

# A Loss Function Approach to Model Specification Testing and Its Relative Efficiency<sup>1</sup>

Yongmiao Hong

*Cornell University & Xiamen University*

Yoon-Jin Lee

*Indiana University*

The generalized likelihood ratio (GLR) test proposed by Fan, Zhang and Zhang (2001) and Fan and Yao (2003) is a generally applicable nonparametric inference procedure. In this paper, we show that although it inherits many advantages of the parametric maximum likelihood ratio (LR) test, the GLR test does not have the optimal power property. We propose a generally applicable test based on loss functions, which measure discrepancies between the null and nonparametric alternative models and are more relevant to decision-making under uncertainty. The new test is asymptotically more powerful than the GLR test in terms of Pitman's efficiency criterion. This efficiency gain holds no matter what smoothing parameter and kernel function are used and even when the true likelihood function is available for the GLR test.

**1. Introduction.** The likelihood ratio (LR) principle is a generally applicable approach to parametric hypothesis testing (e.g., Quang (1989)). The maximum LR test compares the best explanation of data under the alternative with the best explanation under the null hypothesis. It is well known from the Neyman-Pearson Lemma that the maximum LR test has asymptotically optimal power. Moreover, the LR statistic follows an asymptotic null  $\chi^2$  distribution with a known number of degrees of freedom, enjoying the so-called Wilks phenomena that its asymptotic distribution is free of nuisance parameters.

In parametric hypothesis testing, however, it is implicitly assumed that the family of alternative likelihood models contains the true model. When this is not the case, one may fail to reject the null hypothesis erroneously. In many testing problems in practice, while the null hypothesis is well formulated, the alternative is vague. Over the last two decades or so, there has been a growing interest in nonparametric inference, namely, inference for hypotheses on parametric, semiparametric, and nonparametric models against a nonparametric alternative. The nonparametric alternative is very useful when there is no prior information about the true model. Because the nonparametric alternative contains the true model at least for large samples, it ensures the consistency of a test. Nevertheless, there have been few generally applicable nonparametric inference principles. One naive extension would be to develop a nonparametric maximum LR test similar to the parametric maximum LR test. However, the nonparametric maximum likelihood estimator (MLE) usually does not exist, due to the well-known infinite dimensional parameter problem (Bahadur 1958, Le Cam 1979, 1990).

---

<sup>1</sup>Correspondence to: Yongmiao Hong, Department of Economics and Department of Statistical Science, Uris Hall, Cornell University, Ithaca, NY 14850, U.S.A. and Wang Yanan Institute for Studies in Economics, and MOE Key Laboratory of Economics, Xiamen University, Xiamen 361005, Fujian, China. Email: yh20@cornell.edu; and Yoon-Jin Lee, Department of Economics, Indiana University, Wylie Hall, Bloomington, IN 47405, U.S.A. Email: lee243@indiana.edu.

Key Words: Efficiency, Generalized likelihood ratio test, Loss function, Local alternative, Kernel, Pitman efficiency, Smoothing parameter.

Even if it exists, it may be difficult to compute, and the resulting nonparametric maximum LR test is not asymptotically optimal. This is because the nonparametric MLE chooses the smoothing parameter automatically, which limits the choice of the smoothing parameter and renders it impossible for the test to be optimal.

Fan, Zhang and Zhang (2001) and Fan and Yao (2003) proposed a generalized likelihood ratio (GLR) test by replacing the nonparametric MLE with a reasonable nonparametric estimator, attenuating the difficulty of the nonparametric maximum LR test and enhancing the flexibility of the test by allowing for a range of smoothing parameters. The GLR test maintains the intuitive feature of the parametric LR test because it is based on the likelihoods of generating the observed sample under the null and alternative hypotheses. It is generally applicable to various hypotheses involving a parametric, semiparametric or nonparametric null model against a nonparametric alternative. By a proper choice of the smoothing parameter, the GLR test can achieve the asymptotically optimal rate of convergence in the sense of Ingster (1993a, 1993b, 1993c) and Lepski and Spokoiny (1999). Moreover, it enjoys the appealing Wilks phenomena that its asymptotic null distribution is free of nuisance parameters and nuisance functions.

The GLR test is a nonparametric inference procedure based on the empirical Kullback-Leibler Information Criterion (KLIC) between the null model and a nonparametric alternative model. This measure can capture any discrepancy between the null and alternative models, ensuring the consistency of the GLR test. As Fan *et al.* (2001) and Fan and Jiang (2007) point out, it holds an advantage over many discrepancy measures such as the  $L_2$  and  $L_\infty$  measures commonly used in the literature because for the latter the choices of measures and weight functions are often arbitrary, and the null distributions of the test statistics are unknown and generally depend on nuisance parameters. We note that Robinson (1991) developed a nonparametric KLIC test for serial independence and White (1982, p.17) also suggested a nonparametric KLIC test for parametric likelihood models.

The GLR test assumes that stochastic errors follows some parametric distribution which need not contain the true distribution. It is essentially a nonparametric pseudo LR test. Azzalini, Bowman and Härdle (1989), Azzalini and Bowman (1990) and Cai, Fan and Yao (2000) also proposed a nonparametric pseudo-LR test for the validity of parametric regression models.

In this paper, we show that despite its general nature and appealing features, the GLR test does not have the optimal power property of the classical LR test. We first propose a generally applicable nonparametric inference procedure based on loss functions and show that it is asymptotically more powerful than the GLR test in terms of Pitman's efficiency criterion. Loss functions are often used in estimation, model selection, and prediction (e.g., Zellner 1985, Phillips 1996, Weiss 1996, Christoffersen and Diebold 1997, Giacomini and White 2006), but not in testing. A loss function compares the models under the null and alternative hypotheses by specifying a penalty for the discrepancy between the two models. The use of a loss function is often more relevant to decision-making under uncertainty because one can choose a loss function to mimic the objective of the decision maker. In inflation forecasting, for example, central banks may have asymmetric preferences which affect their optimal policies (Peel and Nobay 1998). They may be more concerned with underprediction than

overprediction of inflation rates. In financial risk management, regulators may be more concerned with the left-tailed distribution of portfolio returns than the rest of the distribution. In these circumstances, it is more appropriate to choose an asymmetric loss function to validate an inflation rate model and an asset return distribution model. The admissible class of loss functions for our approach is large, including quadratic, truncated quadratic and asymmetric linex loss functions (Varian 1975, Zellner 1986). They do not require any knowledge of the true likelihood, do not involve any choice of weights, and enjoy the Wilks phenomena that its asymptotic distribution is free of nuisance parameters and nuisance functions. Most importantly, the loss function test is asymptotically more powerful than the GLR test in terms of Pitman’s efficiency criterion, regardless of the choice of the smoothing parameter and the kernel function. This efficiency gain holds even when the true likelihood function is available for the GLR test. Interestingly, all admissible loss functions are asymptotically equally efficient under a general class of local alternatives.

The paper is planned as follows. Section 2 introduces the framework and the GLR principle. Section 3 proposes a class of loss function-based tests. For concreteness, we focus on specification testing for time series regression models, although our approach is applicable to other nonparametric testing problems. Section 4 derives the asymptotic distributions of the loss function test and the GLR test. Section 5 compares their relative efficiency under a class of local alternatives. In Section 6, a simulation study compares the performance between two competing tests in finite samples. Section 7 concludes the paper. All mathematical proofs are collected in an appendix and Supplementary Material.

**2. Generalized Likelihood Ratio Test.** Maximum LR tests are a generally applicable and powerful inference method for most parametric testing problems. However, the classical LR principle implicitly assumes that the alternative model contains the true data generating process (DGP). This is not always the case in practice. To ensure that the alternative model contains the true DGP, one can use a nonparametric alternative model.

Recognizing the fact that the nonparametric MLE may not exist and so cannot be a generally applicable method, Fan *et al.* (2001) and Fan and Yao (2003) proposed the GLR principle as a generally applicable method for nonparametric inference. The idea is to compare a suitable nonparametric estimator with a restricted estimator under the null hypothesis via a LR statistic. Specifically, suppose one is interested in whether a parametric likelihood model  $f_\theta$  is correctly specified for the unknown density  $f$  of the DGP, where  $\theta$  is a finite-dimensional parameter. The null hypothesis of interest is

$$\mathbb{H}_0 : f = f_{\theta_0} \text{ for some } \theta_0 \in \Theta \tag{2.1}$$

where  $\Theta$  is a parameter space. The alternative hypothesis is

$$\mathbb{H}_A : f \neq f_\theta \text{ for all } \theta \in \Theta. \tag{2.2}$$

In testing  $\mathbb{H}_0$  versus  $\mathbb{H}_A$ , a nonparametric model for  $f$  can be used as an alternative, as also suggested in White (1982, p.17). Suppose the log-likelihood function of a random sample is  $\hat{l}(f, \eta)$ ,

where  $\eta$  is a nuisance parameter. Under  $\mathbb{H}_0$ , one can obtain the MLE  $(\hat{\theta}_0, \hat{\eta}_0)$  by maximizing the model likelihood  $\hat{l}(f_\theta, \eta)$ . Under the alternative  $\mathbb{H}_A$ , given  $\eta$ , one can obtain a reasonable smoothed nonparametric estimator  $\hat{f}_\eta$  of  $f$ . The nuisance parameter  $\eta$  can then be estimated by the profile likelihood; that is, to find  $\eta$  to maximize  $l(\hat{f}_\eta, \eta)$ . This gives the maximum profile likelihood  $l(\hat{f}_{\hat{\eta}}, \hat{\eta})$ . The GLR test statistic is then defined as

$$\lambda_n = l(\hat{f}_{\hat{\eta}}, \hat{\eta}) - l(f_{\hat{\theta}_0}, \hat{\eta}_0). \quad (2.3)$$

This is the difference of the log-likelihoods of generating the observed sample under the alternative and null models. A large value of  $\lambda_n$  is evidence against  $\mathbb{H}_0$  since the alternative family of nonparametric models is far more likely to generate the observed data.

The GLR test does not require knowing the true likelihood. This is appealing since nonparametric testing problems do not assume that the underlying distribution is known. For example, in a regression setting one usually does not know the error distribution. Here, one can estimate model parameters by using a quasi-likelihood function  $q(f_\theta, \eta)$ . The resulting GLR test statistic is then defined as

$$\lambda_n = q(\hat{f}_{\hat{\eta}}, \hat{\eta}) - q(f_{\hat{\theta}_0}, \hat{\eta}_0). \quad (2.4)$$

The GLR approach is also applicable to the cases with unknown nuisance functions. This can arise (e.g.) when one is interested in testing whether a function has an additive form which itself is still nonparametric. In this case, one can replace  $f_{\hat{\theta}_0}$  by a nonparametric estimator under the null hypothesis of additivity. Robinson (1991) considers such a case in testing serial independence.

As a generally applicable nonparametric inference procedure, the GLR principle has been used to test a variety of models, including univariate regression models (Fan *et al.* 2001), functional coefficient regression models (Cai *et al.* 2000), spectral density models (Fan and Zhang 2004), varying-coefficient partly linear regression models (Fan and Huang 2005), additive models (Fan and Jiang 2005), diffusion models (Fan and Zhang 2003), and partly linear additive models (Fan and Yao (2003)). Analogous to the classical LR test statistic which follows an asymptotic null  $\chi^2$  distribution with a known number of degrees of freedom, the asymptotic distribution of the GLR statistic  $\lambda_n$  is also a  $\chi^2$  with a known large number of degrees of freedom, in the sense that

$$r\lambda_n \simeq \chi_{\mu_n}^2$$

as a sequence of constants  $\mu_n \rightarrow \infty$  and some constant  $r > 0$ ; namely,

$$\frac{r\lambda_n - \mu_n}{\sqrt{2\mu_n}} \xrightarrow{d} N(0, 1),$$

where  $\mu_n$  and  $r$  are free of nuisance parameters and nuisance functions, although they may depend on the methods of nonparametric estimation and smoothing parameters. Therefore, the asymptotic distribution of  $\lambda_n$  is free of nuisance parameters and nuisance functions. One can use  $\lambda_n$  to make

inference based on the known distribution of  $N(\mu_n, 2\mu_n)$  or  $\chi_{\mu_n}^2$  in large samples. Alternatively, one can simulate the null distribution of  $\lambda_n$  by setting nuisance parameters at any reasonable values, such as the MLE  $\hat{\eta}_0$  or the maximum profile likelihood estimator  $\hat{\eta}$  in (2.3).

The GLR test is powerful under a class of contiguous local alternatives:

$$\mathbb{H}_{an} : f = f_{\theta_0} + n^{-\gamma} g_n,$$

where  $\gamma > 0$  is a constant and  $g_n$  is an unspecified sequence of smooth functions in a large class of function space. It has been shown (Fan *et al.* 2001) that when a local linear smoother is used to estimate  $f$  and the bandwidth is of order  $n^{-2/9}$ , the GLR test can detect local alternatives with rate  $\gamma = 4/9$ , which is optimal according to Ingster (1993a, 1993b, 1993c).

**3. A Loss Function Approach.** In this paper, we will show that while the GLR test enjoys many appealing features of the classical LR test, it does not have the optimal power property of the classical LR test. We will propose a class of loss function-based tests and show that they are asymptotically more powerful than the GLR test under a class of local alternatives. Loss functions measure discrepancies between the null and alternative models and are more relevant to decision making under uncertainty, because the loss function can be chosen to mimic the objective function of the decision maker. The admissible loss functions include but are not restricted to quadratic, truncated quadratic, and asymmetric linex loss functions. Like the GLR test, our tests are generally applicable to various nonparametric inference problems, do not involve choosing any weight function, and their null asymptotic distributions do not depend on nuisance parameters and nuisance functions.

For concreteness, we focus on specification testing for time series regression models. Regression modelling is one of the most important statistical problems, and has been exhaustively studied, particularly in the *i.i.d.* contexts (e.g., Härdle and Mammen 1993). Focusing on testing regression models will provide deep insight into our approach and allow us to provide primitive regularity conditions for formal results. Extension to time series contexts also allows us to expand the scope of applicability of our tests and the GLR test. We emphasize that our approach is applicable to many other nonparametric test problems, such as testing parametric density models.

Suppose  $\{X_t, Y_t\} \in \mathbb{R}^{p+1}$  is a stationary time series with finite second moments, where  $Y_t$  is a scalar,  $p \in \mathbb{N}$  is the dimension of vector  $X_t$ , and  $X_t$  may contain exogenous and/or lagged dependent variables. Then we can write

$$Y_t = g_0(X_t) + \varepsilon_t, \tag{3.1}$$

where  $g_0(X_t) = E(Y_t|X_t)$  and  $E(\varepsilon_t|X_t) = 0$ . The fact that  $E(\varepsilon_t|X_t) = 0$  does not imply that  $\{\varepsilon_t\}$  is a martingale difference sequence. In a time series context,  $\varepsilon_t$  is often assumed to be *i.i.d.*( $0, \sigma^2$ ) and independent of  $X_t$  (e.g., Gao and Gijbels 2008). This implies  $E(\varepsilon_t|X_t) = 0$  but not vice versa, and so it is overly restrictive from a practical point of view. For example,  $\varepsilon_t$  may display volatility clustering (e.g., Engle 1982):

$$\varepsilon_t = z_t \sqrt{h_t},$$

where  $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ . Here, we have  $E(\varepsilon_t|X_t) = 0$  but  $\{\varepsilon_t\}$  is not *i.i.d.* We will allow such an important feature, which is an empirical stylized fact for high-frequency financial time series.

In practice, a parametric model is often used to approximate the unknown function  $g_0(X_t)$ . We are interested in testing validity of a parametric model  $g(X_t, \theta)$ , where  $g(\cdot, \cdot)$  has a known functional form, and  $\theta \in \Theta$  is an unknown finite dimensional parameter. The null hypothesis is

$$\mathbb{H}_0 : \Pr[g_0(X_t) = g(X_t, \theta_0)] = 1 \text{ for some } \theta_0 \in \Theta$$

versus the alternative hypothesis

$$\mathbb{H}_A : \Pr[g_0(X_t) \neq g(X_t, \theta)] < 1 \text{ for all } \theta \in \Theta.$$

An important example is a linear time series model

$$g(X_t, \theta) = X_t' \theta.$$

This is called linearity testing in the time series literature (Granger and Teräsvirta 1993). Under  $\mathbb{H}_A$ , there exists neglected nonlinearity in conditional mean. For discussion on testing linearity in a time series context, see Granger and Teräsvirta (1993), Lee, Granger and White (1993), Hansen (1996), Hjellvik and Tjøstheim (1996), and Hong and Lee (2005).

Because there are many possibilities for departures from a specific functional form, and practitioners usually have no information about the true alternative, it is desirable to construct a test of  $\mathbb{H}_0$  against a nonparametric alternative, which contains the true function  $g_0(\cdot)$  and thus ensures the consistency of the test against  $\mathbb{H}_A$ . For this reason, the GLR test is attractive.

