τ} can be estimated by nonparametric kernels which are shown to be consistent, we need to discuss the issue of estimating the covariance matrix $\mathbf{\Sigma}(\boldsymbol{Z})$. Let's define

$$\mathcal{A}_r^{**}(X,\boldsymbol{Z}) = h_{Z,1}(X)P_{r-1}^{*}(F_1(X|\boldsymbol{Z})) + h_{Z,2}P_{r-1}^{**}(F_q(X|\boldsymbol{Z})) + h_{Z,3}P_{r-1}^{***}(F_1(X|\boldsymbol{Z})).$$

We can define the estimator

$$\widehat{\Sigma}_{r,q}(\boldsymbol{Z}) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{\sigma}(F_1(X_t|\boldsymbol{Z}),F_1(X_s|\boldsymbol{Z}))\mathcal{A}_r^{**}(X_t,\boldsymbol{Z})\mathcal{A}_q^{**}(X_s,\boldsymbol{Z}),$$

where

$$\begin{aligned}
\widehat{\sigma}(F_1(X_t|\boldsymbol{Z}),F_1(X_s|\boldsymbol{Z})) &= [(F_1(X_t|\boldsymbol{Z}),F_1(X_s|\boldsymbol{Z}))^{-} - F_1(X_t|\boldsymbol{Z})F_1(X_s|\boldsymbol{Z})]\\
&+ \sum_{j=2}^{\infty}[\widehat{F}_{1,j}(X_t,X_s|\boldsymbol{Z}) - F_1(X_t|\boldsymbol{Z})F_1(X_s|\boldsymbol{Z})]\\
&+ \sum_{j=2}^{\infty}[\widehat{F}_{1,j}(X_s,X_t|\boldsymbol{Z}) - F_1(X_t|\boldsymbol{Z})F_1(X_s|\boldsymbol{Z})]\\
\widehat{F}_{1,j}(X,Y|\boldsymbol{Z}) &= \widehat{F}_{1,j}(Y|X)F_1(X|\boldsymbol{Z})
\end{aligned}$$

and

$$\widehat{F}_{1,j}(\widetilde{X}|\widetilde{Y}) = \frac{\int_0^{\widetilde{X}}\int_0^{\widetilde{Y}}\widehat{p}_{1,j}(\widetilde{x},\widetilde{y})d\widetilde{x}d\widetilde{y}}{\int_0^{\widetilde{X}}\widehat{p}(\widetilde{x})d\widetilde{x}} \text{ ($\widetilde{X}$ is the monotonic transformation of X from $\mathbb{R}$ into $[0,1]$)}$$

with the kernel estimates[1]

$$\begin{aligned}
\widehat{p}_{1,j}(\widetilde{x}_1,\widetilde{x}_j) &= \frac{1}{(T-j)a(T-j)b(T-j)}\sum_{t=1}^{T-j}\mathcal{K}_1\left(\frac{\widetilde{x}_1-\widetilde{X}_t}{a(T-j)}\right)\mathcal{K}_2\left(\frac{\widetilde{x}_j-\widetilde{X}_{t+j}}{b(T-j)}\right)\\
\widehat{p}(\widetilde{x}_j) &= \frac{1}{Tc(T)}\sum_{t=1}^{T}\mathcal{K}_2\left(\frac{\widetilde{x}_j-\widetilde{X}_t}{c(T)}\right),
\end{aligned}$$

For the consistency of $\widehat{\mathbf{\Sigma}}(\boldsymbol{Z})$, we make some assumptions:

[1] An alternative method for estimating a conditional distribution function is discussed in Hall et al. (1999). Their approach is shown to suffer less bias than the Nadaraya-Watson estimators and produces distribution estimators that always lie between 0 and 1.

Assumption 7. *The true density* $p(x)$ *is continuously differentiable to integral order* $\omega \geq 1$ *on* $[0, 1]$ *and* $\sup_{x\in[0,1]} |\partial^{\mu} p(x)| < \infty \ \forall \ \mu$ *with* $\mu \leq R$.

Assumption 8. $F_{1,j}(X|Y) : \ [0, 1]^2 \longrightarrow [0, 1]$ *is continuously differentiable of order* $\omega + 2$ *with* $\sup_{(x,y)\in[0,1]^2} |D^r F_{1,j}(X|Y)| < \infty$ *for every* $0 \leq r \leq \omega + 2$.

Assumption 9. *The bandwidth parameters* $a(T)$, $b(T)$ *and* $c(T)$ *satisfy* $C_1 a_{1,T} \leq a(T) \leq C_2 a_{2,T}$, $D_1 b_{1,T} \leq b(T) \leq D_2 b_{2,T}$ *and* $E_1 c_{1,T} \leq c(T) \leq E_2 c_{2,T}$ *respectively for some sequences of bounded positive constants* $\{a_{1,T}, b_{1,T}, c_{1,T}\}_{T\geq1}$ *and* $\{a_{2,T}, b_{2,T}, c_{2,T}\}_{T\geq1}$ *and some positive constants* C_1, C_2, D_1, D_2, E_1, E_2. *There exists a sequence of positive constants* $e_T \longrightarrow \infty$ *such that* $e_T/c(T)^{\omega} \longrightarrow \infty$, $e_T/a^{\omega}(T-j) \longrightarrow \infty$, $e_T/b^{\omega}(T-j) \longrightarrow \infty$, *and* $e_T T^{1/2} b(T-j) \longrightarrow \infty$.

Assumption 10 (Standard assumptions on the kernels)**.** *The kernel* $\mathcal{K}$ *satisfies (i)* $\sup_{x\in[0,1]} |\mathcal{K}(x)| < \infty$, $\int_{[0,1]} \mathcal{K}(x)dx = 1$, $\int_{[0,1]} x^r \mathcal{K}(x)dx = 0$ *if* $1 \leq r \leq \omega - 1$ *and* $\int_{[0,1]} x^r \mathcal{K}(x)dx < \infty$ *if* $r = \omega$*; (ii)* $\mathcal{K}$ *has a absolutely integrable Fourier transform, i.e.* $\int_{[0,1]} \exp\{i\tau' x\}\mathcal{K}(x)dx < \infty$.