Suppose we have a random sample  $\{Y_t, X_t\}_{t=1}^n$  of size  $n \in \mathbb{N}$ . Assuming that the error  $\varepsilon_t$  is *i.i.d.*  $N(0, \sigma^2)$ , we obtain the conditional quasi-log-likelihood function of  $Y_t$  given  $X_t$  as follows:

$$\hat{l}(g, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n [Y_t - g(X_t)]^2. \quad (3.2)$$

Let  $\hat{g}(x)$  be a consistent local smoother for  $g_0(x)$ . Examples of  $\hat{g}(x)$  include the Nadaraya-Watson estimator (Härdle 1990, Li and Racine 2007, Pagan and Ullah 1999) and local linear estimator (Fan and Yao 2003). Substituting  $\hat{g}(X_t)$  into (3.2), one obtains the likelihood of generating the observed sample  $\{Y_t, X_t\}_{t=1}^n$  under  $\mathbb{H}_A$  :

$$\hat{l}(\hat{g}, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} SSR_1, \quad (3.3)$$

where  $SSR_1$  is the sum of squared residuals of the nonparametric model; namely,

$$SSR_1 = \sum_{t=1}^n [Y_t - \hat{g}(X_t)]^2.$$

Maximizing the likelihood in (3.3) with respect to nuisance parameter  $\sigma^2$  yields  $\hat{\sigma}^2 = n^{-1}SSR_1$ . Substituting this estimator in (3.3) yields the following likelihood

$$\hat{l}(\hat{g}, \hat{\sigma}^2) = -\frac{n}{2} \ln(SSR_1) - \frac{n}{2} [1 + \ln(2\pi/n)]. \quad (3.4)$$

Using a similar argument and maximizing the model quasi-likelihood function with respect to  $\theta$  and  $\sigma^2$  simultaneously, we can obtain the parametric maximum quasi-likelihood under  $\mathbb{H}_0$ :

$$\hat{l}(\hat{g}_{\hat{\theta}_0}, \hat{\sigma}_0^2) = -\frac{n}{2} \ln SSR_0 - \frac{n}{2} \ln[1 + \ln(2\pi/n)], \quad (3.5)$$

where  $(\hat{\theta}_0, \hat{\sigma}_0^2)$  are the MLE under  $\mathbb{H}_0$ , and  $SSR_0$  is the sum of squared residuals of the parametric regression model, namely,

$$SSR_0 = \sum_{t=1}^n [Y_t - g_0(X_t, \hat{\theta}_0)]^2.$$

Given the *i.i.d.*  $N(0, \sigma^2)$  assumption for  $\varepsilon_t$ ,  $\hat{\theta}_0$  is the least squares estimator that minimizes  $SSR_0$ .

Thus, the GLR statistic is defined as

$$\lambda_n = \hat{l}(\hat{g}, \hat{\sigma}^2) - \hat{l}(\hat{g}_{\hat{\theta}_0}, \hat{\sigma}_0^2) = \frac{n}{2} \ln(SSR_0/SSR_1). \quad (3.6)$$

Under the *i.i.d.*  $N(0, \sigma^2)$  assumption for  $\varepsilon_t$ ,  $\lambda_n$  is asymptotically equivalent to the  $F$  test statistic

$$F = \frac{SSR_0 - SSR_1}{SSR_1}. \quad (3.7)$$

The latter has been proposed by Azzalini *et al.* (1989), Azzalini *et al.* (1993), Hong and White (1995), and Fan and Li (2002) in *i.i.d.* contexts. The asymptotic equivalence between the GLR and  $F$  tests can be seen from a Taylor series expansion of  $\lambda_n$ :

$$\lambda_n = \frac{n}{2} \cdot F + \text{Remainder}.$$

We now propose an alternative approach to testing  $\mathbb{H}_0$  versus  $\mathbb{H}_A$  by comparing the null and alternative models via a loss function  $D : \mathbb{R}^2 \rightarrow \mathbb{R}$ , which measures the discrepancy between the fitted values  $\hat{g}(X_t)$  and  $g(X_t, \hat{\theta}_0)$ :

$$Q_n = \sum_{t=1}^n D \left[ \hat{g}(X_t), g_0(X_t, \hat{\theta}_0) \right]. \quad (3.8)$$

Intuitively, the loss function gives a penalty whenever the parametric model overestimates or underestimates the true model. The latter is consistently estimated by a nonparametric method.

A specific class of loss functions  $D(\cdot, \cdot)$  is given by  $D(u, v) = d(u - v)$ , where  $d(z)$  has a unique minimum at 0, and is monotonically nondecreasing as  $|z|$  increases. Suppose  $d(\cdot)$  is twice continuously differentiable at 0 with  $d(0) = 0$ ,  $d'(0) = 0$  and  $0 < d''(0) < \infty$ . The condition of  $d'(0) = 0$  implies that the first order term in the Taylor expansion of  $d(\cdot)$  around 0 vanishes to 0 identically. This class of

loss functions  $d(\cdot)$  has been called a generalized cost-of-error function in the literature (e.g., Pesaran and Skouras 2000, Granger 1999, Christoffersen and Diebold 1996, 1997, Granger and Pesaran 2000, Weiss 1996). The loss function is closely related to decision-based evaluation, which assesses the economic value of forecasts to a particular decision maker or group of decision makers. For example, in risk management the extreme values of portfolio returns are of particular interest to regulators, while in macroeconomic management the values of inflation or output growth, in the middle of the distribution, may be of concern to central banks. A suitable choice of loss function can mimic the objective of the decision maker.

Infinitely many loss functions  $d(\cdot)$  satisfy the aforementioned conditions, although they may have quite different shapes. To illustrate the scope of this class of loss functions, we consider some examples. The first example of  $d(\cdot)$  is the popular quadratic loss function

$$d(z) = z^2. \quad (3.9)$$

This delivers a statistic based on the sum of squared differences between the fitted values of the null and alternative models:

$$\hat{L}_n^2 = \sum_{t=1}^n [\hat{g}(X_t) - g_0(X_t, \hat{\theta}_0)]^2. \quad (3.10)$$

This statistic is used in Hong and White (1995) and Horowitz and Spokoiny (2001) in an *i.i.d.* setup. It is also closely related to the statistics proposed by Härdle and Mammen (1993) and Pan, Yao and Wang (2007) but different from their statistics,  $\hat{L}_n^2$  in (3.10) does not involve any weighting which suffers from the undesirable feature as pointed out in Fan and Jiang (2007).

A second example of  $d(\cdot)$  is the truncated quadratic loss function

$$d(z) = \begin{cases} \frac{1}{2}z^2 & \text{if } |z| \leq c, \\ c|z| - \frac{1}{2}c^2 & \text{if } |z| > c, \end{cases} \quad (3.11)$$

where  $c$  is a prespecified constant. This loss function is used in robust  $M$ -estimation. It is expected to deliver a test robust to outliers that may cause extreme discrepancies between two estimators.

The quadratic and truncated quadratic loss functions give equal penalty to overestimation and underestimation of same magnitude. They cannot capture asymmetric loss features that may arise in practice. For example, central banks may be more concerned with underprediction than overprediction of inflation rates. For another example, in providing an estimate of the market value of a property of the owner, a real estate agent's underestimation and overestimation may have different consequences. If the valuation is in preparation for a future sale, underestimation may lead to the owner losing money and overestimation to market resistance (Varian 1975).

The above examples motivate using an asymmetric loss function for model validation. Examples of asymmetric loss functions are a class of so-called linex functions

$$d(z) = \frac{\beta}{\alpha^2} [\exp(\alpha z) - (1 + \alpha z)]. \quad (3.12)$$



For each pair of parameters  $(\alpha, \beta)$ ,  $d(z)$  is an asymmetric loss function. Here,  $\beta$  is a scale factor, and  $\alpha$  is a shape parameter. The magnitude of  $\alpha$  controls the degree of asymmetry, and the sign of  $\alpha$  reflects the direction of asymmetry. When  $\alpha < 0$ ,  $d(z)$  increases almost exponentially if  $z < 0$ , and almost linearly if  $z > 0$ , and conversely when  $\alpha > 0$ . Thus, for this loss function, underestimation is more costly than overestimation when  $\alpha < 0$ , and the reverse is true when  $\alpha > 0$ . For small values of  $|\alpha|$ ,  $d(z)$  is almost symmetric and not far from a quadratic loss function. Indeed, if  $\alpha \rightarrow 0$ , the linex loss function becomes a quadratic loss function:

$$d(z) \rightarrow \frac{\beta}{2} z^2.$$

However, when  $|\alpha|$  assumes appreciable values, the linex loss function  $d(z)$  will be quite different from a quadratic loss function. Thus, the linex loss function can be viewed as a generalization of the quadratic loss function allowing for asymmetry. This function was first introduced by Varian (1975) for real estate assessment. Zellner (1986) employs it in the analysis of several central statistical estimation and prediction problems in a Bayesian framework. Granger and Pesaran (1999) also use it to evaluate density forecasts, and Christoffersen and Diebold (1997) analyze the optimal prediction problem under this loss function. Figure 3.1 shows the shapes of the linex function for a variety of choices of  $(\alpha, \beta)$ .

Our loss function approach is by no means only applicable to regression functions. For example, in such contexts as probability density and spectral density estimation, one may compare two nonnegative density estimators, say  $f_{\hat{\theta}}$  and  $\hat{f}$ , using the Hellinger loss function

$$D(f_{\hat{\theta}}, \hat{f}) = \left(1 - \sqrt{f_{\hat{\theta}}/\hat{f}}\right)^2. \quad (3.13)$$

This is expected to deliver a consistent robust test for  $\mathbb{H}_0$  of (2.1). Our approach covers this loss function as well, because when  $f_{\hat{\theta}}$  and  $\hat{f}$  are close under  $\mathbb{H}_0$  of (2.1), we have

$$D(f_{\hat{\theta}}, \hat{f}) = \frac{1}{4} \left(\frac{f_{\hat{\theta}} - \hat{f}}{\hat{f}}\right)^2 + \textit{Remainder},$$

where the first order term in the Taylor expansion vanishes to 0 identically. Interestingly, our approach does not apply to the KLIC loss function

$$D(f_{\hat{\theta}}, \hat{f}) = -\ln\left(f_{\hat{\theta}}/\hat{f}\right), \quad (3.14)$$

which delivers the GLR  $\lambda_n$  in (2.3). This is because the Taylor expansion of (3.14) yields

$$D(f_{\hat{\theta}}, \hat{f}) = -\left(\frac{f_{\hat{\theta}} - \hat{f}}{\hat{f}}\right) + \frac{1}{2} \left(\frac{f_{\hat{\theta}} - \hat{f}}{\hat{f}}\right)^2 + \textit{Remainder}, \quad (3.15)$$

where the first order term in the Taylor expansion does not vanish to 0 identically. Hence, the first

two terms in (3.15) jointly determine the asymptotic distribution of the GLR statistic  $\lambda_n$ . As will be seen below, the presence of the first order term in the Taylor expansion of the KLIC loss function in (3.14) leads to an efficiency loss compared to our loss function approach for which the first order term of a Taylor expansion is identically 0 under the null.

**4. Asymptotic Null Distribution.** Using a local fit with kernel  $K : \mathbb{R} \rightarrow \mathbb{R}$  and bandwidth  $h \equiv h(n)$ , one could obtain a nonparametric regression estimator  $\hat{g}(\cdot)$  and compare it to the parametric model  $g(\cdot, \hat{\theta}_0)$  via a loss function, where  $\hat{\theta}_0$  is a consistent estimator for  $\theta_0$  under  $\mathbb{H}_0$ . To avoid undersmoothing (i.e., to choose  $h$  such that the squared bias of  $\hat{g}(\cdot)$  vanishes to 0 faster than the variance of  $\hat{g}(\cdot)$ ), we estimate the conditional mean of the estimated parametric residual

$$\hat{\varepsilon}_t = Y_t - g(X_t, \hat{\theta}_0),$$

and compare it to a zero function  $E(\varepsilon_t|X_t) = 0$  (implied by  $\mathbb{H}_0$ ) via a loss function criterion

$$\hat{Q}_n = \sum_{t=1}^n D[\hat{m}_h(X_t), 0] = \sum_{t=1}^n d[\hat{m}_h(X_t) - 0] = \sum_{t=1}^n d[\hat{m}_h(X_t)], \quad (4.1)$$

where  $\hat{m}_h(X_t)$  is a nonparametric estimator for  $E(\varepsilon_t|X_t)$ . This is essentially a bias-reduction device. It is proposed in Härdle and Mammen (1993) and also used in Fan and Jiang (2007) for the GLR test. This device helps remove the bias of nonparametric estimation, because there is no bias under  $\mathbb{H}_0$  when we estimate the conditional mean of the estimated model residuals. We note that the bias-reduction device does not lead to any efficiency gain of the loss function test. The same efficiency gain of the loss function approach over the GLR approach is obtained even when we compare estimators for  $E(Y_t|X_t)$ . In the latter case, however, more restrictive conditions on the bandwidth  $h$  are required to ensure that the bias vanishes sufficiently fast under  $\mathbb{H}_0$ .

For simplicity, we use the Nadaraya-Watson estimator

$$\hat{m}_h(x) = \frac{n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t \mathbf{K}_h(x - X_t)}{n^{-1} \sum_{t=1}^n \mathbf{K}_h(x - X_t)}, \quad (4.2)$$

where  $X_t = (X_{1t}, \dots, X_{pt})'$ ,  $x = (x_1, \dots, x_p)'$ , and

$$\mathbf{K}_h(x - X_t) = h^{-p} \prod_{i=1}^p K[h^{-1}(x_i - X_{it})].$$

We note that a local polynomial estimator could also be used, with the same asymptotic results.

To derive the null limit distributions of the loss function test based on  $\hat{Q}_n$  in (4.1) and the GLR statistic  $\lambda_n$  in a time series context, we provide the following regularity conditions:

**Assumption A.1:** (i) For each  $n \in \mathbb{N}$ ,  $\{(Y_t, X_t')' \in \mathbb{R}^{p+1}, t = 1, \dots, n\}$ ,  $p \in \mathbb{N}$ , is a stationary and absolutely regular mixing process with mixing coefficient  $\beta(j) \leq C\rho^j$  for all  $j \geq 0$ , where  $\rho \in (0, 1)$ , and  $C \in (0, \infty)$ ; (ii)  $E|Y_t|^{8+\delta} < C$  for some  $\delta \in (0, \infty)$ ; (iii)  $X_t$  has a compact support  $\mathbb{G} \subset \mathbb{R}^p$

with marginal probability density  $C^{-1} \leq f(x) \leq C$  for all  $x$  in  $\mathbb{G}$ , and  $f(\cdot)$  is twice continuously differentiable on  $\mathbb{G}$ ; (iv) the joint probability density of  $(X_t, X_{t-j})$ ,  $f_j(x, y) \leq C$  for all  $j > 0$  and all  $x, y \in \mathbb{G}$ , where  $C \in (0, \infty)$  does not depend on  $j$ ; (v)  $E(X_{it}^{4(1+\eta)}) \leq C$  for some  $\eta \in (0, \infty)$ ,  $1 \leq i \leq p$ ; (vi)  $\text{var}(\varepsilon_t) = \sigma^2$  and  $\sigma^2(x) = E(\varepsilon_t^2 | X_t = x)$  is continuous on  $\mathbb{G}$ .

**Assumption A.2:** (i) For each  $\theta \in \Theta$ ,  $g(\cdot, \theta)$  is a measurable function of  $X_t$ ; (ii) with probability one,  $g(X_t, \cdot)$  is twice continuously differentiable with respect to  $\theta \in \Theta$ , with  $E \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta} g(X_t, \theta) \right\|^{4+\delta} \leq C$  and  $E \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta \partial \theta'} g(X_t, \theta) \right\|^4 \leq C$ , where  $\Theta_0$  is a small neighborhood of  $\theta_0$  in  $\Theta$ .

**Assumption A.3:** There exists a sequence of constants  $\theta_n^* \in \text{int}(\Theta)$  such that  $n^{1/2}(\hat{\theta}_0 - \theta_n^*) = O_p(1)$ , where  $\theta_n^* = \theta_0$  under  $\mathbb{H}_0$  for all  $n \geq 1$ .

**Assumption A.4:** The kernel  $K : \mathbb{R} \rightarrow [0, 1]$  is a prespecified bounded symmetric probability density which satisfies the Lipschitz condition.

**Assumption A.5:**  $d : \mathbb{R} \rightarrow \mathbb{R}^+$  has a unique minimum at 0 and  $d(z)$  is monotonically nondecreasing as  $|z| \rightarrow \infty$ . Furthermore,  $d(z)$  is twice continuously differentiable at 0 with  $d(0) = 0$ ,  $d'(0) = 0$ ,  $D \equiv \frac{1}{2}d''(0) \in (0, \infty)$ , and  $|d''(z) - d''(0)| \leq C|z|$  for any  $z$  near 0.

Assumptions A.1 and A.2 are conditions on the DGP. For each  $t$ , we allow  $(X_t, Y_t)$  to depend on the sample size  $n$ . This facilitates local power analysis. For notational simplicity, we have suppressed the dependence of  $(X_t, Y_t)$  on  $n$ . We also allow time series data with weak serial dependence. For the  $\beta$ -mixing condition, see (e.g.) Doukan (1995). The compact support for regressor  $X_t$  is assumed in Fan, Zhang and Zhang (2001) for the GLR test to avoid the awkward problem of tackling the KLIC function. This assumption allows us to focus on essentials while maintaining a relatively simple treatment. It could be relaxed in several ways. For example, we could impose a weight function  $\mathbf{1}(|X_t| < C_n)$  in constructing  $Q_n$  and  $\lambda_n$ , where  $\mathbf{1}(\cdot)$  is the indicator function, and  $C_n$  can be either fixed or grow at a suitable rate as the sample size  $n \rightarrow \infty$ .

Assumption A.3 requires a  $\sqrt{n}$ -consistent estimator  $\hat{\theta}_0$  under  $\mathbb{H}_0$ , which need not be asymptotically most efficient. It can be the conditional least squares or quasi-MLE. Also, we do not need to know the asymptotic expansion structure of  $\hat{\theta}_0$ , because the sampling variation in  $\hat{\theta}_0$  does not affect the limit distribution of  $\hat{Q}_n$ . We can estimate  $\hat{\theta}_0$  and proceed as if it were equal to  $\theta_0$ . The replacement of  $\hat{\theta}_0$  with  $\theta_0$  has no impact on the limit distribution of  $\hat{Q}_n$ .

We first derive the limit distribution of the loss function test statistic.

**THEOREM 1 [LOSS FUNCTION TEST]** *Suppose Assumptions A.1–A.5 hold,  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , and  $p < 4$ . Define  $q_n = \hat{Q}_n / \hat{\sigma}_n^2$  where  $\hat{Q}_n$  is given in (4.1) and  $\hat{\sigma}_n^2 = n^{-1}SSR_1 = n^{-1} \sum_{t=1}^n [\hat{\varepsilon}_t - \hat{m}_h(X_t)]^2$ . Then (i) under  $\mathbb{H}_0$ ,  $s(K)q_n \stackrel{d}{\simeq} \chi_{\nu_n}^2$  as  $n \rightarrow \infty$ , in the sense that*

$$\frac{s(K)q_n - \nu_n}{\sqrt{2\nu_n}} \xrightarrow{d} N(0, 1),$$

where  $s(K) = \sigma^2 a(K) \int \sigma^2(x) dx / [Db(K) \int \sigma^4(x) dx]$ ,  $\nu_n = a^2(K) [\int \sigma^2(x) dx]^2 / [h^p b(K) \int \sigma^4(x) dx]$ ,  $a(K) = \int \mathbf{K}^2(\mathbf{u}) d\mathbf{u}$ ,  $b(K) = \int [\int \mathbf{K}(\mathbf{u} + \mathbf{v}) \mathbf{K}(\mathbf{v}) d\mathbf{v}]^2 d\mathbf{u}$ ,  $\mathbf{K}(\mathbf{u}) = \prod_{i=1}^p K(u_i)$ ,  $\mathbf{u} = (u_1, \dots, u_p)'$ .

(ii) Suppose in addition  $\text{var}(\varepsilon_t|X_t) = \sigma^2$  almost surely. Then  $s(K) = a(K)/[Db(K)]$  and  $\nu_n = \Omega a^2(K)/[h^p b(K)]$ , where  $\Omega$  is the Lebesgue's measure of the support of  $X_t$ .

Theorem 1 shows that under (and only under) conditional homoskedasticity, the factors  $s(K)$  and  $\nu_n$  do not depend on nuisance parameters and nuisance functions. In this case, the loss function test statistic  $q_n$ , like the GLR statistic  $\lambda_n$ , also enjoys the Wilks phenomena that its asymptotic distribution does not depend on nuisance parameters and nuisance functions. This offers great convenience in implementing the loss function test.

We note that the condition on the bandwidth  $h$  is relatively mild. In particular, no undersmoothing is required. This occurs because we estimate the conditional mean of the residuals of the parametric model  $g(X_t, \theta)$ . If we directly compared a nonparametric estimator of  $E(Y_t|X_t)$  with  $g(X_t, \theta)$ , we could obtain the same asymptotic distribution for  $q_n$ , but under a more restrictive condition on  $h$  in order to remove the effect of the bias. For simplicity, we consider the case with  $p < 4$ . A higher dimension  $p$  for  $X_t$  could be allowed by suitably modifying factors  $s(K)$  and  $\nu_n$ , but with more tedious expressions.

Theorem 1 also holds for the statistic  $q_n^0 = \hat{Q}_n/\hat{\sigma}_{n,0}^2$ , where  $\hat{\sigma}_{n,0}^2 = n^{-1}SSR_0$ , which is expected to have better sizes than  $q_n$  in finite samples under  $\mathbb{H}_0$  when using asymptotic theory. However,  $q_n$  may have better power than  $q_n^0$ , because  $SSR_0$  may be substantially larger than  $SSR_1$  under  $\mathbb{H}_A$ .

To compare the  $q_n$  and GLR tests, we have to derive the asymptotic distribution of the GLR statistic  $\lambda_n$  in a time series context, a formal result not available in the previous literature, although the GLR test has been widely applied in the time series context (Fan and Yao (2003)).

**THEOREM 2 [GLR TEST IN TIME SERIES]:** Suppose Assumptions A.1–A.5 hold,  $p < 4$ , and  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , and  $p < 4$ . Define  $\lambda_n$  as in (3.6), where  $SSR_1 = \sum_{t=1}^n [\hat{\varepsilon}_t - \hat{m}_h(X_t)]^2$ ,  $SSR_0 = \sum_{t=1}^n \hat{\varepsilon}_t^2$ , and  $\hat{\varepsilon}_t = Y_t - m(X_t, \hat{\theta})$ . Then (i) under  $\mathbb{H}_0$ ,  $r(K)\lambda_n \stackrel{d}{\simeq} \chi_{\mu_n}^2$  as  $n \rightarrow \infty$ , in the sense that

$$\frac{r(K)\lambda_n - \mu_n}{\sqrt{2\mu_n}} \xrightarrow{d} N(0, 1),$$

where  $r(K) = \sigma^2 c(K) \int \sigma^2(x) dx / [d(K) \int \sigma^4(x) dx]$ ,  $\mu_n = [c(K) \int \sigma^2(x) dx]^2 / [h^p d(K) \int \sigma^4(x) dx]$ ,  $c(K) = \mathbf{K}(0) - \frac{1}{2} \int \mathbf{K}^2(\mathbf{u}) d\mathbf{u}$ ,  $d(K) = \int [\mathbf{K}(\mathbf{u}) - \frac{1}{2} \int \mathbf{K}(\mathbf{u}+\mathbf{v})\mathbf{K}(\mathbf{v}) d\mathbf{v}]^2 d\mathbf{u}$ ,  $\mathbf{K}(\mathbf{u}) = \prod_{i=1}^p K(u_i)$ ,  $\mathbf{u} = (u_1, \dots, u_p)$ .

(ii) Suppose in addition  $\text{var}(\varepsilon_t|X_t) = \sigma^2$  almost surely. Then  $r(K) = c(K)/d(K)$  and  $\mu_n = \Omega c^2(K)/[h^p d(K)]$ , where  $\Omega$  is the Lebesgue's measure of the support of  $X_t$ .

Theorem 2 extends the results of Fan *et al.* (2001). We allow  $X_t$  to be a vector and allow time series data. We do not assume that the error  $\varepsilon_t$  is independent of  $X_t$  or the past history of  $\{X_t, Y_t\}$  so conditional heteroskedasticity in a time series context is allowed. This is consistent with the empirical stylized fact of volatility clustering for high frequency financial time series. We note that the proof of the asymptotic normality of the GLR test in a time series context is much more involved than in an *i.i.d.* context. It is interesting to observe that the Wilks phenomena do not hold under conditional heteroskedasticity because the factors  $r(K)$  and  $\mu_n$  involve the nuisance function

$\sigma^2(X_t) = \text{var}(\varepsilon_t|X_t)$ , which is unknown under  $\mathbb{H}_0$ . Conditional homoskedasticity is required to ensure the Wilks phenomena. In this case,  $r(K)$  and  $\mu_n$  are free of nuisance functions.

Like the  $q_n$  test, we also consider the case of  $p < 4$ . A higher dimension  $p$  could be allowed by suitably modifying factors  $r(K)$  and  $\mu_n$ , which would depend on the unknown density  $f(x)$  of  $X_t$  and thus are not free of nuisance functions even under conditional homoskedasticity.

**5. Relative Efficiency.** We now compare the relative efficiency between the loss function test  $q_n$  and the GLR test  $\lambda_n$  under the class of local alternatives

$$\mathbb{H}_n(a_n) : g_0(X_t) = g(X_t, \theta_0) + a_n \delta(X_t), \quad (5.1)$$

where  $\delta : \mathbb{R} \rightarrow \mathbb{R}$  is an unknown continuous function with  $E[\delta^4(X_t)] \leq C$ . The term  $a_n \delta(X_t)$  characterizes the departure of the model  $g(X_t, \theta_0)$  from the true function  $g_0(X_t)$  and the rate  $a_n$  is the speed at which the departure vanishes to 0 as the sample size  $n \rightarrow \infty$ . For notational simplicity, we have suppressed the dependence of  $g_0(X_t)$  on  $n$  here. Without loss of generality, we assume that  $\delta(X_t)$  is uncorrelated with  $X_t$ , namely  $E[\delta(X_t)X_t] = 0$ .

**THEOREM 3 [LOCAL POWER]** *Suppose Assumptions A.1–A.5 hold,  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , and  $p < 4$ . Then (i) under  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2}h^{-p/4}$ , we have*

$$\frac{s(K)q_n - \nu_n}{\sqrt{2\nu_n}} \xrightarrow{d} N(\psi, 1) \text{ as } n \rightarrow \infty,$$

where  $\psi = \sigma^2 E[\delta^2(X_t)] / \sqrt{2b(K) \int \sigma^4(x) dx}$ , and  $s(K)$  and  $\nu_n$  are as in Theorem 1. Suppose in addition  $\text{var}(\varepsilon_t|X_t) = \sigma^2$  almost surely. Then  $\psi = E[\delta^2(X_t)] / \sqrt{2b(K)\Omega}$ .

(ii) under  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2}h^{-p/4}$ , we have

$$\frac{r(K)\lambda_n - \mu_n}{\sqrt{2\mu_n}} \xrightarrow{d} N(\xi, 1) \text{ as } n \rightarrow \infty,$$

where  $\xi = \sigma^2 E[\delta^2(X_t)] / [2\sqrt{2d(K) \int \sigma^4(x) dx}]$ , and  $r(K)$  and  $\mu_n$  are as in Theorem 2. Suppose in addition  $\text{var}(\varepsilon_t|X_t) = \sigma^2$  almost surely. Then  $\xi = E[\delta^2(X_t)] / [2\sqrt{2d(K)\Omega}]$ .

When  $X_t$  is a scalar (i.e.,  $p = 1$ ) and  $h = n^{-2/9}$ , the factor  $a_n = n^{-1/2}h^{-p/4} = n^{-4/9}$  achieves the optimal rate in the sense of Ingster (1993a, 1993b, 1993c). Following a similar reasoning to Fan *et al.* (2001), we can show that the  $q_n$  test can also detect local alternatives with the optimal rate  $n^{-2k/(4k+p)}$  in the sense of Ingster (1993a, 1993b, 1993c), for the function space  $\mathcal{F}_k = \{\delta \in L^2 : \int \delta^{(k)}(x)^2 dx \leq C\}$ . For  $p = 1$  and  $k = 2$ , this is achieved by setting  $h = n^{-2/9}$ .

It is interesting to note that the noncentrality parameter  $\psi$  of the  $q_n$  test is independent of the curvature parameter  $D = d''(0)/2$  of the loss function  $d(\cdot)$ . This implies that all loss functions satisfying Assumption A.5 are asymptotically equally efficient under  $\mathbb{H}_n(a_n)$  in terms of Pitman's efficiency criterion (Pitman 1979, Ch.7), although their shapes may be different.

While the  $q_n$  and  $\lambda_n$  tests achieve the same optimal rate of convergence in the sense of Ingster (1993a, 1993b, 1993c), Theorem 4 below shows that under the same set of regularity conditions (including the same bandwidth  $h$  and the same kernel  $K(\cdot)$  for both tests),  $q_n$  is asymptotically more efficient than  $\lambda_n$  under  $\mathbb{H}_n(a_n)$ .

**THEOREM 4 [RELATIVE EFFICIENCY]:** *Suppose Assumptions A.1–A.5 hold,  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , and  $p < 4$ . Then Pitman’s relative efficiency of the  $q_n$  test over the GLR  $\lambda_n$  test under  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2}h^{-p/4}$  is given by*

$$\text{ARE}(q_n : \lambda_n) = \left\{ \frac{\int [2\mathbf{K}(\mathbf{u}) - \int \mathbf{K}(\mathbf{u} + \mathbf{v}) \mathbf{K}(\mathbf{v}) d\mathbf{v}]^2 d\mathbf{u}}{\int [\int \mathbf{K}(\mathbf{u} + \mathbf{v}) \mathbf{K}(\mathbf{v}) d\mathbf{v}]^2 d\mathbf{u}} \right\}^{\frac{1}{2-p\omega}}, \quad (5.2)$$

where  $\mathbf{K}(\mathbf{u}) = \prod_{i=1}^p K(u_i)$ ,  $\mathbf{u} = (u_1, \dots, u_p)'$ . The asymptotic relative efficiency  $\text{ARE}(q_n : \lambda_n)$  is larger than 1 for any kernel satisfying Assumption A.4 and the condition of  $K(\cdot) \leq 1$ .

Theorem 4 holds under both conditional heteroskedasticity and conditional homoskedasticity. It suggests that although the GLR  $\lambda_n$  test is a natural extension of the classical parametric LR test and is a generally applicable nonparametric inference procedure with many appealing features, it does not have the optimal power property of the classical LR test. In particular, the GLR test is always asymptotically less efficient than the loss function test under  $\mathbb{H}_n(a_n)$  whenever they use the same kernel  $K(\cdot)$  and the same bandwidth  $h$ , including the optimal kernel and the optimal bandwidth (if any) for the GLR test. The relative efficiency gain of the loss function test over the GLR test holds even if the GLR test  $\lambda_n$  uses the true likelihood function. This result is in sharp contrast to the classical LR test in a parametric setup, which is asymptotically most powerful according to the Neyman-Pearson Lemma.

Insight into the relative efficiency between  $q_n$  and  $\lambda_n$  can be obtained by a Taylor expansion of the  $\lambda_n$  statistic:

$$\lambda_n = \frac{1}{2} \frac{SSR_0 - SSR_1}{(SSR_1/n)} + \text{Remainder}, \quad (5.3)$$

where the remainder term is an asymptotically negligible higher order under  $\mathbb{H}_n(a_n)$ . This is equivalent to use of the loss function

$$D(g, g_\theta) = [Y_t - g(X_t, \theta)]^2 - [Y_t - g(X_t)]^2. \quad (5.4)$$

When  $g(X_t)$  is close to  $g(X_t, \theta_0)$ , the first order term in a Taylor expansion of  $D(g, g_\theta)$  around  $g_{\theta_0}$  does not vanish to 0 under  $\mathbb{H}_0$ . More specifically, the asymptotic distribution of  $\lambda_n$  is determined by the dominant term :

$$\frac{1}{2}[SSR_0 - SSR_1] = \frac{1}{2} \left[ \sum_{t=1}^n \hat{\varepsilon}_t^2 - \sum_{t=1}^n [\hat{\varepsilon}_t - \hat{m}_h(X_t)]^2 \right] = \sum_{t=1}^n \hat{\varepsilon}_t \hat{m}_h(X_t) - \frac{1}{2} \sum_{t=1}^n \hat{m}_h^2(X_t). \quad (5.5)$$

The first term in (5.5) corresponds to the first order term of a Taylor expansion of (5.4). It is a

second order  $V$ -statistic (Serfling 1980), and after demeaning, it can be approximated as a second order degenerate  $U$ -statistic. The second term in (5.5) corresponds to the second order term of a Taylor expansion of (5.4). It is a third order  $V$ -statistic, and can be approximated by a second order degenerate  $U$ -statistic after demeaning. These two degenerate  $U$ -statistics are of the same order of magnitude and jointly determine the asymptotic distribution of  $\lambda_n$ . In particular, the asymptotic variance of  $\lambda_n$  is determined by the variances of these two  $U$ -statistics and their covariance. In contrast, under Assumption A.5, a Taylor expansion suggests that the asymptotic distribution of the  $q_n$  statistic is determined by

$$\hat{Q}_n = D \sum_{t=1}^n \hat{m}_h^2(X_t) + \text{Remainder}, \quad (5.6)$$

which corresponds to the second term in the expansion of  $SSR_0 - SSR_1$  in (5.5). As it turns out, the asymptotic variance of this term alone is always smaller than the variance of the difference of the two terms in (5.5). This leads to a more efficient test than the GLR test, as is shown in Theorem 4. We note that the first term in (5.5), which causes an efficiency loss for the GLR test relative to the  $q_n$  test, is always present no matter whether we use the bias-reduction device (i.e., estimating the conditional mean of the estimated model residuals).

To assess the magnitude of the relative efficiency gain of the  $q_n$  test over the  $\lambda_n$  test, we consider a few commonly used multiweight kernels: the uniform, Epanechnikov, biweight, and triweight kernels (see Table 5.1 below). Suppose the bandwidth rate parameter  $\omega = 1/5, 2/9$  respectively in the univariate case (i.e.,  $p = 1$ ). The rate of  $\omega = 1/5$  gives the optimal bandwidth rate for estimating  $g_0(\cdot)$ , and the rate of  $\omega = 2/9$  achieves the optimal convergence rate in the sense of Ingster (1993a, 1993b, 1993c). Table 5.1 reports Pitman's asymptotic relative efficiencies (ARE). The efficiency gain of using the  $q_n$  test is substantial no matter the bandwidth  $h$  is of the order of  $n^{-1/5}$  or  $n^{-2/9}$ . Furthermore, there is little difference in the asymptotic relative efficiency between the two choices of  $h$ . These are confirmed in our simulation study below.