Assumption 11. $\sup_{\boldsymbol{Z}\in[0,1]^k} \|h_{\boldsymbol{Z},i}(X, \boldsymbol{Z})\|_4 < \infty \ \forall \ i = 1, 2, 3.$

Theorem 5. *Under Assumptions 3 and 7-11, the estimate* $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{Z})$ *is consistent, i.e.*

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{Z}) \overset{P}{\Longrightarrow} \boldsymbol{\Sigma}(\boldsymbol{Z}).$$

Proof. As shown by Andrews (1991), Assumptions 3, 7, 8-10 implies that

$$\sup_{x\in[0,1]} |\widehat{p}(x) - p(x)| = O_p(T^{-1/2}c^{-1}(T)) + O_p(c^{\omega}(T)).$$

Moreover, by replicating the proof in Komunjer and Vuong (2006), we can show that

$$\sup_{(x,y)\in[0,1]^2} |\widehat{p}_{1,j}(x, y) - p_{1,j}(x, y)| = O_p(T^{-1/2}b^{-1}(T-j)) + O_p(a^{\omega}(T-j)) + O_p(b^{\omega}(T-j)).$$

Hence, we have

$$\begin{aligned} e_T^{-1} \sup_{x\in[0,1]} |\widehat{p}(x) - p(x)| &= o_p(1), \\ e_T^{-1} \sup_{(x,y)\in[0,1]^2} |\widehat{p}_{1,j}(x,y) - p_{1,j}(x,y)| &= o_p(1). \end{aligned}$$

It follows that

$$\begin{aligned} e_T^{-1} \left| \int_0^{\widetilde{X}} \widehat{p}(x)dx - \int_0^{\widetilde{X}} \widehat{p}(x)dx \right| &\le \widetilde{X} e_T^{-1} \sup_{x\in[0,1]} |\widehat{p}(x) - p(x)| \\ &= o_p(1), \qquad (2.18) \\ e_T^{-1} \left| \int_0^{\widetilde{X}} \int_0^{\widetilde{Y}} \widehat{p}_{1,j}(x,y)dxdy - \int_0^{\widetilde{X}} \int_0^{\widetilde{Y}} p_{1,j}(x,y)dxdy \right| &\le \widetilde{X}\widetilde{Y} e_T^{-1} \sup_{(x,y)\in[0,1]^2} |\widehat{p}(x,y) - p(x,y)| \\ &= o_p(1). \qquad (2.19) \end{aligned}$$

Equations (2.18) and (2.19) and the Slutsky lemma yield $|\widehat{F}_{1,j}(\widetilde{X}|\widetilde{Y}) - F_{1,j}(\widetilde{X},\widetilde{Y})| = o_p(1)$. Thus, $|\widehat{\sigma}(\bullet,\bullet) - \sigma(\bullet,\bullet)| = o_p(1)$.

Since

$$\begin{aligned} \left|\widehat{\Sigma}_{r,q}(\boldsymbol{Z}) - \Sigma_{r,q}(\boldsymbol{Z})\right| &\le T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left|[\widehat{\sigma}(\bullet,\bullet) - \sigma(\bullet,\bullet)]\mathcal{A}_r^{**}(X_t,\boldsymbol{Z})\mathcal{A}_q^{**}(X_s,\boldsymbol{Z})\right| \\ &\le \sup_{(t,s)\in[1,T]^2} \|\widehat{\sigma}(\bullet,\bullet) - \sigma(\bullet,\bullet)\|_2 T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \|\mathcal{A}_r^{**}(X_t,\boldsymbol{Z})\mathcal{A}_q^{**}(X_s,\boldsymbol{Z})\|_2 \\ &\le \sup_{(t,s)\in[1,T]^2} \|\widehat{\sigma}(\bullet,\bullet) - \sigma(\bullet,\bullet)\|_2 \|\mathcal{A}_q^{**2}(X_0,\boldsymbol{Z})\|_2 T^{-2} \sum_{t,s=1}^{T} \|E[\mathcal{A}_r^{**2}(X_{t-s},\boldsymbol{Z})] \\ &+ 2\phi^{1/2}(t-s)\|\mathcal{A}_r^{**2}(X_{t-s},\boldsymbol{Z})\|_2, \end{aligned}$$

where the third inequality follows from Lemma 8.

The dominated convergence theorem implies that: $\sup_{(t,s)\in[1,T]^2} \|\widehat{\sigma}(\bullet,\bullet) - \sigma(\bullet,\bullet)\|_2 = o_p(1)$.

Assumptions 3 and 11 imply that

$$T^{-2}\sum_{t,s=1}^{T}\|E[\mathcal{A}_r^{**2}(X_{t-s},\boldsymbol{Z})]+2\phi^{1/2}(t-s)\|\mathcal{A}_r^{**2}(X_{t-s},\boldsymbol{Z})\|_2<\infty.$$

Hence, the theorem follows. □

2.4 Optimal Number of L-statistics

Because given a particular sample size T the test statistics $\mathcal{M}_g$, $\mathcal{M}^*$, and $\mathcal{M}^{**}$ depends on the number of L-statistics (N) implied by the Stein equation as given in Lemma 1, the power functions of the tests are sensitive to the choice of N. Hence, N can be chosen in a such way that the sample size needed for the tests at a given level to have the power not less than a given bound at a point of an alternative hypothesis is minimum. In this section, we apply the principle proposed by Bahadur (1960, 1967) to infer the optimal number of the GMLM conditions used in our test statistics for independent data.

Consider the problem of testing the hypothesis

$$H_0: \ (X,Z)\sim F_0(X|Z)$$

(i.e., the conditional distribution is $F_0(X|Z)$) against the alternative

$$H_1: \ (X,Z)\sim G_\theta(X|Z)$$

(i.e., the conditional distribution is $G_\theta(X|Z)$ for $\theta\in\Theta$[2]). Let $\mathfrak{M}_{T,N}(F_T)$ denotes a test statistic depending on the empirical distribution of X for a given Z ($F_T(X|Z)$). $\mathfrak{M}_{T,N}(F_T)$ can be one of the tests $\mathcal{M}_g$, $\mathcal{M}^*$, and $\mathcal{M}^{**}$ as defined in Section 2.3 with the critical region given by

$$\{\mathfrak{M}_{T,N}(F_T)\geq c\},\ \text{where } c \text{ is some real number.}$$

[2] $\lim_{\theta\longrightarrow\infty}G_\theta(X|Z)=F_0(X|Z)$ for nested hypotheses and $G_\theta(X|Z)\neq F_0(X|Z)$, $\forall\,\theta\in\Theta$ for non-nested hypotheses

The *power function* of this test is the quantity $P\{\mathfrak{M}_{T,N}(F_T) \geq c\}$ that varies with respect to the sequence of alternatives $G_\theta(X|Z)$, $\forall\theta \in \Theta$; and its size is $P\{\mathfrak{M}_{T,N}(F_T) \geq c| \ H_0 \text{ is true}\}$. Let $P\{\mathfrak{M}^{G_\theta}_{T,N}(F_T) \geq c\}$ and $P\{\mathfrak{M}^{F_0}_{T,N}(F_T) \geq c\}$ denotes the power function and the size of this test respectively.

Now define for any $\beta \in (0,1)$ a real sequence $c_T := c_T(\beta, \theta)$ which satisfies the double inequality

$$P\{\mathfrak{M}^{G_\theta}_{T,N}(F_T) > c_T\} \leq \beta < P\{\mathfrak{M}^{G_\theta}_{T,N}(F_T) \geq c_T\}.$$

Then

$$\alpha_T(\beta, \theta, N) := P\{\mathfrak{M}^{F_0}_{T,N} \geq c_T\}$$

is the minimum size of the test $\mathfrak{M}_{T,N}$ for which the power at a point $\theta \in \Theta$ is not less than β. Thus, it is clear that the positive number

$$N_{\mathfrak{M}}(\alpha, \beta, \theta, N) := \min\{\tau : \ \alpha_T(\beta, N) \leq \alpha \text{ for all } m \geq \tau\}$$

is the minimum sample size necessary for the test at a level α to attain the power not less than β at a point in the set of alternatives.