**TABLE 5.1: Asymptotic Relative Efficiency of the Loss Function Test over the GLR test**

	Uniform	Epanechnikov	Biweight	Triweight
$K(u)$	$\frac{1}{2}1( u  \leq 1)$	$\frac{3}{4}[1 - u^2]1( u  \leq 1)$	$\frac{15}{16}[1 - u^2]^21( u  \leq 1)$	$\frac{35}{32}[1 - u^2]^31( u  \leq 1)$
$ARE_1$	2.80	2.04	1.99	1.98
$ARE_2$	2.84	2.06	2.01	1.99

Note:  $ARE$  denotes Pitman's asymptotic relative efficiency of the loss function  $q_n$  test to the GLR  $\lambda_n$  test.  $ARE_1$  is for  $h = cn^{-1/5}$  and  $ARE_2$  is for  $h = cn^{-2/9}$ , for  $0 < c < \infty$ .

We emphasize that Theorem 4 does not imply that the GLR test should be abandoned. Indeed, it is a natural extension of the classical LR test and has many appealing features. It will remain as a useful general nonparametric inference procedure in practice.

While the relative efficiency of the loss function  $q_n$  test over the GLR  $\lambda_n$  test holds whenever the same bandwidth  $h$  and the same kernel  $K(\cdot)$  are used, the choice of an optimal bandwidth remains as important issue for each test. Theorems 1-4 allow for a wide range of the choices of  $h$ , but they do not provide a practical guidance on how to choose  $h$ . In practice, a simple rule of thumb is to choose  $h = S_X n^{-1/5}$  or  $h = S_X n^{-2/9}$ , where  $S_X^2$  is the sample variance of  $\{X_t\}_{t=1}^n$ . One could also choose a data-driven bandwidth using a cross-validation procedure, i.e., choose  $h = \arg \min_{c_1 n^{-1/(p+4)} \leq h \leq c_2 n^{-1/(p+4)}} \sum_{t=1}^n [\hat{\varepsilon}_t - \hat{m}_{h,t}(X_t)]^2$  for some prespecified constants  $0 < c_1 < c_2 < \infty$ , where for each given  $t$ ,  $\hat{m}_{h,t}(X_t)$  is the leave-one-out estimator that is based on the sample  $\{\hat{\varepsilon}_s, X_s\}_{s=1, s \neq t}^n$ . The bandwidth based on cross-validation is asymptotically optimal for estimation in terms of mean squared errors, but it may not be optimal for the  $q_n$  and  $\lambda_n$  tests. For testing problems, the central concern is the Type I error or Type II error, or both. Based on the Edgeworth expansion of the asymptotic distribution of a test statistic, Gao and Gijbels (2008) show that the choice of  $h$  affects both the Type I and Type II errors of a closely related nonparametric test, and usually there exists a tradeoff between Type I and Type II errors when choosing  $h$ . A sensible optimal rule is to choose  $h$  to maximize the power of a test given a significance level. Gao and Gijbels (2008) derive the leading terms of the size and power functions of their test statistic, and then choose a bandwidth to maximize the power under a class of local alternatives similar to (5.1) under a controlled significance level, i.e., to choose  $h = \max_{h \in B_n(\alpha)} \beta_n(h)$ , where  $B_n(\alpha) = \{h : \alpha - c_{\min} < \alpha_n(h) < \alpha + c_{\min}\}$  for some prespecified small constant  $c_{\min} \in (0, \alpha)$ , and  $\alpha_n(h)$  and  $\beta_n(h)$  are the size and power functions of the nonparametric test. They then propose a data-driven bandwidth in combination with a bootstrap and show that it works well in finite samples. Unfortunately, Gao and Gijbels' (2008) results cannot be directly applied to either the  $q_n$  or  $\lambda_n$  test, because the higher order terms of  $\alpha_n(h)$  and  $\beta_n(h)$  depend on the form of test statistic, the DGP, the kernel  $K$  and the bandwidth  $h$ , among many other things. However, it is possible to extend their approach to the  $q_n$  and  $\lambda_n$  tests to obtain their optimal bandwidths respectively. As the associated technicality is quite involved, we leave this important problem for subsequent work. We note that Sun, Phillips and Jin (2008), in a different context, also consider a data-driven bandwidth by minimizing a weighted average of the Type I and Type II errors of a test, namely choose  $h = \arg \min_h \left( \frac{w_n}{1+w_n} e_n^I + \frac{1}{1+w_n} e_n^{II} \right)$ , where  $e_n^I$  and  $e_n^{II}$  are the Type I and Type II errors respectively, and  $w_n$  is a weight function that reflects the relative importance of  $e_n^I$  and  $e_n^{II}$ .

**6. Monte Carlo Evidence.** We now compare the finite sample performance of the loss function  $q_n$  test and the GLR  $\lambda_n$  test. To examine the sizes of the tests, we consider the following a null linear regression model in a time series context:

DGP 0 [Linear Regression]:

$$\begin{cases} Y_t = 1 + X_t + \varepsilon_t, \\ X_t = 0.5X_{t-1} + v_t, \\ v_t \sim i.i.d.N(0, 1). \end{cases}$$

Here,  $X_t$  is truncated within its two standard deviations. To examine robustness of the tests, we



consider a variety of distributions for the error  $\varepsilon_t$  : (i)  $\varepsilon_t \sim i.i.d. N(0, 1)$ , (ii)  $\varepsilon_t \sim i.i.d. \text{Student-}t_5$ , (iii)  $\varepsilon_t \sim i.i.d. U[0, 1]$ , (iv)  $\varepsilon_t \sim i.i.d. \ln N(0, 1)$ , and (v)  $\varepsilon_t \sim i.i.d. \chi_1^2$ , where the  $\varepsilon_t$  in (iii)–(v) have been scaled to have mean 0 and variance 1.

Because the asymptotic normal approximation for the  $q_n$  and  $\lambda_n$  tests might not perform well in finite samples, we also use a conditional bootstrap procedure based on the Wilks phenomena:

*Step 1:* Obtain the parameter estimator  $\hat{\theta}_0$  (e.g., OLS) of the null linear regression model, and the nonparametric estimator  $\hat{g}(X_t)$ .

*Step 2:* Compute the  $q_n$  statistic and the residual  $\hat{\varepsilon}_t = Y_t - \hat{g}(X_t)$  from the nonparametric model.

*Step 3:* Conditional on each  $X_t$ , draw a bootstrap error  $\varepsilon_t^*$  from the centered empirical distribution of  $\hat{\varepsilon}_t$  and compute  $Y_t^* = X_t' \hat{\theta}_0^* + \varepsilon_t^*$ . This forms a conditional bootstrap sample  $\{X_t, Y_t^*\}_{t=1}^n$ .

*Step 4:* Use the conditional bootstrap sample  $\{X_t, Y_t^*\}_{t=1}^n$  to compute a bootstrap statistic  $q_n^*$ , using the same kernel  $K(\cdot)$  and the same bandwidth  $h$  as in Step 2.

*Step 5:* Repeat Steps 3 and 4 for a total of  $B$  times, where  $B$  is a large number. We then obtain a collection of bootstrap test statistics,  $\{q_{nl}^*\}_{l=1}^B$ .

*Step 6:* Compute the bootstrap  $P$  value  $P^* = B^{-1} \sum_{l=1}^B \mathbf{1}(q_n < q_{nl}^*)$ . Reject  $\mathbb{H}_0$  at a prespecified significance level  $\alpha$  if and only if  $P^* < \alpha$ .

When conditional heteroskedasticity exists, we can modify Step 2 by using a wild bootstrap for  $\{\hat{\varepsilon}_t^*\}$ . If  $X_t$  contains lagged dependent variables, we can use a recursive simulation method; see (e.g.) Franke *et al.* (2002). For space, we do not justify the validity of the bootstrap here. Fan and Jiang (2007, Theorem 7) show the consistency of the bootstrap for the GLR test in an *i.i.d.* context. We could establish the consistency of the bootstrap for our loss function test by following the approaches of Fan and Jiang (2007) and Gao and Gijbels (2008).

We consider two versions of the loss function test, one is to standardize  $\hat{Q}_n$  by  $\hat{\sigma}_n^2 = n^{-1}SSR_1$ , where  $SSR_1$  is the sum of squared residuals of the nonparametric regression estimates. This is denoted as  $q_n$ . The other version is to standardize  $\hat{Q}_n$  by  $\hat{\sigma}_{n,0}^2 = n^{-1}SSR_0$ , where  $SSR_0$  is the sum of squared residuals of the null linear model. This is denoted as  $q_n^0$ . It is expected that  $q_n$  may be more powerful than  $q_n^0$  in finite samples under  $\mathbb{H}_A$ , because  $SSR_0$  is expected to be significantly larger than  $SSR_1$  under  $\mathbb{H}_A$ . To construct the  $q_n$  and  $q_n^0$  tests, we choose the family of linex loss functions in (3.12), with  $(\alpha, \beta) = (0, 1), (0.2, 1), (0.5, 1)$  and  $(1, 1)$  respectively (see Figure 3.1 for their shapes). The choice of  $(\alpha, \beta) = (0, 1)$  corresponds to the symmetric quadratic loss function, while the degree of asymmetry of the loss function increases as  $\alpha$  increases (the choice of  $\beta$  has no impact on the  $q_n$  tests). Various choices of  $(\alpha, \beta)$  thus allow us to examine sensitivity of the power of the  $q_n$  tests to the choices of the loss function. Rather conveniently, when using the bootstrap procedure, there is no need to compute the centering and scaling factors for the  $q_n$  and  $q_n^0$  tests; it suffices to compare the statistic  $q_n$  or  $q_n^0$  with their bootstrap counterparts. We choose  $B = 99$ . The same bootstrap is used for the GLR test  $\lambda_n$ .

To examine the power of the tests, we consider three nonlinear DGP's:

DGP 1 [*Quadratic Regression*]:

$$Y_t = 1 + X_t + \theta X_t^2 + \varepsilon_t,$$

DGP 2 [*Threshold Regression*]:

$$Y_t = 1 + X_t \mathbf{1}(X_t > 0) + (1 + \theta) X_t \mathbf{1}(X_t \leq 0) + \varepsilon_t,$$

DGP 3 [*Smooth Transition Regression*]:

$$Y_t = 1 + X_t + [1 - \theta F(X_t)] X_t + \varepsilon_t,$$

where  $F(X_t) = [1 + \exp(-X_t)]^{-1}$ .

We consider various values for  $\theta$  in each DGP to examine how the power of the tests changes as the value of  $\theta$  changes.

To examine sensitivity of all tests to the choices of  $h$ , we consider  $h = S_X n^{-\omega}$  for  $\omega = \frac{2}{9}$  and  $\frac{1}{5}$  respectively, where  $S_X$  is the sample standard deviation of  $\{X_t\}_{t=1}^n$ . These correspond to the optimal rate of convergence in the sense of Ingster (1993a, 1993b, 1993c) and the optimal rate of estimation in terms of mean squared errors respectively. The results are similar. Here, we focus our discussion on the results with  $h = S_X n^{-2/9}$ , as reported in Tables 6.1-6.5. The results with  $h = S_X n^{-1/5}$  are reported in Tables S.1-S.5 of Supplementary Material. We use the uniform kernel  $K(z) = \frac{1}{2} \mathbf{1}(|z| \leq 1)$  for all tests. We have also used the biweight kernel, and the results are very similar (so, not reported here).

Tables 6.1 and 6.2 report the empirical rejection rates of the tests under  $\mathbb{H}_0$  (DGP 0) at the 10% and 5% levels, using both asymptotic and bootstrap critical values respectively. We first examine the size of the tests using asymptotic critical values, with  $n = 100, 250$  and  $500$  respectively. Table 6.1 shows that all tests,  $\lambda_n, q_n$  and  $q_n^0$ , have reasonable sizes in finite samples, and they are robust to various error distributions, but they all show some underrejection, particularly at the 10% level. The  $q_n$  and  $\lambda_n$  tests have similar sizes in most cases, whereas  $q_n^0$  shows a bit more underrejection. Overall, the sizes of the  $q_n, q_n^0$  and  $\lambda_n$  tests display some underrejections in most cases in finite samples, but they are not unreasonable.

Next, we examine the size of the tests based on the bootstrap. Table 6.2 shows that overall, the rejection rates of all tests based on the bootstrap are close to the significance levels (10% and 5%), indicating the gain of using the bootstrap in finite samples. The sizes of all tests are robust to a variety of error distributions, confirming the Wilks phenomena that the asymptotic distribution of both the  $q_n$  and  $\lambda_n$  tests are distribution free. For the loss function tests  $q_n$  and  $q_n^0$ , the sizes are very similar for different choices of parameters  $(\alpha, \beta)$  governing the shape of the linex loss function. We note that when asymptotic critical values are used, the sizes of the tests with  $h = S_X n^{-2/9}$  are slightly better than with  $h = S_X n^{-1/5}$ . When bootstrap critical values are used, however, the sizes of all tests with  $h = S_X n^{-2/9}$  and  $h = S_X n^{-1/5}$  respectively are very similar.

Next, we turn to the powers of the tests under  $\mathbb{H}_A$ . Since the sizes of the tests using asymptotic critical values are different in finite samples, we use the bootstrap procedure only, which delivers similar sizes close to significance levels and thus provides a fair ground for comparison. Tables 6.3-6.5 report the empirical rejection rates of the tests under DGP 1 (quadratic regression), DGP 2 (threshold regression), and DGP 3 (smooth transition regression), respectively. For all DGPs, the loss function tests  $q_n$  and  $q_n^0$  are more powerful than the GLR test, confirming our asymptotic efficiency analysis. Interestingly, for the two loss function tests,  $q_n$ , which is standardized by the nonparametric  $SSR_1$ , is roughly equally powerful to  $q_n^0$ , which is standardized by the parametric  $SSR_0$ , although asymptotic analysis suggests that  $q_n$  should be more powerful than  $q_n^0$  under  $\mathbb{H}_A$ , because  $SSR_0$  is significantly larger than  $SSR_1$  under  $\mathbb{H}_A$ . Obviously, this is due to the use of the bootstrap. Since the bootstrap statistics  $q_n^*$  and  $q_n^{0*}$  are standardized by  $SSR_0^*$  and  $SSR_1^*$  respectively, where  $SSR_1^* < SSR_0^*$ , the ranking between  $q_n$  and  $q_n^*$  remains more or less similar to the ranking between  $q_n^0$  and  $q_n^{0*}$ , and therefore  $q_n^*$  and  $q_n^{0*}$  have similar power. Under DGP 1, the powers of  $q_n$  and  $q_n^0$  increase as the degree of asymmetry of the linex loss function, which is indexed by  $\alpha$ , increases. When  $\alpha = 1$ , the powers of  $q_n$  and  $q_n^0$  are substantially higher than the GLR test. Under DGP 2, there is some tendency that the powers of  $q_n$  and  $q_n^0$  increase in  $\alpha$  for  $\theta < 0$ , whereas they decrease in  $\alpha$  for  $\theta > 0$ . When  $\theta$  is close to 0,  $q_n$  and  $q_n^0$  have similar power to the GLR test, but as  $|\theta| > 0$  increases, they are more powerful than the GLR test. Similarly, under DGP 3, the powers of  $q_n$  and  $q_n^0$  increase in  $\alpha$  for  $\theta < 0$ , whereas they decrease in  $\alpha$  for  $\theta > 0$ . Nevertheless, by and large, the powers of both  $q_n$  and  $q_n^0$  do not change much for the different choices of parameters  $(\alpha, \beta)$  governing the shape of the linex loss function. Although the shape of the loss function changes dramatically when  $\alpha$  changes from 0 to 1, the powers of  $q_n$  and  $q_n^0$  remain relatively robust.

All tests become more powerful as the departures from linearity increases (as characterized by the value of  $\theta$  in each DGP), and as the sample size  $n$  increases.

**7. Conclusion.** The GLR test has been proposed as a generally applicable method for nonparametric testing problems. It inherits many advantages of the maximum LR test for parametric models. In this paper, we have shown that despite its general nature and many appealing features, the GLR test does not have the optimal power property of the classical LR test. We propose a loss function test in a time series context. The new test enjoys the same appealing features as the GLR test, but is more powerful in terms of Pitman's asymptotic efficiency. This holds no matter what kernel and bandwidth are used, and even when the true likelihood function is available for the GLR test. The efficiency gain, together with more relevance to decision making under uncertainty of using a loss function, suggests that the loss function approach can be a generally applicable and powerful nonparametric inference procedure alternative to the GLR principle.

## MATHEMATICAL APPENDIX

Throughout the appendix, we let  $\tilde{m}_h(x)$  be defined in the same way as  $\hat{m}_h(x)$  in (4.2) with  $\{\varepsilon_t = Y_t - g_0(X_t)\}_{t=1}^n$  replacing  $\{\hat{\varepsilon}_t = Y_t - g(X_t, \hat{\theta}_0)\}_{t=1}^n$ . Also,  $C \in (1, \infty)$  denotes a generic bounded

constant. This appendix provides the structure of our proof strategy. We leave the detailed proofs of most technical lemmas and propositions to a Supplementary Material.