Suppose that the sample $\{X^s_t, Z_s\}^T_{t=1}$ for $s = 1, \ldots, T$ are independent identically distributed draws from a distribution for given contexts Z_s under the alternative hypothesis. In other words, the sampling scheme is done in two steps: 1) draw an IID sample $\{Z_s\}^T_{s=1}$; 2) for each Z_s draw an IID sample of $\{X^s_t\}^T_{t=1}$. Bahadur (1960) shows that

$$N_{\mathfrak{M}}(\alpha, \beta, \theta, N) \sim \frac{2\ln(1/\alpha)}{c_{\mathfrak{M}}(\theta, N)}, \tag{2.20}$$

where $c_{\mathfrak{M}}(\theta, N)$ is the *Bahadur exact slope of the sequence of the test* $\mathfrak{M}_{T,N}(F_T)$.

The *optimal choice of* N is the positive integer that minimizes $N_{\mathfrak{M}}(\alpha, \beta, \theta, N)$ or maximizes the

Bahadur exact slope $c_{\mathfrak{M}}(\theta, N)$, i.e.

$$N^* = \arg\max_N c_{\mathfrak{M}}(\theta, N).$$

The Bahadur exact slope can be approximated by using the following theorem

Theorem 6 (Bahadur (1960, 1967)). *Let for a sequence $\{\mathfrak{M}_{T,N}\}$ the following two conditions be fulfilled;*

$$\mathfrak{M}_{T,N}^{G_\theta} \xrightarrow{G_\theta} b(\theta, N) \tag{2.21}$$

$$\lim_{T\longrightarrow\infty} T^{-1} \ln P\{\mathfrak{M}_{T,N}^{F_0} \geq t\} = -k(t) \tag{2.22}$$

for each t from an open interval I on which f is continuous and $\{b(\theta, N),\ \theta \in \Theta\} \in I$. Then

$$c_{\mathfrak{M}}(\theta, N) = 2k(b(\theta, N)).$$

The first part of Theorem 6 (Equation 2.21) easily follows from the central limit theorem for L-statistics. To derive the second part (Equation 2.22), an application of the large deviations theorem for functionals of empirical distributions as given in Hoadley (1967) results

$$\lim_{T\longrightarrow\infty} T^{-1} \ln P\{\mathfrak{M}_{T,N}^{F_0} \geq a + u_T\} = -KL(\Omega_a, F_0), \tag{2.23}$$

where $\Omega_a := \{F \in \Lambda_1 :\ \mathfrak{M}_{\infty,N}^{F_0}(F) \geq a\}$; Λ_1 is the space of conditional distribution functions on $\mathbb{R}$; $u_T \longrightarrow 0$; $KL(Q, P)$ is the Kullback-Leibler information distance defined as

$$KL(Q,P) = \begin{cases} \int \ln(\frac{dQ}{dP})dQ \text{ if } Q \text{ is absolutely continuous with respect to } P \\ +\infty \text{ Otherwise} \end{cases}$$

and $KL(\Omega_a, F_0) = \inf\{KL(Q, F_0) :\ Q \in \Omega_a\}$. Keep in mind that $K(\varnothing, P) = +\infty$.

Now, we use the transformation $F_0(X|Z):\ \mathbb{R}\longrightarrow[0,1]$ to map the domain of distribution functions on the space Λ_1 on to $[0,1]$, i.e.,

$$\begin{aligned} Q(X|Z) &= Q(F_0^{-1}(t|Z)) = Q^*(t|Z) \\ dQ^*(t) &= q^*(t)dt, \end{aligned}$$

where F_0^{-1} is the inverse function of F_0. Let's define the following quantities

$$\begin{aligned} Q^\circ(t,\boldsymbol{A}) &= \sum_{k=0}^{\mathbb{K}} \frac{(-1)^k}{k!}\phi^{(k)}(t)(\boldsymbol{a}_k^{'}Z) \\ q^\circ(t,\boldsymbol{A}) &= \sum_{k=0}^{\mathbb{K}} \frac{(-1)^k}{k!}\phi^{(k+1)}(t)(\boldsymbol{a}_k^{'}Z), \end{aligned}$$

where $\boldsymbol{A}=\{\boldsymbol{a}_0^{'},\ldots,\boldsymbol{a}_{\mathbb{K}}^{'}\}$ is a $K\times\dim(Z)$ vector ($\dim(Z)$ is the dimension of Z and $\boldsymbol{a}_0^{'}Z=1$); $\phi^{(k)}(t)$ is the k-th derivative of the standard normal c.d.f $\phi(t)=(2\phi)^{-1}\int_{-\infty}^{x}\exp(-y^2/2)dy$; and

$$\begin{aligned} f_{\mathbb{K},0}(\boldsymbol{A}) &= \int_{[0,1]} q^\circ(t,\boldsymbol{A})\ln q^\circ(t,\boldsymbol{A})dt \\ f_{\mathbb{K},1}(\boldsymbol{A}) &= \int_{[0,1]} q^\circ(t,\boldsymbol{A})-1 \\ f_{\mathbb{K},2}(\boldsymbol{A},N) &= \int_{[0,1]} \left\{\mathfrak{m}_{\infty,N}(Q^\circ(t,\boldsymbol{A}))-\mathfrak{m}_{\infty,N}(G_\theta(t|Z))\right\}dt, \end{aligned}$$

where $\mathfrak{M}_{\infty,N}(F)=\int_{[0,1]}\mathfrak{m}_{\infty,N}(F(t))dt$ and $G_\theta(t|Z)=G_\theta(F_0^{-1}(t|Z))$.

In the light of Theorem 6 and Equation 2.23, we state the following theorem

Theorem 7. *The optimal number of L-statistics is given by*

$$N^{**} := \arg\lim_{\mathbb{K}\longrightarrow\infty}\max[f_{\mathbb{K},0}(\boldsymbol{A}(N_f^*)), f_{\mathbb{K},0}(\boldsymbol{A}(N_c^*))],$$

where N_f^* and N_c^* are the floor and ceiling values of N^* which is the unique solution to the following *linear programming problem;*

$$\text{(KL)} \lim_{\mathbb{K}\longrightarrow\infty} \max_{N} \min_{\boldsymbol{A}(N)} f_{\mathbb{K},0}(\boldsymbol{A})$$

s.t.

$$\begin{array}{rcl} \lim_{\mathbb{K}\longrightarrow\infty} f_{\mathbb{K},1}(\boldsymbol{A}) & = & 0 \\ \lim_{\mathbb{K}\longrightarrow\infty} f_{\mathbb{K},2}(\boldsymbol{A}, N) & \geq & 0. \end{array}$$

Proof. See Appendix. □

An application of linear programming principle, the problem (KL) can be solved by the following equation:

$$\begin{cases} \widehat{\lambda}_0 \frac{\partial}{\partial \boldsymbol{A}} f_{\mathbb{K},0}(\boldsymbol{A}, r) + \widehat{\lambda}_1 \frac{\partial}{\partial \boldsymbol{A}} f_{\mathbb{K},1}(\boldsymbol{A}, r) + \widehat{\lambda}_2 f_{\mathbb{K},2}(\boldsymbol{A}, N) & = 0 \\ \widehat{\lambda}_0 & \geq 0 \\ \widehat{\lambda}_2 & \leq 0 \\ \widehat{\lambda}_2 f_{\mathbb{K},2}(\widehat{\boldsymbol{A}}(N), N) & = 0 \\ \widehat{\lambda}_0 \frac{\partial}{\partial N} f_{\mathbb{K},0}(\widehat{\boldsymbol{A}}(N), N) + \widehat{\lambda}_1 \frac{\partial}{\partial N} f_{\mathbb{K},1}(\widehat{\boldsymbol{A}}(N)) + \widehat{\lambda}_2 \frac{\partial}{\partial N} f_{\mathbb{K},2}(\widehat{\boldsymbol{A}}(N)) & = 0 \end{cases}$$

as $\mathbb{K}$ is sufficiently large.

3 Power Comparison by Simulation

In this section, we shall use the Monte Carlo approach to compare the powers and sizes in small IID samples of the GMLM test $\mathcal{M}_g$ and 1) the tests based empirical distribution: the Cramér-von Mises statistics W^2, the Watson statistics U^2, the Anderson-Darling statistics A^2 (see e.g. Stephens

(1974, 1978)); 2) the tests based on likelihood ratio Z_K, Z_A and Z_C proposed by Zhang (2002);

$$
\begin{aligned}
Z_K &= \max_{1\leq t\leq T}\left((t-1/2)\log\left\{\frac{t-1/2}{TF(X_{t:T})}\right\} + (T-t+1/2)\log\left\{\frac{T-t+1/2}{T(1-F(X_{t:T}))}\right\}\right)\\
Z_A &= -\sum_{t=1}^{T}\left\{\frac{\log(F(X_{t:T}))}{T-t+1/2} + \frac{\log(1-F(X_{t:T}))}{t-1/2}\right\}\\
Z_C &= \sum_{t=1}^{T}\left\{\log\left(\frac{F(X_{t:T})^{-1}-1}{(T-1/2)/(t-3/4)-1}\right)\right\}^2
\end{aligned}
$$

3) the tests based on conventional moments: the BM test and Kiefer & Salmon test (Kiefer and Salmon, 1983).

First, note that the IID version of the GMLM test $\mathcal{M}_g$ has the following form:

$$
\begin{aligned}
\mathcal{M}_g &= T\mathfrak{N}_T'\widehat{\boldsymbol{\Sigma}}^{-1}\mathfrak{N}_T, \text{ where}\\
\mathfrak{N}_{T,r} &= T^{-1}\sum_{t=1}^{T}X_{t:T}\left\{f_{\theta,1}(X_{t:T})[P^{*}_{r-1}(t/T) - P^{**}_{r-1}(t/T)] + \frac{f'_{\theta,1}(X_{t:T})}{f_{\theta,1}(X_{t:T})}P^{***}_{r-1}(t/T)\right\}\ \forall\, r=1,\ldots,N,
\end{aligned}
$$

and $\widehat{\boldsymbol{\Sigma}}$ is an estimate of the covariance matrix of the vector of L-statistics $T^{1/2}\{\mathfrak{N}_{T,r}\}_{r=1}^{N}$. Since there is always loss in the power or distortion in the size of tests in the case when the simulation sizes are quite small, thus it is of limited practical value to use the exact asymptotic covariance matrix as given in Corollary 2[3]. We shall rely on the exact bootstrap variance of the L-statistics $T^{1/2}\{\mathfrak{N}_{T,r}\}_{r=1}^{N}$ as proposed by Hutson and Ernst (2000).

The sample sizes for simulation are 15, 20, 25, and 30. The significance level or the probability of type I error for testing the goodness of fit is 0.05, at which level the critical values of the GMLM test, the BM test and the KS test are taken from the table of the χ^2 distribution. Meanwhile, the critical values at the level of 0.05 of the tests W^2, U^2, A^2 are given in Stephens (1977); and the critical values at the 0.05 level of Z_K, Z_A, and Z_C are computed with simulations of 1 million replicates (see e.g. Zhang (2002).)

[3]In the IID case, the covariance kernel $\sigma(F,G) = (F,G)^{-} - FG$.

3.1 Example 1: $H_0 : \{X_t\}_{t=1}^T \sim^{\text{IID}} N(0,1)$ versus $H_1 : \{X_t\}_{t=1}^T \sim^{\text{IID}} 0.5N(0,1) + 0.5/\tau N(0,\tau^2)$

The purpose of this experiment is to examine the performance of the tests for the null of normality when the true distribution deviates slightly from normality. The degree of divergence from Normality is controlled by τ. Simulation results are reported in Tables 1a, 1b and 1c.

We assume that the underlying distribution is the standard normal under the null hypothesis. The alternative hypothesis is a mixture of Gaussian. For the sample size of 15, the powers and sizes of the BM and GMLM tests are almost similar; both of them outperform the tests based on the empirical distribution. However, the latter has better size which is very close to theoretical size that can be attained in a large sample. For the sample size of 25 and 30, the GMLM test does outperform the BM and KS tests in term of the powers while the sizes are almost the same. In general, the GMLM test performs better than the tests based on the empirical distribution and the moments in this study.

3.2 Example 2: $H_0 : \{X_t\}_{t=1}^T \sim^{\text{IID}} N(0,1)$ versus $H_1 : \{X_t\}_{t=1}^T \sim^{\text{IID}} N(0,1) + \text{Bin}(p)\text{Poisson}(\mu)$

The purpose of this experiment is to examine how robust the tests under our consideration against outliers in the data. The number of outliers in a sample is controlled by the binomial probability p. Simulation results are reported in Tables 2a, 2b and 2c.

As before, we assume that the null hypothesis is Normality. Then we generate samples of mixed normal-Poisson random variables with sizes 15, 20, 25, 30, and 50 respectively. The probability that outliers appear in the samples is controlled by p and the magnitudes of outliers are determined by μ (the mean of Poisson distribution). For the case $\mu = 2$ and $p = 0.2$, when the simulation size is equal to 15 the highest power that the GMLM test can attain is 0.875 with 5 L-moments while the highest powers are 0.61, 0.38, and 0.50 for the BM test, the KS test, and the Z_C test respectively. When μ is set equal to 6, the powers of the GMLM test still dominates all the other tests. Note that the powers of the GMLM test, the KS test and the tests based on the empirical distribution are very sensitive to μ while the BM is not very sensitive to μ; (the reason that the KS is very sensitive to μ is that the moments of Hermite polynomials which are extremely sensitive to outliers are normalized by their asymptotic covariance matrix). This suggests that the tests based on the empirical distribution especially the GMLM test are very robust in *a small sample with outliers*. Moreover, as either the probability of jumps (p) or the sample size increases, in general the powers of the tests converges to 1. However, the speed of convergence of the GMLM test is very fast.