**Proof of Theorem 1:** Theorem 1 follows as a special case of Theorem 3(i) with  $\delta(X_t) = 0$ . ■

**Proof of Theorem 2:** Theorem 2 follows as a special case of Theorem 3(ii) with  $\delta(X_t) = 0$ . ■

**Proof of Theorem 3(i):** We shall first derive the asymptotic distribution of  $q_n$  under  $\mathbb{H}_n(a_n)$ . From Lemmas 3.1 and 3.2 and Propositions 3.1 and 3.2 below, we can obtain

$$\frac{h^{\frac{p}{2}} D^{-1} q_n - h^{-\frac{p}{2}} \sigma^{-2} \int K^2(u) du \int \sigma^2(x) dx}{\sqrt{2\sigma^{-4} \int [\int K(u)K(u+v)du]^2 dv \int \sigma^4(x) dx}} \xrightarrow{d} N(\psi, 1).$$

The desired result of Theorem 3(i) then follows immediately.

**Lemma 3.1:** Under the conditions of Theorem 3,  $\hat{Q}_n = D \sum_{t=1}^n \hat{m}_h^2(X_t) + o_p(h^{-p/2})$ .

**Lemma 3.2:** Under the conditions of Theorem 3,  $\sum_{t=1}^n \hat{m}_h^2(X_t) = n \int \hat{m}_h^2(x) f(x) dx + o_p(h^{-p/2})$ .

**Proposition 3.1:** Under the conditions of Theorem 3,  $n \int \hat{m}_h^2(x) f(x) dx = n \int \tilde{m}_h^2(x) f(x) dx + h^{-p/2} E[\delta^2(X_t)] + o_p(h^{-p/2})$ .

**Proposition 3.2:** Under the conditions of Theorem 3, and  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2} h^{-p/4}$ ,

$$\left[ nh^{p/2} \int \tilde{m}_h^2(x) f(x) dx - h^{-\frac{p}{2}} a(K) \int \sigma^2(x) dx \right] / \sqrt{2b(K) \int \sigma^4(x) dx} \xrightarrow{d} N(\psi, 1).$$

**Proof of Lemma 3.1:** Given in Supplementary Material.

**Proof of Lemma 3.2:** Given in Supplementary Material.

**Proof of Proposition 3.1:** Given in Supplementary Material. ■

**Proof of Proposition 3.2:** Proposition 3.2 follows from Lemmas 3.3 and 3.4 below.

**Lemma 3.3:** Put  $\hat{H}_q = n^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} H_n(Z_t, Z_s)$ , where  $Z_t = (\varepsilon_t, X_t)'$ ,  $H_n(Z_t, Z_s) = 2\varepsilon_t \varepsilon_s W_h(X_t, X_s)$ , and

$$W_h(X_t, X_s) = \int \frac{\mathbf{K}_h(X_t - x) \mathbf{K}_h(X_s - x)}{f(x)} dx.$$

Suppose Assumptions A.1 and A.4 hold,  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , and  $p < 4$ . Then

$$n \int \hat{m}_h^2(x) g(x) dx = h^{-p} \int \mathbf{K}^2(\mathbf{u}) d\mathbf{u} \int \sigma^2(x) dx + \hat{H}_q + o_p(h^{-p/2}).$$

**Lemma 3.4:** Suppose Assumptions A.1 and A.4 hold, and  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ . Define  $V_q = 2 \int [\int \mathbf{K}(\mathbf{v}) \mathbf{K}(\mathbf{u} + \mathbf{v}) d\mathbf{v}]^2 du \int \sigma^4(x) dx$ . Then  $V_q^{-1/2} h^{p/2} \hat{H}_q \xrightarrow{d} N(\psi, 1)$ .

Since Lemmas 3.3 and 3.4 are the key results for deriving the asymptotic distributions of the proposed  $q_n$  statistic when  $\{\varepsilon_t\}$  may not be an *i.i.d.* sequence nor martingale difference sequence, we provide detailed proofs for them below.

**Proof of Lemma 3.3:** Let  $\hat{F}_n(x)$  be the empirical distribution function of  $\{X_t\}_{t=1}^n$ . We have

$$\begin{aligned} & n \int \hat{m}_h^2(x) f(x) dx \\ &= n \int \frac{[n^{-1} \sum_{s=1}^n \varepsilon_s \mathbf{K}_h(X_t - X_s)]^2}{f(x)} dx + \int \left[ n^{-1} \sum_{s=1}^n \varepsilon_s \mathbf{K}_h(X_t - X_s) \right]^2 \left[ \frac{1}{\hat{f}^2(x)} - \frac{1}{f^2(x)} \right] f(x) dx \end{aligned}$$

$$\begin{aligned}
&= n^{-1} \sum_{t=1}^n \sum_{s=1}^n \varepsilon_t \varepsilon_s \int \frac{\mathbf{K}_h(X_t - x) \mathbf{K}_h(X_s - x)}{f(x)} dx + O_p(n^{-1} h^{-p}) O_p(n^{-1/2} h^{-p/2} \ln n + h^2) \\
&= n^{-1} \sum_{t=1}^n \varepsilon_t^2 \int \frac{\mathbf{K}_h^2(X_t - x)}{f(x)} dx + n^{-1} \sum_{1 \leq s < t \leq n} 2 \varepsilon_t \varepsilon_s \int \frac{\mathbf{K}_h(X_t - x) \mathbf{K}_h(X_s - x)}{f(x)} + o_p(h^{-p/2}) \\
&= \hat{C}_q + \hat{H}_q + o_P(h^{-p/2}), \tag{A.1}
\end{aligned}$$

where we have made use of the fact that  $\sup_{x \in \mathbb{G}} |\hat{f}(x) - f(x)| = O_p(n^{-1/2} h^{-p/2} \ln n + h^2)$  given Assumption A.2, and  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ .

By change of variable, the law of iterated expectations, and Assumption A.1, we can obtain

$$E(\hat{C}_q) = \iint \sigma^2(x) \frac{\mathbf{K}_h^2(y-x)}{f(x)} f(y) dx dy = h^{-p} \int \mathbf{K}^2(u) du \int \sigma^2(x) dx [1 + O(h^2)]. \tag{A.2}$$

On the other hand, by Chebyshev's inequality and the fact that  $E(\hat{C}_q - E\hat{C}_q)^2 = O_p(n^{-1} h^{-2p})$  given Assumption A.1, we have

$$\hat{C}_q = E(\hat{C}_q) + O_p(n^{-1/2} h^{-p}). \tag{A.3}$$

Combining (A.1)–(A.3) and  $p < 4$  then yields the desired result of Lemma 3.3. ■

**Proof of Lemma 3.4:** Because  $E[H_n(Z_t, z)] = E[H_n(z', Z_s)] = 0$  for all  $z, z', \hat{H}_q \equiv n^{-1} \sum_{1 \leq s < t \leq n} H_n(Z_t, Z_s)$  is a degenerate  $U$ -statistic. Following Tenreiro's (1997) central limit theorem for degenerate  $U$ -statistics of a time series context process, we have  $[n^{-2} \sum_{1 \leq s < t \leq n} E[h^p H_n^2(Z_t, Z_s)]]^{-1/2} h^{p/2} \hat{H}_q \xrightarrow{d} N(0, 1)$  as  $n \rightarrow \infty$  if the following conditions are satisfied: For some constants  $\delta_0 > 0, \gamma_0 < \frac{1}{2}$  and  $\gamma_1 > 0$ , (i)  $u_n(4 + \delta_0) = O(n^{\gamma_0})$ , (ii)  $v_n(2) = o(1)$ , (iii)  $w_n(2 + \frac{\delta_0}{2}) = o(n^{\frac{1}{2}})$ , and (iv)  $z_n(2)n^{\gamma_1} = O(1)$ , where

$$\begin{aligned}
u_n(r) &= h^{p/2} \max \left\{ \max_{1 \leq t \leq n} \|H_n(Z_t, Z_0)\|_r, \|H_n(Z_0, \bar{Z}_0)\|_r \right\}, \\
v_n(r) &= h^p \max \left\{ \max_{1 \leq t \leq n} \|G_{n0}(Z_t, Z_0)\|_r, \|G_{n0}(Z_0, \bar{Z}_0)\|_r \right\}, \\
w_n(r) &= h^p \|G_{n0}(Z_0, Z_0)\|_r, \\
z_n(r) &= h^p \max_{0 \leq t \leq n, 1 \leq s \leq n} \max \left\{ \|G_{ns}(Z_t, Z_0)\|_r, \|G_{ns}(Z_0, Z_t)\|_r, \|G_{ns}(Z_0, \bar{Z}_0)\|_r \right\},
\end{aligned}$$

$G_{ns}(u, v) = E[H_n(Z_s, u)H_n(Z_0, v)]$  for  $s \in \mathbb{N}$  and  $u, v \in \mathbb{R}^p$ ,  $\bar{Z}_0$  is an independent copy of  $Z_0$ , and  $\|\xi\|_r = E_r^{1/r} |\xi|^r$ .

We first show  $n^{-2} \sum_{1 \leq s < t \leq n} h^p E[H_n^2(Z_t, Z_s)] \rightarrow V_q$  as  $n \rightarrow \infty$ . By change of variables and Assumption A.1, it is straightforward to calculate

$$\begin{aligned}
n^{-2} \sum_{1 \leq s < t \leq n} h^p E[H_n^2(Z_t, Z_s)] &= 4h^p n^{-2} \sum_{1 \leq s < t \leq n} E[\varepsilon_t^2 \varepsilon_s^2 W_h^2(X_t, X_s)] \\
&\rightarrow 2 \int \left[ \int \mathbf{K}(v) \mathbf{K}(u+v) dv \right]^2 du \int \sigma^4(x) dx \equiv V_q. \tag{A.4}
\end{aligned}$$

We now verify Conditions (i)–(iv). We first consider Condition (i). By the Cauchy-Schwarz inequality and change of variables, we have for all  $t \geq 0$ ,

$$\begin{aligned}
E \left| h^{p/2} H_n(Z_t, Z_0) \right|^r &= 2^\gamma h^{\frac{p}{2}r} E \left| \varepsilon_t^r \varepsilon_0^r W_h^r(X_t, X_0) \right| \\
&\leq 2^\gamma h^{\frac{p}{2}r} (E \varepsilon_0^{2cr})^{\frac{1}{c}} (E |W_h(X_t, X_0)|^{cr})^{\frac{1}{c}} \\
&\leq Ch^{\frac{p}{2}r} \left[ \int |W_h(x, x_0)|^{cr} f_{X_t, X_0}(x, x_0) dx dx_0 \right]^{\frac{1}{c}} \\
&\leq Ch^{\frac{p}{2}r} (h^{-pcr} h^p)^{\frac{1}{c}} \leq Ch^{-\frac{r}{2}p + \frac{p}{c}},
\end{aligned}$$

for all  $c > 1$ , and given  $E(\varepsilon_t^{8+\delta}) \leq C$ . We obtain  $\|h^{p/2} H_n(Z_t, Z_0)\|_r = \left(Ch^{-\frac{r}{2}p + \frac{p}{c}}\right)^{\frac{1}{r}} \leq Ch^{-\frac{p}{2} + \frac{p}{cr}}$ . Given  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , we have  $\|h^{p/2} H_n(Z_t, Z_0)\|_r \leq Cn^{\omega p(\frac{1}{2} - \frac{2}{8+\delta})}$ , with  $c = \frac{8+\delta}{2r}$  and if  $r < 4 + \frac{\delta}{2}$ . By a similar argument and replacing  $f_{X_t, X_0}(x, x_0)$  with  $f(x)f(x_0)$ , we can obtain the same order of magnitude for  $\|h^{p/2} H_n(Z_0, \bar{Z}_0)\|_r$ . Hence, we obtain  $u_n(r) \leq Cn^{\omega p(\frac{1}{2} - \frac{2}{8+\delta})}$ , and Condition (i) holds by setting  $\gamma_0 = \omega p(\frac{1}{2} - \frac{2}{8+\delta})$ .

Now we verify Condition (ii). Note that for all  $s \geq 0$ , we have

$$\begin{aligned}
G_{ns}(z, z') &= E [H_n(Z_s, z) H_n(Z_0, z')] \\
&= 4E [\varepsilon_t \varepsilon W_h(X_s, x) \varepsilon_0 \varepsilon' W_h(X_0, x')] \\
&= 4\varepsilon \cdot \varepsilon' E [\varepsilon_s \varepsilon_0 W_h(X_s, x) W_h(X_0, x')],
\end{aligned}$$

where  $z = (\varepsilon, x)$  and  $z' = (\varepsilon', x')$ . To compute the order of magnitude for  $v_n(r)$ , we first consider the case of  $s = 0$ . We have

$$\begin{aligned}
G_{n0}(z, z') &= 4\varepsilon \varepsilon' E_0 [\bar{\varepsilon}_0^2 W_h(\bar{X}_0, x) W_h(\bar{X}_0, x')] \\
&= 4\varepsilon \varepsilon' E_0 [\sigma^2(\bar{X}_0) W_h(\bar{X}_0, x) W_h(\bar{X}_0, x')],
\end{aligned}$$

where  $E_0(\cdot)$  is an expectation taken over  $(\bar{X}_0, \bar{\varepsilon}_0)$ . By the Cauchy-Schwarz inequality and change of variables, we have

$$\begin{aligned}
E |h^p G_{n0}(Z_t, Z_0)|^2 &= 16E |h^{2p} \varepsilon_t^2 \varepsilon_0^2 E_0^2 [\sigma^2(\bar{X}_0) W_h(\bar{X}_0, X_t) W_h(\bar{X}_0, X_0)]| \\
&\leq 16h^{2p} E \left| \varepsilon_t^2 \varepsilon_0^2 [E_0 \sigma^{2c}(\bar{X}_0)]^{\frac{2}{c}} [E_0 W_h^c(\bar{X}_0, X_t) W_h^c(\bar{X}_0, X_0)]^{\frac{2}{c}} \right| \\
&\leq 16h^{2p} C [E |\varepsilon_t^4 \varepsilon_0^4|]^{\frac{1}{2}} \left\{ E |E_0 (W_h^c(\bar{X}_0, X_t) W_h^c(\bar{X}_0, X_0))|^{\frac{4}{c}} \right\}^{\frac{1}{2}} \\
&\leq Ch^{2p} \left\{ E |h^{-2cp+p} \mathbf{A}_{c,h}(X_t, X_0)|^{\frac{4}{c}} \right\}^{\frac{1}{2}} \\
&= O \left( h^{2p} \left[ h^{(-2cp+p)\frac{4}{c}} h^p \right]^{\frac{1}{2}} \right) = O \left( h^{(\frac{2}{c} - \frac{3}{2})p} \right)
\end{aligned}$$

for any  $c > 1$ , where

$$\begin{aligned}
E_0[W_h^c(\bar{X}_0, X_t) W_h^c(\bar{X}_0, X_0)] &= h^{-2cp+p} \int W_h^c(\bar{x}_0, X_t) W_h^c(\bar{x}_0, X_0) f(\bar{x}_0) d\bar{x}_0 \\
&= h^{-2cp+p} \mathbf{A}_{c,h}(X_t, X_0)
\end{aligned}$$

by change of variable, where  $\mathbf{A}_{c,h}(X_t, X_0)$  is a function similar to  $\mathbf{K}_h(X_t - X_0)$ . Thus, we obtain

$\|h^p G_{n0}(Z_t, Z_0)\|_2 \leq Ch^{(\frac{1}{c}-\frac{3}{4})p}$ . By a similar argument, we obtain the same order of magnitude for  $\|h^p G_{n0}(Z_t, \bar{Z}_0)\|_2$ . Thus, we have  $v_n(r) \leq Ch^{(\frac{1}{c}-\frac{3}{4})p}$ , and Condition (ii) holds, i.e.,  $v_n(2) = o(1)$ , with  $1 < c < \frac{4}{3}$ .

Next, to verify Condition (iii), we shall evaluate  $\|h^p G_{n0}(Z_0, \bar{Z}_0)\|_r$  for  $r < 2 + \frac{\delta_0}{4}$ . By the Cauchy-Schwarz inequality and change of variables, we have

$$\begin{aligned} E|h^p G_{n0}(Z_0, Z_0)|^r &= 4^\gamma E|h^{rp} \varepsilon_0^r \varepsilon_0^r E_0^r [\sigma^2(\bar{X}_0) W_h(\bar{X}_0, X_0) W_h(\bar{X}_0, X_0)]| \\ &\leq 4^\gamma h^{2p} E|\varepsilon_0^{2r} \sigma^{2c}(\bar{X}_0)^{\frac{r}{c}} [E_0 W_h^{2c}(\bar{X}_0, X_0)]^{\frac{r}{c}}| \\ &\leq Ch^{rp} (E\varepsilon_0^{4r})^{\frac{1}{2}} \left[ E|E_0 W_h^{2c}(\bar{X}_0, X_0)|^{\frac{2r}{c}} \right]^{\frac{1}{2}} \\ &= O\left( h^{rp} \left[ h^{(1-2c)p \frac{2r}{c}} \right]^{\frac{1}{2}} \right) = O\left( h^{rp(\frac{1}{c}-1)} \right), \end{aligned}$$

where  $E_0[W_h^{2c}(\bar{X}_0, X_0)] = \int W_h^{2c}(\bar{x}_0, X_0) f_{\bar{X}_0}(\bar{x}_0) d\bar{x}_0 = O(h^{(1-2c)p})$  by change of variable. Thus, we obtain  $\|h^p G_{n0}(Z_0, Z_0)\|_r \leq Ch^{p(\frac{1}{c}-1)} = Cn^{\omega p(1-\frac{1}{c})}$  given  $h \propto n^{-\omega}$ . Thus, Condition (iii) holds by choosing  $c$  sufficiently small subject to the constraint of  $c > 1$ .

Finally, we verify Condition (iv). We first consider the case with  $t = 0$  and  $s \neq 0$ . We have, by the Cauchy-Schwarz inequality and change of variables,

$$\begin{aligned} E|h^p G_{ns}(Z_0, Z_0)|^2 &= 16E|h^{2p} \varepsilon_0^2 \varepsilon_0^2 E_0^2 [\bar{\varepsilon}_s \bar{\varepsilon}_0 W_h(\bar{X}_s, X_0) W_h(\bar{X}_0, X_0)]| \\ &\leq 16h^{2p} E|\varepsilon_0^4 (E_0 \bar{\varepsilon}_s^c \bar{\varepsilon}_0^c)^{\frac{2}{c}} [E_0 W_h^c(\bar{X}_s, X_0) W_h^c(\bar{X}_0, X_0)]^{\frac{2}{c}}| \\ &\leq 16h^{2p} (E|\varepsilon_0|^8)^{\frac{1}{2}} \left[ E|E_0^{\frac{4}{c}} W_h^c(\bar{X}_s, X_0) W_h^c(\bar{X}_0, X_0)| \right]^{\frac{1}{2}} \\ &= O\left( h^{2p} \left[ h^{2(1-c)p \frac{4}{c}} \right]^{\frac{1}{2}} \right) = O\left( h^{2(\frac{2}{c}-1)p} \right), \end{aligned}$$

where  $E_0[W_h^c(\bar{X}_s, X_0) W_h^c(\bar{X}_0, X_0)] = \int W_h^c(\bar{x}, X_0) W_h^c(\bar{x}_0, X_0) f_{\bar{X}_s \bar{X}_0}(\bar{x}, \bar{x}_0) d\bar{x} d\bar{x}_0 = O(h^{2(1-c)p})$  by change of variable. Thus, we have  $\|h^p G_{ns}(Z_0, Z_0)\|_2 \leq \left[ Ch^{2(\frac{2}{c}-1)p} \right]^{\frac{1}{2}} = Ch^{(\frac{2}{c}-1)p}$ , and so  $n^{\gamma_1} \|h^p G_{ns}(Z_0, Z_0)\|_2 = n^{(1-\frac{2}{c})\omega p + \gamma_1}$  if  $h = O(n^{-\omega})$ . Therefore, we obtain  $\|h^p G_{ns}(Z_0, Z_0)\|_2 = O(n^{-\gamma_1})$  with  $\gamma_1 = (\frac{c}{2} - 1)\omega p$ , if we choose  $c$  small enough for  $1 < c < 2$ . For the case with  $t \neq 0$  and  $s \neq 0$ , by a similar argument, we have  $\|h^p G_{ns}(Z_t, Z_0)\|_2 \leq \left[ Ch^{2(\frac{2}{c}-1)p} \right]^{\frac{1}{2}} = O\left( h^{(\frac{2}{c}-1)p} \right)$ . Thus, Condition (iv) holds with  $\gamma_1 = (\frac{c}{2} - 1)\omega p$ , provided we choose  $c$  small enough with  $1 < c < 2$ . Since all Conditions (i)–(iv) hold, we have  $V_q^{-1/2} h^{p/2} \hat{H}_q \xrightarrow{d} N(0, 1)$  by Tenreiro's (1997) central limit theorem. ■

**Proof of Theorem 3(ii):** We shall now derive the asymptotic distribution of  $\lambda_n$  under  $\mathbb{H}_n(a_n)$ . From Lemmas 3.5 and 3.6 and Propositions 3.3 and 3.4 below, we have under  $\mathbb{H}_n(a_n)$ ,

$$\left[ \lambda_n - h^{-p} \sigma^{-2} c(K) \int \sigma^2(x) dx \right] / \sqrt{2\sigma^{-4} d(K) \int \sigma^4(x) dx} \xrightarrow{d} N(\xi, 1).$$

**Lemma 3.5:** Under the conditions of Theorem 3,  $\lambda_n = \frac{n}{2} \frac{SSR_0 - SSR_1}{SSR_1} + o_p(h^{-p/2})$  under  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2} h^{-p/4}$ .

**Lemma 3.6:** Under the conditions of Theorem 3,  $\hat{\sigma}_n^2 \equiv n^{-1} SSR_1 = \sigma^2 + O_p(n^{-1/2})$  under  $\mathbb{H}_n(a_n)$

with  $a_n = n^{-1/2}h^{-p/4}$ .

**Proposition 3.3:** Let  $\widetilde{SSR}_0$  and  $\widetilde{SSR}_1$  be defined in the same way as  $SSR_0$  and  $SSR_1$  respectively, with  $\{\varepsilon_t\}_{t=1}^n$  replacing  $\{\hat{\varepsilon}_t\}_{t=1}^n$ . Then under the conditions of Theorem 3,  $SSR_0 - SSR_1 = \widetilde{SSR}_0 - \widetilde{SSR}_1 + h^{-p/2}E[\delta^2(X_t)] + o_p(h^{-p/2})$  under  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2}h^{-p/4}$ .

**Proposition 3.4:** Under the conditions of Theorem 3 and  $\mathbb{H}_n(a_n)$  with  $a_n = n^{-1/2}h^{-p/4}$ ,

$$\left[ \frac{\widetilde{SSR}_0 - \widetilde{SSR}_1}{2\sigma^2} - h^{-p}\sigma^{-2}c(K) \int \sigma^2(x)dx \right] / \sqrt{2\sigma^{-4}d(K) \int \sigma^4(x)dx} \xrightarrow{d} N(\xi, 1).$$

**Proof of Lemma 3.5:** Given in Supplementary Material.

**Proof of Lemma 3.6:** Given in Supplementary Material.

**Proof of Proposition 3.3:** Given in Supplementary Material.

**Proof of Proposition 3.4:** Proposition 3.4 follows from Lemmas 3.7 and Lemma 3.8 below.

**Lemma 3.7:** Put  $\hat{H}_\lambda = n^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} H_n(Z_t, Z_s)$ , where  $Z_t = (\varepsilon_t, X_t)'$ ,  $H_n(Z_t, Z_s) = \varepsilon_t \varepsilon_s W_h(X_t, X_s)$ , and

$$W_h(X_t, X_s) = \left[ \frac{1}{f(X_t)} + \frac{1}{f(X_s)} \right] \mathbf{K}_h(X_t - X_s) - \int \frac{\mathbf{K}_h(X_t - x)\mathbf{K}_h(X_s - x)}{f(x)} dx.$$

Suppose Assumptions A.1 and A.4 hold,  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ , and  $p < 4$ . Then

$$\widetilde{SSR}_0 - \widetilde{SSR}_1 = h^{-p} \left[ 2\mathbf{K}(0) - \int \mathbf{K}^2(\mathbf{u})d\mathbf{u} \right] \int \sigma^2(x)dx + \hat{H}_\lambda + o_p(h^{-p/2}).$$

**Lemma 3.8:** Suppose Assumptions A.1 and A.4 hold, and  $h \propto n^{-\omega}$  for  $\omega \in (0, 1/2p)$ . Define

$$V_\lambda = 2 \int \left[ \mathbf{K}(\mathbf{u}) - \frac{1}{2} \int \mathbf{K}(\mathbf{v})\mathbf{K}(\mathbf{u} + \mathbf{v})d\mathbf{v} \right]^2 du \int \sigma^4(x)dx.$$

Then  $V_\lambda^{-1/2}h^{p/2}\hat{H}_\lambda \xrightarrow{d} N(\xi, 1)$ .

**Proof of Lemma 3.7:** Given in Supplementary Material.

**Proof of Lemma 3.8:** Given in Supplementary Material.

**Proof of Theorem 4:** Pitman's asymptotic relative efficiency of the  $q_n$  test over the  $\lambda_n$  test is the limit of the ratio of the sample sizes required by the two tests to have the same asymptotic power at the same significance level, under the same local alternative (see Pitman 1979, Ch 7). Supposed  $n_1$  and  $n_2$  are the sample sizes required for the  $q_n$  and  $\lambda_n$  tests respectively. Then Pitman's asymptotic relative efficiency of  $q_n$  to  $\lambda_n$  is defined as

$$ARE(q_n : \lambda_n) = \lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{n_2} \tag{A.5}$$

under the condition that  $\lambda_n$  and  $q_n$  have the same asymptotic power under the same local alternatives  $n_1^{-\frac{1}{2}}h_1^{-\frac{p}{4}}\delta_1(x) \sim n_2^{-\frac{1}{2}}h_2^{-\frac{p}{4}}\delta_2(x)$  in the sense that

$$\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1^{-\frac{1}{2}}h_1^{-\frac{p}{4}}\delta_1(x)}{n_2^{-\frac{1}{2}}h_2^{-\frac{p}{4}}\delta_2(x)} = 1.$$

Given  $h_i = cn_i^{-\omega}$ ,  $i = 1, 2$ , we have  $n_1^{-2\gamma}E[\delta_1^2(X_t)] \sim n_2^{-2\gamma}E[\delta_2^2(X_t)]$ , where  $\gamma = \frac{2-\omega p}{4}$ . Hence,



$$\lim_{n_1, n_2 \rightarrow \infty} \left( \frac{n_1}{n_2} \right)^{2\gamma} = \frac{E[\delta_2^2(X_t)]}{E[\delta_1^2(X_t)]}. \quad (\text{A.6})$$

On the other hand, from Theorem 3(ii), we have

$$\frac{\gamma(K)\lambda_{n_1} - \mu_{n_1}}{\sqrt{2\mu_{n_1}}} \xrightarrow{d} N(\xi, 1),$$

under  $\mathbb{H}_{n_1}(a_{n_1}) : g_0(X_t) = g(X_t, \theta_0) + n_1^{-\frac{1}{2}} h_1^{-\frac{1}{4}} \delta_1(X_t)$ , where  $\xi = E[\delta_1^2(X_t)]/[2\sigma^{-2}\sqrt{2d(K)\int\sigma^4(x)dx}]$ . Also, from Theorem 3(i), we have

$$\frac{q_{n_2} - \nu_{n_2}}{\sqrt{2\nu_{n_2}}} \xrightarrow{d} N(\psi, 1)$$

under  $\mathbb{H}_{n_2}(a_{n_2}) : g_0(X_t) = g(X_t, \theta_0) + n_2^{-\frac{1}{2}} h_2^{-\frac{1}{4}} \delta_2(X_t)$ , where  $\psi = E[\delta_2^2(X_t)]/\sigma^{-2}\sqrt{2b(K)\int\sigma^4(x)dx}$ . To have the same asymptotic power, the noncentrality parameters must be equal; namely  $\xi = \psi$ , or

$$\frac{E[\delta_1^2(X_t)]}{2\sqrt{2d(K)\int\sigma^4(x)dx}} = \frac{E[\delta_2^2(X_t)]}{\sqrt{2b(K)\int\sigma^4(x)dx}}. \quad (\text{A.7})$$

Combining (A.5)–(A.7) yields

$$\begin{aligned} ARE(q_n, \lambda_n) &= \left[ \frac{2\sqrt{d(K)}}{\sqrt{b(K)}} \right]^{\frac{1}{2\gamma}} = \left[ \frac{4d(K)}{b(K)} \right]^{\frac{1}{4\gamma}} \\ &= \left[ \frac{\int (2\mathbf{K}(u) - \int \mathbf{K}(u)\mathbf{K}(u+v)du)^2 dv}{\int (\int \mathbf{K}(u)\mathbf{K}(u+v)du)^2 dv} \right]^{\frac{1}{2-\omega p}}. \end{aligned}$$

Finally, we show  $ARE(q_n : \lambda_n) \geq 1$  for any positive kernels with  $K(\cdot) \leq 1$ . For this purpose, it suffices to show

$$\int \left[ 2\mathbf{K}(u) - \int \mathbf{K}(u)\mathbf{K}(u+v)du \right]^2 dv \geq \int \left[ \int \mathbf{K}(u)\mathbf{K}(u+v)du \right]^2 dv$$

or equivalently,

$$\int \mathbf{K}^2(v)dv \geq \iint \mathbf{K}(u)\mathbf{K}(v)\mathbf{K}(u+v)dudv.$$

This last inequality follows from Zhang and Dette (2004, Lemma 2). This completes the proof. ■

**Acknowledgments.** We would like to thank the Editor, an Associate Editor and two anonymous referees for insightful comments and suggestions. We also thank participants of 2008 International Symposium on Recent Developments in Time Series Econometrics, Xiamen, China, and 2008 Symposium on Econometric Theory and Applications (SETA), Seoul, Korea, 2009 Time Series Conference, CIREQ, Montreal, Canada, Midwest Econometrics Group Meeting, Purdue, and seminar participants at Cornell University, Hong Kong University of Science and Technology, Korea University, Ohio State University, Texas A&M, and UC Riverside for their helpful comments and discussions.

## SUPPLEMENTARY MATERIAL

In this supplement, we present the detailed proofs of Theorems 1-4 and report the simulation results with the bandwidth  $h = S_X n^{-1/5}$ .

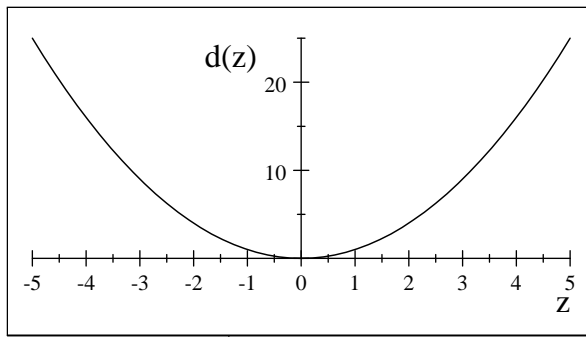
## REFERENCES

- Azzalini, A. and A. Bowman** (1990), “A Look at Some Data On the Old Faithful Geyser,” *Applied Statistics* 39, 357-365.
- Azzalini, A. and A. Bowman** (1993), “On the Use of Nonparametric Regression for Checking Linear Relationships,” *Journal of Royal Statistical Society, Series B* 55, 549-557.
- Azzalini, A., A. Bowman and W. Härdle** (1989), “On the Use of Nonparametric Regression for Model Checking,” *Biometrika* 76, 1-11.
- Bahadur, R.R.** (1958), “Examples of Inconsistency of Maximum Likelihood Estimates”, *Sankhya* 20, 207-210.
- Bentkus, V., F. Gotse and A. Tikhomirov** (1997), “Berry-Esseen Bounds for Statistics of Weakly Dependent Samples,” *Bernoulli* 3, 329-349.
- Cai, Z., J. Fan and Q. Yao** (2000), “Functional-Coefficient Regression Models for Nonlinear Time Series,” *Journal of American Statistical Association* 95, 941-956.
- Christoffersen, P.F. and F.X. Diebold** (1997), “Optimal Prediction under Asymmetric Loss,” *Econometric Theory* 13, 808-817.
- Doukan, P.** (1995), *Mixing: Properties and Examples*, Springer-Verlag.
- Engle, R.** (1982), “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica* 50, 987-1007.
- Fan, J. and T. Huang** (2005), “Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models,” *Bernoulli* 11, 1031-1057.
- Fan, J. and J. Jiang** (2005), “Nonparametric Inference for Additive Models,” *Journal of American statistical Association* 100, 890-907.
- Fan, J. and J. Jiang** (2007), “Nonparametric Inference with Generalized Likelihood Ratio Test,” *Test* 16, 409-444.
- Fan, Y. and Q. Li** (2002), “A Consistent Model Specification Test Based on the Kernel Sum of Squares of Residuals,” *Econometric Reviews* 21, 337-352.
- Fan, J. and C. Zhang** (2003), “A Reexamination of Diffusion Estimators With. Applications to Financial Model Validation,” *Journal of American statistical Association* 98, 118-134.
- Fan, J. and W. Zhang** (2004), “Generalized Likelihood Ratio Tests for Sepctral Density,” *Biometrika* 91, 195-209.
- Fan, J. and Q. Yao** (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag.
- Fan, J., C. Zhang and J. Zhang** (2001), “Generalized Likelihood Ratio Statistics and Wilks Phenomenon,” *Annals of Statistics* 29, 153-193.
- Fan, J. and J. Zhang** (2004), “Sieve Empirical Likelihood Ratio Tests for Nonparametric Functions,” *Annals of Statistics* 32, 1858-1907.
- Franke, J., J.-P. Kreiss and E. Mammen** (2002), “Bootstrap of Kenel Smoothing in Nonlinear Time Series,” *Bernoulli* 8, 1-37.