Appendices

Proof of Theorem 1. Let $\xi_t = F(X_t)$; and since the functions of mixing random variables of finite lags are mixing of the same size, thus ξ_t satisfies Assumption 3. In view of $t/T = F_T(X_{t:T})$ we can rewrite $\mathfrak{T}_T$ as

$$\begin{aligned}
\mathfrak{T}_T &= \sum_{t=1}^{T}\sum_{j=1}^{K} P_{T,j}(F_T(X_{t:T}))h_j(X_{t:T})\Delta F_T(X_{t:T}) \\
&\overset{(1)}{=} \sum_{t=1}^{T}\sum_{j=1}^{K} \int_{X_{t:T}}^{X_{t+1:T}} P_{T,j}(F_T(X_{t:T}))h_j(X_{t:T})\Delta F_T(X_{t:T}) \\
&\overset{(2)}{=} \sum_{j=1}^{K} \int_{-\infty}^{+\infty} P_{T,j}(F_T(x))h_j(x)dF_T(x) \\
&\overset{(3)}{=} \sum_{j=1}^{K} \int_{-\infty}^{+\infty} P_{T,j}(U_T(u))h_j(F^{-1}(u))dU_T(u)
\end{aligned}$$

where (2) follows because the range of X_t is $(-\infty, +\infty)$; (3) follows because $U_T(u) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(\xi_t \leq u) = F_T(F^{-1}(u))$. Let

$$\mathfrak{T} = \sum_{j=1}^{K} \int_{-\infty}^{+\infty} P_j(F(x))h_j(x)dF(x),$$

we have

$$
\begin{aligned}
I_T &= T^{1/2}(\mathfrak{T}_T - \mathfrak{T}) \\
&= T^{1/2}\sum_{j=1}^{K}\int_0^1 \left(P_{T,j}(U_T(u)) - P_j(u)\right) h(F^{-1}(u))dU_T(u) \\
&+ T^{1/2}\sum_{j=1}^{K}\left(\int_0^1 P_j(u)h(F^{-1}(u))dU_T(u) - \int_0^1 P_j(u)h_j(F^{-1}(u))du\right).
\end{aligned}
$$

By the mean value theorem, we have

$$
\begin{aligned}
P_{T,j}(U_T(u)) - P_j(u) &= P_{T,j}(U_T(u)) - P_{T,j}(u) + P_{T,j}(u) - P_j(u) \\
&= P^{'}_{T,j}(U^*_T(u))(U_T - u) + P_{T,j}(u) - P_j(u),
\end{aligned}
$$

where $U^*_T(u) = \theta U_T(u) + (1-\theta)u$ and $\theta \in (0,1)$. Therefore I_T can be decomposed as

$$
I_T = I_{T,1} + I_{T,2} + I_{T,3} + I_{T,4} + I_{T,5},
$$

where

$$
\begin{aligned}
I_{T,1} &= \sum_{j=1}^{K}\int_0^1 \left[P^{'}_j(u)h(F^{-1}(u))W_T(u)du + P_j(u)h(F^{-1}(u))dW_T(u)\right] \\
I_{T,2} &= \sum_{j=1}^{K}\int_0^1 h(F^{-1}(u))W_T(u)[P^{'}_{T,j}(U^*_T(u)) - P^{'}_j(u)]dU_T(u) \\
I_{T,3} &= T^{1/2}\sum_{j=1}^{K}\int_0^1 h(F^{-1}(u))W_T(u)P^{'}_j(u)dW_T(u) \\
I_{T,4} &= \sum_{j=1}^{K}\int_0^1 [P_{T,j}(u) - P_j(u)]h(F^{-1}(u))dW_T(u) \\
I_{T,5} &= T^{1/2}\sum_{j=1}^{K}\int_0^1 [P_{T,j}(u) - P_j(u)]h(F^{-1}(u))du,
\end{aligned}
$$

where $W_T(u) = T^{1/2}(U_T(u) - u)$. Assumption 4 implies that

$$\lim_{u \longrightarrow 0+} uh(F^{-1}(u)) = \lim_{u \longrightarrow 1-} (1-u)h(F^{-1}(u)) = 0,$$

after integration by parts we have

$$I_{T,1} = T^{-1/2} \sum_{t=1}^{T} \int_0^1 \sum_{j=1}^{K} [P_j(u) \mathbf{1}(\xi_t \leq u) dh_j(F^{-1}(u))].$$

In view of Assumption 3, by the central limit theorem for stationary mixing processes, we obtain the limit distribution of $I_{T,1}$ which is the main result of the theorem. Using the same argument as Mehra and Rao (1975), we can show that $I_{T,2}$, $I_{T,3}$, $I_{T,4}$, and $I_{T,5}$ converges to zero with probability 1. □

Lemma 7 (Scorgo and Revesz, 1981, Lemma 7.2.2)**.** *Let A be any event of positive probability. Then*

$$\left\{ \frac{S_{[nt]}}{\sqrt{n}};\ 0 \leq t \leq 1 \middle| A \right\} \Longrightarrow \{W(t);\ 0 \leq t \leq 1\}.$$

This means that the sequence of conditional probability measures generated by the process $\frac{S_{[nt]}}{\sqrt{n}}$ given A, converges to the Wiener measure.

Lemma 8. *If the process X_t is uniform mixing with the coefficient $\phi(h)$, then for given measurable and bounded functions $g_1(X_t)$ and $g_2(X_t)$ such that $\|g_i(X_t)\|_q < \infty$ for $i = 1, 2$ and some $q \geq 2$ we have*

$$E[g_1(X_t)g_2(X_{t+h})] \leq \|g_1(X_t)\|_2 \left\{ E|g_2(X_{t+h})| + 2\phi(h)^{1-1/q} \|g_2(X_{t+h})\|_q \right\}.$$

Proof.

$$
\begin{aligned}
E[g_1(X_t)g_2(X_{t+h})] &= E\left[E[g_1(X_t)g_2(X_{t+h})|\mathcal{F}_{-\infty}^t]\right] \\
&\overset{(1)}{\leq} \|g_1(X_t)\|_2\|E[g_2(X_{t+h})|\mathcal{F}_{-\infty}^t]\|_2 \\
&\overset{(2)}{\leq} \|g_1(X_t)\|_2\left\{\|E[g_2(X_{t+h})|\mathcal{F}_{-\infty}^t] - E[g_2(X_{t+h})]\|_2 + E[g_2(X_{t+h})]\right\} \\
&\overset{(3)}{\leq} \|g_1(X_t)\|_2\{E[g_2(X_{t+h})] + 2\phi(h)^{1-1/q}\|g_2(X_{t+h})\|_q\}
\end{aligned}
$$

for $q \geq 2$. The inequality (1) follows form the Cauchy-Schwartz inequality; the inequality (2) follows from the Minskowski inequality; and the inequality (3) follows form Lemma 6.16 in Stinchcombe and White (1998) and Theorem 14.1 in Davidson (2002) which asserts that the processes $g_i(X_t)$, $\forall\ i = 1, 2$ are also mixing of the same sizes. ☐

Proof of Theorem 7. In view of Theorem 6, the approximation of the Bahadur exact slope can be done by solving the Kullback-Leibler entropy minimization problem as given in the RHS of Equation 2.23. Let's rewrite this entropy minimization problem as

$$
\text{(KL-0)} \quad \min_{q^*} \int_{[0,1]} q^* \ln(q^*) dt
$$

s.t.

$$
\begin{aligned}
\int_{[0,1]} q^*(t)dt &= 1 \\
\int_{[0,1]} \{\mathfrak{m}_{\infty,N}(Q^*(t|Z)) - \mathfrak{m}_{\infty,N}(G_\theta(t|Z))\}\, dt &\geq 0.
\end{aligned}
$$

The problem (KL-0) can be solved by setting up a Lagrangian equation–the optimization of this equation gives rise to a nonlinear integral equation because $\mathfrak{m}_{\infty,N}(Q^*(t|Z))$ is a highly nonlinear function of $Q^*(t|Z)$. Thus, it is difficult to solve or approximate the solution of this nonlinear integral equation. We apply asymptotic expansions of the conditional distribution $Q^*(t|Z)$ and its density $q^*(t|Z)$. Since

$$I(U \leq t) = \lim_{\mathbb{K} \longrightarrow \infty} \sum_{k=0}^{\mathbb{K}} \frac{(-1)^k}{k!} \phi^{(k)}(t) H_k(U), \tag{3.1}$$

where $I(U \leq t)$ is an indicator function which is equal to 1 if $U \leq t$ and 0 otherwise; $H_k(U)$ is the k-th order Hermite polynomial. Apply conditional expectation operations on both sides of Equation 3.1, we obtain

$$\begin{aligned} Q^*(t|Z) &:= \lim_{\mathbb{K} \longrightarrow \infty} \sum_{k=0}^{\mathbb{K}} \frac{(-1)^k}{k!} \phi^{(k)}(t) \mu_k(Z) \\ q^*(t|Z) &:= \lim_{\mathbb{K} \longrightarrow \infty} \sum_{k=0}^{\mathbb{K}} \frac{(-1)^k}{k!} \phi^{(k+1)}(t) \mu_k(Z), \end{aligned}$$

where $\mu_k(Z) = E_{Q^*}[H_k(U)|Z]$. Furthermore, suppose that the conditional expectation μ_k can be estimated with a linear regression $\mu_k(Z) = \boldsymbol{a}_k^{'} Z$.

Since $Q^*(t|Z)$ can be approximated with a linear functional with the coefficient $\boldsymbol{A}$ or the function $Q^\circ(t, \boldsymbol{A})$ as defined in Theorem 7. It is clear that the problem (KL-0) can be approximated by the problem (KL). □

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* **59**(3): 817=858.

Andrews, D. W. K. (1997). A conditional kolmogorov test, *Econometrica* **65**(5): 1097–1128.

Bahadur, R. R. (1960). Stochastic comparison of tests, *Annal of Mathematical Statistics* **31**: 276–295.

Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics, *Annal of Mathematical Statistics* **38**: 303–324.

Bai, J. (2003). Testing parametric conditional distributions of dynamic models, *The Review of Economics and Statistics* **85**(3): 531–549.

Bai, J. and Chen, Z. (2006). Testing multivariate distributions, *Working paper, New York University* .

Barbour, A. D. and Chen, L. H. Y. (2005). *Stein's Method and Applications*, World Scientific, London, Singapore, Hong Kong.

Bierens, H. J. (1990). A consistent conditional moment test of functional form, *Econometrica* **58**(6): 1443–1458.

Billingsley, P. (1995). *Probability and Measure*, Wiley & Sons, New York.

Bontemps, C. and Meddahi, N. (2005). Testing normality: A gmm approach, *Journal of Econometrics* **124**: 149–186.

Bontemps, C. and Meddahi, N. (2006). Testing distributional assumptions: A GMM approach, *Working paper, Universite de Montreal* .

Borovkova, S., Burton, R. and Dehling, H. (2001). Limit theorems for functional of mixing processes with applications to u-statistics and dimension estimation, *Transactions of the American Mathematical Society* **20**(11): 4261–4318.

Bose, A. (1998). A glivenko-cantelli theorem and strong laws for l-statistics, *Journal of Theoretical Probability* **11**(4): 921–933.

Chow, Y. S. and Teicher, H. (1997). *Probability Theory*, Springer-Verlag.

Corradi, V. and Swanson, N. R. (2006). Bootstrap conditional distribution tests in the presence of dynamic misspecification, *Journal of Econometrics, forthcoming* .

Davidson, J. (2002). *Stochastic Limit Theory*, Oxford University Press.

Davydov, Y. (1996). Weak convergence of discontinuous processes to continuous ones, *Theory of Probab. and Math. Stat., Proceedings of the sem. deducated to memory of Kolmogorov, Ibragimov I. and Zaitsev A. (Eds.)*, Gordon and Breach, pp. 15–18.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*, Cambridge University Press.

Fiorentini, G., Sentana, E. and Calzolari, G. (2004). On the validity of the jarque-bera normality test in conditionally heteroskedastic dynamic regression models, *Economic Letters* **83**(3): 307–312.

Hall, P., Wolff, R. C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function, *Journal of American Statistical Association* **94**: 154–163.

Hoadley, A. B. (1967). On the probability of large deviations of functions of several empirical c.d.f.'s, *Annal of Mathematical Statistics* **38**: 360–381.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society, Series B* **52**(1): 105–124.

Hosking, J. R. M. (1992). Moments or l moments? an example comparing two measures of distributional shape, *The American Statistician* **46**(3): 186–189.

Hutson, A. D. and Ernst, M. D. (2000). The exact bootrap mean and variance of an l-estimator, *Journal of Royal Statistical Society B* **62**: 89–94.

Jlrech, T. J., Velis, D. R., Woodbury, A. D. and Sacchi, M. D. (2000). L-moments and c-moments, *Stochastic Environmental Research and Risk Assessment* **14**: 50–68.

Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-tests, *Theory of Probability and its Applications* **XXVI**: 240–257.

Kiefer, N. M. and Salmon, M. (1983). Testing normality in econometric models, *Economics Letters* **11**: 123–127.

Komunjer, I. and Vuong, Q. (2006). Efficient conditional quantile estimation: The time series case, *Working paper* .

Lehmann, E. L. (2000). *Testing Statistical Hypotheses*, Springer-Verlag.

Li, F. and Tkacz, G. (2006). A consistent bootstrap test for conditional density functions with time-series data, *Journal of Econometrics, forthcoming* .

Lin, C. (2006). A robust test for parameter homogeneity in panel data models, *Working paper, University of Michigan* .

Lin, Z. and Lu, C. (1989). *Limit Theory for Mixing Dependent Random Variables*, Science Press, New York, Beijing.

Mason, D. M. and Shorack, G. R. (1992). Necessary and sufficient conditions for asymptotic normality of l-statistics, *The Annals of Probability* **20**(4): 1779–1804.