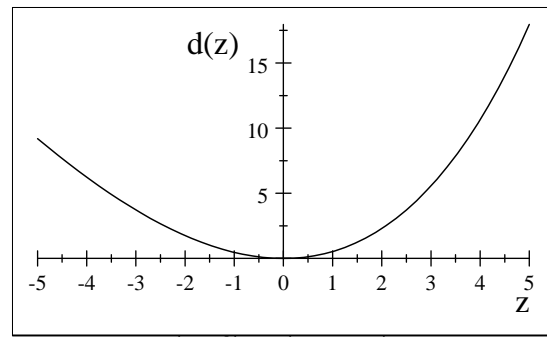
- Gao, J. and Gijbels** (2008), “Bandwidth Selection in Nonparametric Kernel Testing,” *Journal of American Statistical Association* 103, 1584-1594.
- Giacomini, R. and H. White** (2006), “Tests of Conditional Predictive Ability,” *Econometrica* 74, 1545-1578.
- Granger, C.W.J. and M.H. Pesaran** (1999), “Economic and Statistical Measures of Forecast Accuracy,” Cambridge Working Papers in Economics 9910, Faculty of Economics, University of Cambridge.
- Granger, C.W.J. and T. Teräsvirta** (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press: New York.
- Härdle, W.** (1990), *Applied Nonparametric Regression*, Cambridge.
- Härdle, W. and E. Mammen** (1993), “Comparing Nonparametric Versus Parametric Regression Fits,” *Annals of Statistics* 21, 1926-1947.
- Hjellvik, V. and D. Tjøstheim** (1996), “Nonparametric Statistics for Testing of Linearity and Serial Independence,” *Journal of Nonparametric Statistics* 6, 223-251.
- Hong, Y. and Y. Lee** (2005), “Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroskedasticity of Unknown Form,” *Review of Economic Studies* 72, 499-541.
- Hong, Y. and Y. Lee** (2012), Supplement to “A Loss Function Approach to Model Specification Testing and Its Relative Efficiency,” DOI:
- Hong, Y. and H. White** (1995), “Consistent Specification Testing Via Nonparametric Series Regression,” *Econometrica* 63, 1133-1159.
- Horowitz, J. and V.G. Spokoiny** (2001), “An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model against a Nonparametric Alternative,” *Econometrica* 69, 599-631.
- Ingster, Y.I.** (1993a), “Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives I-III,” *Mathematical Methods of Statistics* 2, 85-114.
- Ingster, Y.I.** (1993b), “Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives I-III,” *Mathematical Methods of Statistics* 3, 1715-189.
- Ingster, Y.I.** (1993c), “Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives I-III,” *Mathematical Methods of Statistics* 4, 249-268.
- Le Cam, L.** (1979), *Maximum Likelihood: An Introduction*, Lecture Notes in Statistics, No. 18, University of Maryland, College Park, Md.
- Lee, T.-H., H. White and C.W.J. Granger** (1993), “Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests,” *Journal of Econometrics* 56, 269-290.
- Lepski O.V. and V.G. Spokoiny** (1999), “Minimax Nonparametric Hypothesis Testing: The Case of An Inhomogeneous Alternative,” *Bernoulli* 5, 333-358.
- Li, Q. and J. Racine** (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- Pagan, A. and A. Ullah** (1999), *Nonparametric Econometrics*, Cambridge.

- Pan, J., Q. Yao and H. Wang** (2000), "Weighted least absolute deviations estimation for ARMA models with infinite variance," *Econometric Theory* 23, 852-879.
- Peel, D.A., and A.R. Nobay** (1998), "Optimal Monetary Policy in A Model of Asymmetric Central Bank Preferences," *FMG Discussion Paper* 0306.
- Phillips, P. C.** (1996), "Econometric Model Determination," *Econometrica* 64, 763-812.
- Pitman, E. J. G.** (1979), *Some basic theory for statistical inference*, Chapman and Hall: New York.
- Robinson, P.** (1991), "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies* 58, 437-453.
- Serfling, R. J.** (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons.
- Sun, Phillips, and Jin** (2008), "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica* 76, 175-194.
- Tenreiro, C.** (1997), "Loi Asymptotique Des Erreurs Quadratiques Intégrées des Estimateurs à Noyau De La Densité et de la Regression, Sou des Conditions de Dépendance," *Portugaliae Math.* 54, 187-213.
- Van Der Vaart, A.W.** (1998), *Asymptotic Statistics*, Cambridge University Press.
- Varian, H.** (1975), "A Bayesian Approach to Real Estate Assessment," *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J.Savage*, eds. Stephen E. Fienberg and Arnold Zellner, Amsterdam: North-Holland, 195-208.
- Vuong, Q.** (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica* 57, 307-333.
- Weiss, A.A.** (1996), "Estimating Time Series Models Using the Relevant Cost Function," *Journal of Applied Econometrics* 11, 539-560.
- White, H.** (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-25.
- Zhang, C. and H. Dette** (2004), "A Power Comparison Between Nonparametric Regression Tests," *Statistics and Probability Letters* 66, 289-301.
- Zellner, A.** (1986), "Bayesian Estimation and Prediction Using Asymmetric Loss Functions," *Journal of American Statistical Association* 81, 446-451.

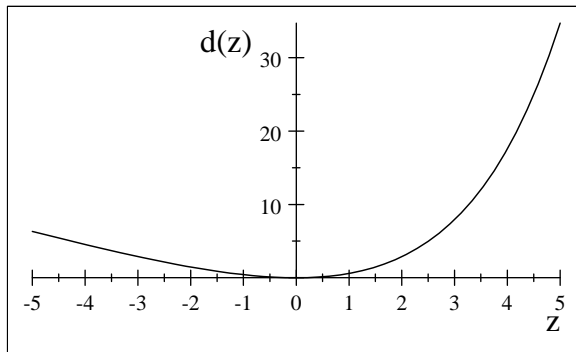
**FIGURE 3.1: The LINEX Loss Function**  $d(z) = \frac{\beta}{\alpha^2}[\exp(\alpha z) - (1 + \alpha z)]$ .



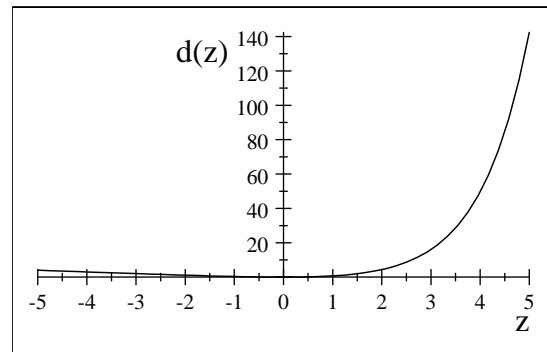
$\alpha^+ \rightarrow 0, \beta = 1.0$



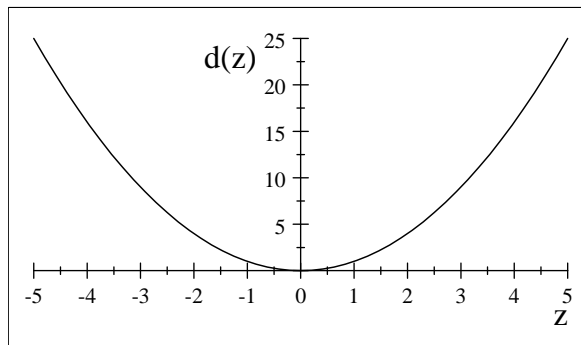
$(\alpha, \beta) = (0.2, 1.0)$



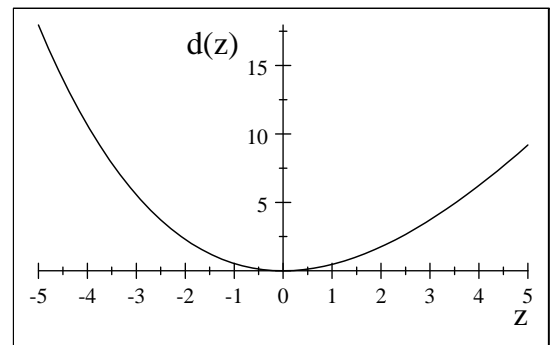
$(\alpha, \beta) = (0.5, 1.0)$



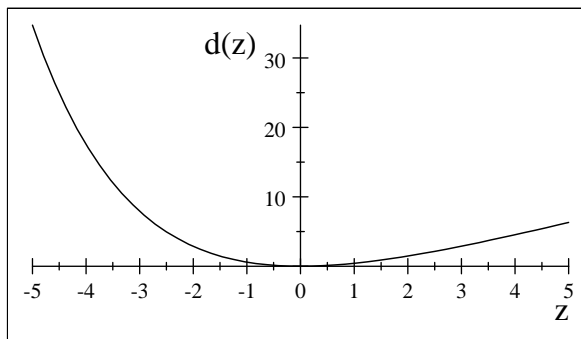
$(\alpha, \beta) = (1.0, 1.0)$



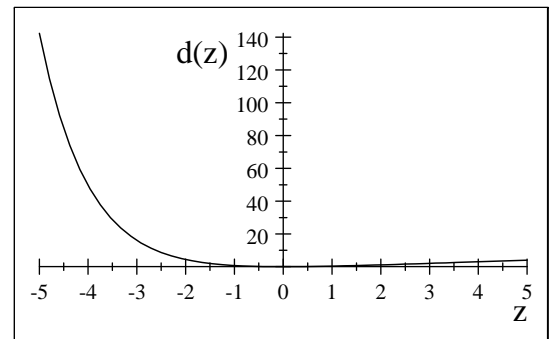
$\alpha^- \rightarrow 0, \beta = 1.0$



$(\alpha, \beta) = (-0.2, 1.0)$



$(\alpha, \beta) = (-0.5, 1.0)$



$(\alpha, \beta) = (-1.0, 1.0)$

TABLE 6.1. Empirical Sizes of Tests Using Asymptotic Critical Values

$(a, \beta)$	$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$	
	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	$n = 100$						$n = 250$						$n = 500$					
	DGP S.1: <i>i.i.d.</i> Normal Errors																	
(0.0,1.0)	7.7	5.0	4.5	2.9	7.4	5.1	6.0	3.6	4.6	2.6	7.3	4.5	6.8	4.8	6.3	4.0	7.2	3.8
(0.2,1.0)	8.1	5.2	4.5	3.0	7.4	5.1	6.1	3.8	4.7	2.5	7.3	4.5	6.6	4.9	6.2	4.1	7.2	3.8
(0.5,1.0)	8.4	5.2	4.9	3.5	7.4	5.1	6.6	3.7	4.8	2.6	7.3	4.5	6.6	4.8	6.1	4.5	7.2	3.8
(1.0,1.0)	9.4	5.8	5.7	4.3	7.4	5.1	6.8	4.3	5.2	3.0	7.3	4.5	6.9	5.0	6.0	4.6	7.2	3.8
	DGP S.2: <i>i.i.d.</i> Student- $t_5$ Errors																	
(0.0,1.0)	6.8	4.3	4.0	2.5	6.4	3.4	6.0	3.3	4.4	2.4	4.9	2.7	5.4	3.1	4.3	2.4	7.7	4.5
(0.2,1.0)	6.7	4.3	4.1	2.6	6.4	3.4	5.8	3.8	4.4	2.4	4.9	2.7	5.4	3.2	4.6	2.5	7.7	4.5
(0.5,1.0)	6.9	4.5	4.3	2.6	6.4	3.4	5.9	3.9	4.3	2.7	4.9	2.7	5.6	3.3	4.8	2.6	7.7	4.5
(1.0,1.0)	8.5	5.1	5.0	3.0	6.4	3.4	6.4	4.4	5.0	3.4	4.9	2.7	6.2	3.5	5.1	3.0	7.7	4.5
	DGP S.3: <i>i.i.d.</i> Uniform Errors																	
(0.0,1.0)	7.1	5.2	4.3	2.7	6.2	3.5	6.8	4.2	5.5	2.9	6.6	4.0	6.4	4.2	5.6	3.5	6.3	3.2
(0.2,1.0)	7.1	5.4	4.5	2.5	6.2	3.5	6.8	4.4	5.4	2.9	6.6	4.0	6.2	4.2	5.3	3.5	6.3	3.2
(0.5,1.0)	7.4	5.4	4.9	2.7	6.2	3.5	7.0	4.5	5.6	3.1	6.6	4.0	6.2	4.5	5.5	3.6	6.3	3.2
(1.0,1.0)	9.0	6.0	5.7	3.7	6.2	3.5	7.1	5.1	5.9	3.4	6.6	4.0	6.7	4.6	6.1	3.7	6.3	3.2
	DGP S.4: <i>i.i.d.</i> Log-Normal Errors																	
(0.0,1.0)	9.4	7.2	6.8	4.3	8.5	6.6	6.1	4.2	4.8	2.8	6.7	3.6	6.9	5.1	6.3	4.4	6.9	4.1
(0.2,1.0)	9.9	7.9	7.6	5.2	8.5	6.6	6.5	4.7	4.9	3.3	6.7	3.6	7.5	5.6	6.6	4.5	6.9	4.1
(0.5,1.0)	10.4	8.7	8.2	6.5	8.5	6.6	8.1	4.8	5.9	4.3	6.7	3.6	7.9	5.9	7.1	5.3	6.9	4.1
(1.0,1.0)	12.4	10.2	9.5	7.9	8.5	6.6	9.2	6.9	7.6	5.5	6.7	3.6	9.5	7.1	8.3	6.2	6.9	4.1
	DGP S.5: <i>i.i.d.</i> Chi-Square Errors																	
(0.0,1.0)	7.6	5.9	5.6	3.5	7.3	5.2	6.3	4.0	4.6	2.8	6.3	3.1	5.2	3.5	4.4	2.8	5.4	2.9
(0.2,1.0)	8.1	6.3	6.0	3.8	7.3	5.2	6.5	4.2	5.1	3.0	6.3	3.1	5.2	3.7	4.9	3.1	5.4	2.9
(0.5,1.0)	8.8	7.2	7.0	4.6	7.3	5.2	7.0	4.8	5.6	3.7	6.3	3.1	5.5	3.9	5.3	3.3	5.4	2.9
(1.0,1.0)	10.8	9.2	8.9	6.2	7.3	5.2	7.8	5.6	6.3	4.8	6.3	3.1	6.3	4.6	5.9	4.1	5.4	2.9

Notes : (i) 1000 iterations;

(ii)  $GLR$ , the generalized likelihood ratio test,  $q_n$  and  $q_n^0$ , loss function-based tests;

(iii)  $q_n$  is standardized by  $SSR_1$ , the sum of squared residuals of the nonparametric regression estimates, and  $q_n^0$  is standardized by  $SSR_0$ , the sum of squared residuals of the null linear model;

(iv) The uniform kernel is used for  $GLR$ ,  $q_n$  and  $q_n^0$ ; the bandwidth  $h = S_X n^{-2/9}$ , where  $S_X$  is the sample standard deviation of  $\{X_t\}_{t=1}^n$ ;

(v) The  $q_n$  tests are based on the linex loss function:  $d(z) = \frac{\beta}{\alpha^2}[\exp(\alpha z) - 1 - \alpha z]$ ;

(vi)  $Y_t = 1 + X_t + \varepsilon_t$ ,  $X_t = 0.5X_{t-1} + v_t$ ,  $v_t \sim i.i.d. N(0, 1)$ , where DGP S.1:  $\varepsilon_i \sim i.i.d. N(0, 1)$ ; DGP S.2:  $\varepsilon_i \sim i.i.d. \text{Student-}t_5$ ; DGP S.3:  $\varepsilon_i \sim i.i.d. U[0, 1]$ ; DGP S.4:  $\varepsilon_i \sim i.i.d. \log N(0, 1)$ ; DGP S.5:  $\varepsilon_i \sim i.i.d. \chi_1^2$ .