Mehra, K. L. and Rao, S. (1975). On functions of order statistics for mixing processes, *The Annals of Statistics* **3**(4): 874–883.

Perez, N. H. (2005). A theoretical comparison between moments and l-moments, *Working paper, University of Toronto* .

Rosenbaltt, M. (1952). Remarks on a multivariate transformation, *Annals of Mathematical Statistics* **23**: 470–472.

Ruymgaart, F. H. and Zuijlen, M. C. A. V. (1977). Asymptotic normality of linear combinations of functions of order statistics in the non-iid case, *Indag. Math.* **80**: 432–447.

Sankarasubramanian, A. and Srinivasan, K. (1999). Investigation and comparison of sampling properties of l-moments and conventional moments, *Journal of Hydrology* **218**: 13–34.

Scorgo, M. and Revesz, P. (1981). *Strong Approximations in Probability and Statistics*, Academic Press, New York, San Francisco, London.

Sen, P. K. (1978). An invariance principle for linear combinations of order statistics, *Z. Wahrscheinlichkeitstheorie Verw. Geibiete* **42**: 327–340.

Serfling, R. J. (1984). Generalized l-, m-, and r-statistics, *The Annals of Statistics* **12**(1): 76–86.

Shorack, G. R. (1972). Functions of order statistics, *The Annals of Mathematical Statistics* **43**(2): 412–427.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *Proc. Sixth Berkeley Symp. Math. Statist. Probab. 2* pp. 583–602.

Stein, C. (1986). *Approximate computation of expectations*, Institute of Mathematical Statistics, Hayward, California.

Stephens, M. A. (1974). Edf statistics for goodness-of-fit and some comparisons, *Journal of American Statistical Association* **69**: 730–737.

Stephens, M. A. (1977). Goodness of fit the extreme value distribution, *Biometrika* **64**: 583–588.

Stephens, M. A. (1978). Goodness of fit with special reference to tests for exponentiality, *Technical Report, Stanford University* .

Stinchcombe, M. B. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative, *Econometric Theory* **14**: 295–325.

White, H. (2001). *Asymptotic Theory for Econometricians*, 2 edn, Academic Press, New York, London, Tokyo.

Wu, W. B. (2005). Empirical processes of dependent random variables, *Technical report, University of Chicago* .

Yoshihara, K. (1976). Limiting behaviour of u-statistics for stationary, absolutely regular processes, *Z. Wahrscheinlichkeitstheorie Verw. Geibiete* **35**: 237–252.

Zhang, J. (2002). Powerful goodness-of-fit tests based on likelihood ratio, *Journal of Royal Statistical Society B* **64**: 281–294.

Zheng, J. X. (2000). A consistent test of conditional parametric distributions, *Econometric Theory* **16**: 667–691.

The GMLM test					
Sample size is 15					
Number of moments	τ				
	0.2	*0.4*	*0.6*	*0.8*	*1*
2	0.74	0.735	0.71	0.7	0.42
3	0.825	0.82	0.84	0.76	0.52
4	0.79	0.795	0.8	0.79	0.52
5	0.855	0.855	0.83	0.82	0.58
Sample size is 20					
2	0.82	0.805	0.79	0.72	0.49
3	0.89	0.895	0.88	0.84	0.6
4	0.89	0.88	0.86	0.85	0.66
5	0.95	0.93	0.91	0.9	0.69
Sample size is 25					
2	0.84	0.85	0.79	0.75	0.47
3	0.915	0.91	0.9	0.82	0.6
4	0.918	0.916	0.91	0.86	0.69
5	0.924	0.92	0.915	0.88	0.61
Sample size is 30					
2	0.855	0.84	0.84	0.8	0.31
3	0.93	0.92	0.92	0.89	0.35
4	0.95	0.92	0.9	0.88	0.3
5					0.4

Table 1a: The power of the GMLM test under the null hypothesis of Normality against the alternative hypothesis of the mixture of Normalities: $0.5N(0,1)+0.5/\tau N(0,\tau)$

The tests based on empirical distribution					
Sample size is 15					
Statistics	τ				
	0.2	0.4	0.6	0.8	1
Cramer-von Mises (W)	0.334	0.284	0.232	0.22	0.04
Watson (U)	0.398	0.32	0.262	0.236	0.036
Anderson-Darling (A)	0.428	0.35	0.298	0.282	0.046
Z_K	0.342	0.29	0.244	0.246	0.006
Z_A	0.112	0.084	0.07	0.072	0
Z_C^A	0.342	0.284	0.254	0.242	0.008
Sample size is 25					
Cramer-von Mises (W)	0.406	0.334	0.282	0.26	0.05
Watson (U)	0.462	0.376	0.308	0.266	0.048
Anderson-Darling (A)	0.494	0.414	0.356	0.334	0.048
Z_K	0.426	0.362	0.336	0.296	0.012
Z_A	0.116	0.09	0.074	0.082	0
Z_C^A	0.408	0.364	0.326	0.29	0.006
Sample size is 50					
Cramer-von Mises (W)	0.536	0.418	0.368	0.328	0.046
Watson (U)	0.592	0.488	0.394	0.38	0.058
Anderson-Darling (A)	0.608	0.502	0.45	0.416	0.054
Z_K	0.598	0.532	0.482	0.418	0.014
Z_A	0.132	0.104	0.078	0.08	0
Z_C^A	0.604	0.546	0.448	0.382	0.01

Table 1b: The powers of the tests based on the empirical distribution under the null hypothesis of Normality against the alternative hypothesis of the mixture of Normalities: $0.5N(0,1)+0.5/\tau N(0,\tau)$

Number of moments	The tests based on moments									
	Sample size is 15									
	τ									
	0.2		0.4		0.6		0.8		1	
	BM	KS	BM	KS	BM	KS	BM	KS	BM	KS
2	0.56	0.36	0.53	0.37	0.56	0.34	0.45	0.31	0.13	0.06
3	0.62	0.38	0.64	0.38	0.58	0.34	0.5	0.32	0.29	0.06
4	0.83	0.37	0.84	0.37	0.85	0.33	0.79	0.31	0.67	0.05
5	0.85	0.37	0.87	0.38	0.85	0.35	0.84	0.32	0.82	0.06
15	0.27	0.37	0.37	0.38	0.34	0.34	0.39	0.32	0.36	0.05
	Sample size is 25									
2	0.57	0.4	0.55	0.42	0.53	0.39	0.5	0.36	0.09	0.07
3	0.69	0.41	0.7	0.42	0.65	0.4	0.53	0.36	0.21	0.05
4	0.83	0.41	0.88	0.42	0.85	0.4	0.82	0.36	0.53	0.05
5	0.87	0.4	0.91	0.42	0.9	0.4	0.86	0.36	0.7	0.06
15	0.61	0.38	0.76	0.41	0.64	0.38	0.75	0.36	0.78	0.03
	Sample size is 50									
2	0.73	0.47	0.64	0.48	0.64	0.45	0.54	0.44	0.04	0.05
3	0.84	0.46	0.81	0.47	0.77	0.45	0.63	0.43	0.09	0.06
4	0.87	0.48	0.92	0.49	0.89	0.46	0.87	0.45	0.3	0.04
5	0.87	0.46	0.91	0.47	0.9	0.45	0.88	0.43	0.43	0.04
15	0.87	0.46	0.87	0.47	0.93	0.41	0.93	0.38	1	0.02

Table 1c: The powers of the Bontemps & Meddahi test and the Kiefer & Salmon test under the null hypothesis of Normality against the alternative hypothesis of the mixture of Normalities: $0.5N(0,1)+0.5/\tau N(0,\tau)$

The GMLM test										
	Sample size is 15									
Number of L moments	p = 0.2		p=0.4		p = 0.6		p = 0.8		p = 0.95	
	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6
2	0.775	0.95	0.94	0.995	0.96	1	0.97	1	1	1
3	0.815	0.965	0.94	1	0.99	1	0.97	1	1	1
4	0.865	0.97	0.97	0.995	0.97	1	0.98	1	1	1
5	0.875	0.98	0.975	1	0.97	1	0.98	1	1	1
	Sample size is 20									
2	0.84	0.985	0.975	1	1	1	1	1	1	1
3	0.87	0.99	0.985	1	1	1	1	1	1	1
4	0.91	0.995	0.99	1	1	1	1	1	1	1
5	0.915	0.99	0.99	1	1	1	1	1	1	1
	Sample size is 25									
2	0.9	1	0.981	1	1	1	1	1	1	1
3	0.915	1	1	1	1	1	1	1	1	1
4	0.945	1	1	1	1	1	1	1	1	1
5	0.96	1	1	1	1	1	1	1	1	1
	Sample size is 30									
2	0.96	0.98	0.991	1	1	1	1	1	1	1
3	0.97	0.99	1	1	1	1	1	1	1	1
4	0.98	1	1	1	1	1	1	1	1	1
5	0.99	1	1	1	1	1	1	1	1	1

Table 2a: The power of the GMLM test under the null hypothesis of Normality against the alternative hypothesis of the mixture of Normality and Poisson jumps: $N(0,1) + Bin(p)Poi(\mu)$

The tests based on the empirical distribution										
	Sample size is 15									
Statistics	p									
	0.2		0.4		0.6		0.8		0.95	
	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6
Cramer-von Mises (W)	0.188	0.272	0.454	0.736	0.766	0.962	0.95	1	0.99	1
Watson (U)	0.114	0.202	0.32	0.672	0.638	0.95	0.85	0.998	0.972	1
Anderson-Darling (A)	0.378	0.78	0.748	0.994	0.946	1	0.992	1	1	1
Z_K	0.324	0.836	0.676	0.996	0.89	1	0.976	1	0.996	1
Z_A	0	0	0	0	0	0	0	0	0	0
Z_C	0.502	0.888	0.794	0.998	0.956	1	0.988	1	0.996	1
	Sample size is 25									
Cramer-von Mises (W)	0.268	0.408	0.7	0.914	0.944	1	0.994	1	1	1
Watson (U)	0.162	0.294	0.502	0.866	0.846	0.996	0.964	1	0.998	1
Anderson-Darling (A)	0.546	0.924	0.91	1	0.99	1	0.998	1	1	1
Z_K	0.534	0.968	0.91	1	0.986	1	0.996	1	0.998	1
Z_A	0	0	0	0	0	0	0	0	0	0
Z_C	0.69	0.982	0.95	1	0.992	1	0.998	1	1	1
	Sample size is 50									
Cramer-von Mises (W)	0.422	0.694	0.936	1	1	1	1	1	1	1
Watson (U)	0.244	0.548	0.8	0.996	0.986	1	0.998	1	1	1
Anderson-Darling (A)	0.75	0.996	0.994	1	1	1	1	1	1	1
Z_K	0.828	1	0.992	1	1	1	1	1	1	1
Z_A	0	0	0	0	0	0	0	0	0	0
Z_C	0.898	1	1	1	1	1	1	1	1	1

Table 2b: The powers of the tests based on the empirical distribution under the null hypothesis of Normality against the alternative hypothesis of the mixture of Normality and Poisson jumps: $N(0,1)+Bin(p)Poi(\mu)$

Number of C moments	The tests based on moments																			
	Sample size is 15																			
	p = 0.2				p = 0.4				p = 0.6				p = 0.8				p = 0.95			
	BM		KS		BM		KS		BM		KS		BM		KS		BM		KS	
	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6	μ=2	μ=6
2	0.17	0.18	0.58	0.88	0.3	0.46	0.85	0.99	0.57	0.87	0.95	1	0.84	0.97	0.96	1	0.92	1	0.98	1
3	0.25	0.24	0.57	0.89	0.27	0.38	0.85	0.99	0.51	0.74	0.95	1	0.72	0.96	0.96	1	0.9	1	0.98	1
4	0.4	0.36	0.58	0.88	0.44	0.34	0.84	0.99	0.52	0.66	0.95	1	0.67	0.86	0.96	1	0.83	1	0.98	1
5	0.61	0.51	0.57	0.88	0.62	0.39	0.83	0.99	0.69	0.54	0.96	1	0.78	0.83	0.96	1	0.88	1	0.98	1
15	0.35	0.29	0.57	0.88	0.41	0.4	0.8	0.99	0.47	0.41	0.96	1	0.46	0.46	0.96	1	0.5	0.43	0.98	1
	Sample size is 25																			
2	0.16	0.27	0.76	1	0.51	0.85	0.97	1	0.86	1	1	1	0.98	1	1	1	1	1	1	1
3	0.14	0.26	0.74	1	0.4	0.78	0.97	1	0.79	1	1	1	0.99	1	1	1	1	1	1	1
4	0.3	0.26	0.74	1	0.49	0.6	0.97	1	0.79	0.98	1	1	0.93	1	1	1	0.99	1	1	1
5	0.53	0.41	0.74	1	0.6	0.62	0.97	1	0.71	0.94	1	1	0.89	1	1	1	0.99	1	1	1
15	0.84	0.8	0.74	1	0.9	0.68	0.95	1	0.91	0.8	1	1	0.91	0.75	1	1	0.83	0.86	1	1
	Sample size is 50																			
2	0.35	0.78	0.96	1	0.92	1	1	1	0.99	1	1	1	1	1	1	1	1	1	1	1
3	0.26	0.67	0.96	1	0.89	1	1	1	0.99	1	1	1	1	1	1	1	1	1	1	1
4	0.34	0.54	0.95	1	0.86	0.99	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	0.43	0.56	0.95	1	0.81	1	1	1	0.97	1	1	1	1	1	1	1	1	1	1	1
15	1	0.92	0.95	1	1	0.98	1	1	1	0.98	1	1	1	0.97	1	1	1	0.98	1	1

Table 2c: The powers of the Bontemps & Meddahi test and the Kiefer & Salmon test under the null hypothesis of Normality against the alternative hypothesis of the mixture of Normality and Poisson jumps: $N(0,1) + Bin(p)Poi(\mu)$