TABLE 6.2. Empirical Sizes of Tests Using Bootstrap Critical Values

$(a, \beta)$	$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$	
	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
$n = 100$						$n = 250$						$n = 500$						
DGP S.1: <i>i.i.d.</i> Normal Errors																		
(0.0,1.0)	10.0	5.4	10.0	5.3	10.4	4.4	11.1	6.1	11.8	5.9	11.5	5.0	11.4	6.1	11.6	6.1	10.9	6.5
(0.2,1.0)	10.0	5.4	10.0	5.6	10.4	4.4	11.8	6.2	12.1	6.3	11.5	5.0	11.4	6.1	11.4	6.0	10.9	6.5
(0.5,1.0)	9.4	4.9	9.7	4.9	10.4	4.4	11.4	5.9	11.7	5.8	11.5	5.0	11.4	6.4	11.6	6.2	10.9	6.5
(1.0,1.0)	9.2	4.6	9.7	4.5	10.4	4.4	11.3	5.8	11.2	5.5	11.5	5.0	11.7	6.4	12.2	6.1	10.9	6.5
DGP S.2: <i>i.i.d.</i> Student- $t_5$ Errors																		
(0.0,1.0)	9.1	4.1	8.8	4.4	9.5	3.8	8.8	4.3	8.9	4.5	9.3	3.9	10.1	4.7	10.7	5.2	11.5	5.7
(0.2,1.0)	8.8	4.2	8.7	4.2	9.5	3.8	8.8	4.3	9.0	4.6	9.3	3.9	10.6	5.0	10.8	4.8	11.5	5.7
(0.5,1.0)	9.4	4.0	8.7	4.4	9.5	3.8	9.1	4.5	8.8	4.5	9.3	3.9	10.5	5.1	10.3	5.0	11.5	5.7
(1.0,1.0)	9.8	3.9	10.2	4.5	9.5	3.8	8.8	5.0	9.2	4.7	9.3	3.9	10.6	5.4	10.5	5.2	11.5	5.7
DGP S.3: <i>i.i.d.</i> Uniform Errors																		
(0.0,1.0)	10.3	5.0	10.1	5.3	9.1	4.2	11.1	5.8	11.2	5.8	10.9	6.5	9.6	6.0	9.5	6.1	10.6	5.3
(0.2,1.0)	10.3	5.1	10.3	5.5	9.1	4.2	11.1	5.8	10.9	5.7	10.9	6.5	9.6	6.0	9.5	6.0	10.6	5.3
(0.5,1.0)	10.8	5.1	10.7	5.4	9.1	4.2	11.2	5.6	11.2	5.7	10.9	6.5	9.4	6.0	9.4	6.0	10.6	5.3
(1.0,1.0)	10.3	5.5	10.6	5.2	9.1	4.2	10.7	5.7	11.0	5.8	10.9	6.5	9.5	6.0	9.4	6.2	10.6	5.3
DGP S.4: <i>i.i.d.</i> Log-Normal Errors																		
(0.0,1.0)	11.0	5.8	10.9	6.1	10.2	5.5	9.7	4.6	9.6	5.1	10.8	4.5	10.3	5.3	10.4	5.4	11.0	5.9
(0.2,1.0)	10.7	5.9	10.6	6.0	10.2	5.5	10.0	4.7	9.5	4.9	10.8	4.5	9.9	5.3	10.3	5.4	11.0	5.9
(0.5,1.0)	10.9	5.8	10.8	5.6	10.2	5.5	9.5	4.7	9.4	4.7	10.8	4.5	9.9	5.6	10.0	5.4	11.0	5.9
(1.0,1.0)	10.7	5.8	11.0	5.9	10.2	5.5	9.7	4.5	9.5	4.4	10.8	4.5	10.1	5.6	10.4	5.4	11.0	5.9
DGP S.5: <i>i.i.d.</i> Chi-Square Errors																		
(0.0,1.0)	9.4	4.4	9.2	4.4	9.7	5.3	10.2	4.8	10.2	4.9	9.3	4.6	7.9	3.7	8.2	3.7	9.0	4.1
(0.2,1.0)	9.3	4.5	9.0	4.3	9.7	5.3	10.2	4.9	10.0	4.9	9.3	4.6	7.7	3.7	8.0	3.7	9.0	4.1
(0.5,1.0)	9.2	4.9	9.1	4.8	9.7	5.3	10.2	4.7	10.2	4.9	9.3	4.6	7.8	3.7	7.9	3.7	9.0	4.1
(1.0,1.0)	8.7	5.2	8.5	4.8	9.7	5.3	10.3	3.7	9.9	3.6	9.3	4.6	7.8	3.6	7.7	3.7	9.0	4.1

Notes : (i) 1000 iterations;

(ii)  $GLR$ , the generalized likelihood ratio test,  $q_n$  and  $q_n^0$ , loss function-based tests;

(iii)  $q_n$  is standardized by  $SSR_1$ , the sum of squared residuals of the nonparametric regression estimates, and  $q_n^0$  is standardized by  $SSR_0$ , the sum of squared residuals of the null linear model;

(iv) The uniform kernel is used for  $GLR$ ,  $q_n$  and  $q_n^0$ ; the bandwidth  $h = S_X n^{-2/9}$ , where  $S_X$  is the sample standard deviation of  $\{X_t\}_{t=1}^n$ ;

(v) The  $q_n$  tests are based on the linex loss function:  $d(z) = \frac{\beta}{\alpha^2}[\exp(\alpha z) - 1 - \alpha z]$ ;

(vi)  $Y_t = 1 + X_t + \varepsilon_t$ ,  $X_t = 0.5X_{t-1} + v_t$ ,  $v_t \sim i.i.d. N(0, 1)$ , where DGP S.1:  $\varepsilon_i \sim i.i.d. N(0, 1)$ ; DGP S.2:  $\varepsilon_i \sim i.i.d.$  Student- $t_5$ ; DGP S.3:  $\varepsilon_i \sim i.i.d. U[0, 1]$ ; DGP S.4:  $\varepsilon_i \sim i.i.d. \log N(0, 1)$ ; DGP S.5:  $\varepsilon_i \sim i.i.d. \chi_1^2$ .

**TABLE 6.3. Empirical Powers of Tests Using Bootstrap Critical Values**

		DGP P.1: Quadratic Regression																	
$(a, \beta)$	$\theta$	$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
		$n = 100$						$n = 250$						$n = 500$					
(0.0,1.0)	0.1	22.6	12.8	23.0	12.4	20.2	12.4	38.4	27.2	39.2	27.8	29.4	20.4	62.0	50.6	62.2	50.8	47.4	33.6
	0.2	53.2	39.8	53.6	39.6	45.2	34.8	90.6	83.4	91.0	83.4	80.4	70.6	99.4	99.0	99.4	99.0	97.4	95.6
	0.3	85.2	77.0	85.2	77.2	76.8	65.8	99.6	99.2	99.6	99.2	98.8	97.2	100.0	100.0	100.0	100.0	100.0	100.0
	0.5	99.4	98.4	99.4	98.8	98.6	96.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
(0.2,1.0)	0.1	23.4	13.6	23.6	13.0	20.2	12.4	39.8	28.0	40.0	28.4	29.4	20.4	63.8	51.8	64.0	52.0	47.4	33.6
	0.2	55.8	41.2	55.8	41.0	45.2	34.8	91.0	84.0	91.4	83.8	80.4	70.6	99.4	99.2	99.4	99.2	97.4	95.6
	0.3	86.2	78.4	86.2	78.6	76.8	65.8	99.8	99.2	99.6	99.2	98.8	97.2	100.0	100.0	100.0	100.0	100.0	100.0
	0.5	99.4	98.6	99.4	99.2	98.6	96.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
(0.5,1.0)	0.1	24.8	14.4	24.2	14.6	20.2	12.4	42.6	30.0	41.4	29.2	29.4	20.4	64.8	53.4	64.6	53.6	47.4	33.6
	0.2	58.0	43.8	59.0	43.4	45.2	34.8	91.4	85.2	92.0	85.8	80.4	70.6	99.6	99.4	99.6	99.4	97.4	95.6
	0.3	87.0	80.0	86.6	80.2	76.8	65.8	99.8	99.4	99.6	99.4	98.8	97.2	100.0	100.0	100.0	100.0	100.0	100.0
	0.5	99.4	98.8	99.4	99.2	98.6	96.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
(1.0,1.0)	0.1	27.6	16.2	26.2	18.0	20.2	12.4	44.6	32.4	45.2	31.4	29.4	20.4	66.4	66.0	66.0	56.6	47.4	33.6
	0.2	60.8	46.4	62.2	46.4	45.2	34.8	92.2	87.0	92.6	88.2	80.4	70.6	99.6	99.4	99.8	99.4	97.4	95.6
	0.3	88.8	81.6	88.4	80.6	76.8	65.8	99.8	99.6	99.6	97.2	98.8	97.2	100.0	100.0	100.0	100.0	100.0	100.0
	0.5	99.6	99.0	99.6	99.2	98.6	96.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	1.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes : (i) 1000 iterations;

(ii)  $GLR$ , the generalized likelihood ratio test,  $q_n$  and  $q_n^0$ , loss function-based tests;

(iii)  $q_n$  is standardized by  $SSR_1$ , the sum of squared residuals of the nonparametric regression estimates, and  $q_n^0$  is standardized by  $SSR_0$ , the sum of squared residuals of the null linear model;

(iv) The uniform kernel is used for  $GLR$ ,  $q_n$  and  $q_n^0$ ; the bandwidth  $h = S_X n^{-2/9}$ , where  $S_X$  is the sample standard deviation of  $\{X_t\}_{t=1}^n$ ;

(v) The  $q_n$  tests are based on the linex loss function:  $d(z) = \frac{\beta}{\alpha^2} [\exp(\alpha z) - 1 - \alpha z]$ ;

(vi) DGP P.1,  $Y_t = 1 + X_t + \theta X_t^2 + \varepsilon_t$ , where  $\{\varepsilon_i\} \sim i.i.d.N(0, 1)$ .



TABLE 6.4. Empirical Powers of Tests Using Bootstrap Critical Values Using Bootstrap Critical Values

		DGP P.2: Threshold Regression																	
		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$	
$(a, \beta)$	$\theta$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
		$n = 100$						$n = 250$						$n = 500$					
(0.0,1.0)	-1.0	73.0	60.2	73.2	61.4	60.8	46.4	98.8	96.2	98.8	96.4	95.0	90.8	100.0	100.0	100.0	100.0	100.0	100.0
	-0.5	28.0	18.0	28.4	18.4	24.2	15.6	51.8	38.2	52.6	39.8	39.2	28.6	81.6	69.0	81.8	69.4	62.6	51.6
	-0.2	13.4	7.2	13.4	7.2	13.0	7.8	17.0	8.8	16.4	8.8	13.4	6.8	24.8	15.4	25.2	15.6	18.6	12.2
	0.2	11.0	5.8	11.4	6.4	12.6	5.8	14.6	8.0	15.0	8.4	12.6	7.2	23.2	14.6	23.4	14.4	18.6	11.0
	0.5	25.0	15.6	25.2	14.4	21.0	11.8	50.4	36.8	50.4	37.0	37.2	25.4	79.2	69.6	79.0	69.4	64.8	51.8
	1.0	71.4	59.0	72.2	58.6	60.2	43.4	97.6	94.6	97.6	94.6	92.6	87.2	100.0	100.0	100.0	100.0	100.0	100.0
(0.2,1.0)	-1.0	74.0	61.6	74.8	62.2	60.8	46.4	99.0	96.8	98.8	96.8	95.0	90.8	100.0	100.0	100.0	100.0	100.0	100.0
	-0.5	29.4	18.2	29.8	18.6	24.2	15.6	53.0	40.2	53.6	39.8	39.2	28.6	81.0	70.4	81.8	69.8	62.6	51.6
	-0.2	14.0	7.4	13.8	7.2	13.0	7.8	17.6	8.4	17.2	8.2	13.4	6.8	25.2	15.2	25.2	15.6	18.6	12.2
	0.2	10.8	6.2	11.2	6.0	12.6	5.8	14.0	7.6	14.2	7.8	12.6	7.2	23.4	14.8	23.4	14.2	18.6	11.0
	0.5	24.0	13.8	23.8	12.8	21.0	11.8	49.0	36.6	49.2	36.2	37.2	25.4	78.6	68.6	78.4	69.0	64.8	51.8
	1.0	70.2	56.8	70.4	56.4	60.2	43.4	97.2	94.2	97.4	94.6	92.6	87.2	100.0	100.0	100.0	100.0	100.0	100.0
(0.5,1.0)	-1.0	76.2	62.2	76.8	63.0	60.8	46.4	99.0	96.8	98.8	97.0	95.0	90.8	100.0	100.0	100.0	100.0	100.0	100.0
	-0.5	30.4	19.2	31.2	19.8	24.2	15.6	55.0	41.2	56.4	40.4	39.2	28.6	81.4	71.6	81.8	71.4	62.6	51.6
	-0.2	14.0	7.8	14.0	8.2	13.0	7.8	18.0	9.4	18.4	10.0	13.4	6.8	26.2	16.2	25.6	15.8	18.6	12.2
	0.2	9.4	6.0	9.6	5.8	12.6	5.8	13.0	6.8	13.8	7.0	12.6	7.2	23.0	14.0	22.4	14.0	18.6	11.0
	0.5	21.4	12.4	21.0	12.0	21.0	11.8	48.8	35.2	48.6	35.0	37.2	25.4	78.4	66.8	78.0	66.8	64.8	51.8
	1.0	68.8	54.6	67.8	52.8	60.2	43.4	96.8	94.0	97.0	94.0	92.6	87.2	100.0	100.0	100.0	100.0	100.0	100.0
(1.0,1.0)	-1.0	77.0	63.6	77.8	64.2	60.8	46.4	99.0	96.8	99.0	96.8	95.0	90.8	100.0	100.0	100.0	100.0	100.0	100.0
	-0.5	31.2	21.4	32.4	22.0	24.2	15.6	57.6	44.0	57.8	44.0	39.2	28.6	82.6	74.2	83.2	73.8	62.6	51.6
	-0.2	14.0	8.0	14.8	8.2	13.0	7.8	18.6	11.0	19.6	10.8	13.4	6.8	26.0	16.0	26.6	17.0	18.6	12.2
	0.2	9.2	6.0	9.4	6.0	12.6	5.8	12.2	6.4	12.4	6.8	12.6	7.2	21.4	12.6	21.0	13.0	18.6	11.0
	0.5	19.2	10.0	18.4	9.6	21.0	11.8	47.2	31.6	46.8	32.2	37.2	25.4	76.4	64.0	75.8	65.2	64.8	51.8
	1.0	63.8	48.2	62.2	46.2	60.2	43.4	95.8	92.8	96.0	93.0	92.6	87.2	100.0	100.0	100.0	100.0	100.0	100.0

Notes : (i) 1000 iterations;

(ii)  $GLR$ , the generalized likelihood ratio test,  $q_n$  and  $q_n^0$ , loss function-based tests;

(iii)  $q_n$  is standardized by  $SSR_1$ , the sum of squared residuals of the nonparametric regression estimates, and  $q_n^0$  is standardized by  $SSR_0$ , the sum of squared residuals of the null linear model;

(iv) The uniform kernel is used for  $GLR$ ,  $q_n$  and  $q_n^0$ ; the bandwidth  $h = S_X n^{-2/9}$ , where  $S_X$  is the sample standard deviation of  $\{X_t\}_{t=1}^n$ ;

(v) The  $q_n$  tests are based on the linex loss function:  $d(z) = \frac{\beta}{\alpha^2} [\exp(\alpha z) - 1 - \alpha z]$ ;

(vi) DGP P.2,  $Y_t = 1 + X_t 1(X_t > 0) + (1 + \theta) X_t 1(X_t \leq 0) + \varepsilon_t$ , where  $\{\varepsilon_i\} \sim i.i.d.N(0, 1)$ .

**TABLE 6.5. Empirical Powers of Tests**

		DGP P.3: Smooth Transition Regression																	
		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$		$q_n$		$q_n^0$		$GLR$	
$(a, \beta)$	$\theta$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
		$n = 100$						$n = 250$						$n = 500$					
(0.0,1.0)	-1.0	53.2	40.0	54.0	41.6	43.8	33.6	90.2	84.6	90.2	84.2	78.8	69.8	99.6	99.2	99.6	99.2	97.8	95.8
	-0.5	22.6	13.0	22.8	13.8	20.8	11.8	37.8	27.2	38.8	27.4	27.8	20.4	61.2	49.2	60.2	49.8	47.6	35.2
	0.5	20.6	10.4	19.8	11.4	16.6	9.6	36.2	25.0	35.6	26.2	29.2	17.0	60.8	49.4	60.2	50.6	46.8	35.0
	1.0	52.2	38.2	51.4	37.6	44.0	29.8	87.4	79.6	86.8	79.8	78.2	69.0	99.4	98.6	99.4	98.6	98.4	96.0
	1.5	85.6	75.2	86.6	77.0	75.6	65.6	99.6	99.4	99.6	99.2	98.2	96.8	100.0	100.0	100.0	100.0	100.0	100.0
(0.2,1.0)	-1.0	55.0	40.4	56.0	42.6	43.8	33.6	90.6	85.0	90.4	85.6	78.8	69.8	99.6	99.2	99.6	99.2	97.8	95.8
	-0.5	23.8	14.0	23.0	14.2	20.8	11.8	39.4	28.4	40.2	28.2	27.8	20.4	61.6	49.8	60.8	51.0	47.6	35.2
	0.5	19.4	10.2	19.4	10.4	16.6	9.6	35.4	24.4	35.2	25.2	29.2	17.0	59.2	48.8	59.0	49.2	46.8	35.0
	1.0	50.4	37.2	49.8	36.2	44.0	29.8	86.8	79.0	85.6	78.4	78.2	69.0	99.4	98.6	99.4	98.6	98.4	96.0
	1.5	84.6	74.2	85.0	75.2	75.6	65.6	99.4	99.4	99.6	99.0	98.2	96.8	100.0	100.0	100.0	100.0	100.0	100.0
(0.5,1.0)	-1.0	56.6	43.0	57.6	44.0	43.8	33.6	86.2	63.6	86.0	64.6	78.8	69.8	99.6	99.2	99.6	99.2	97.8	95.8
	-0.5	25.0	14.4	24.0	14.4	20.8	11.8	40.4	29.4	40.6	29.2	27.8	20.4	62.4	51.4	61.4	51.8	47.6	35.2
	0.5	18.4	9.4	18.4	8.8	16.6	9.6	34.6	23.0	34.0	23.2	29.2	17.0	58.0	47.4	57.8	48.0	46.8	35.0
	1.0	48.8	34.0	47.6	33.2	44.0	29.8	85.0	77.6	85.0	77.8	78.2	69.0	99.2	98.2	99.4	98.4	98.4	96.0
	1.5	83.0	70.2	82.6	72.0	75.6	65.6	99.4	98.8	99.6	99.0	98.2	96.8	100.0	100.0	100.0	100.0	100.0	100.0
(1.0,1.0)	-1.0	58.4	46.4	58.8	46.0	43.8	33.6	92.0	87.4	91.8	86.8	78.8	69.8	99.6	99.2	99.6	99.2	97.8	95.8
	-0.5	26.2	16.2	26.2	16.4	20.8	11.8	43.6	30.8	43.8	30.6	27.8	20.4	64.6	52.4	64.0	52.2	47.6	35.2
	0.5	14.6	7.4	15.0	7.6	16.6	9.6	31.0	20.2	31.8	20.6	29.2	17.0	56.8	44.4	56.6	45.2	46.8	35.0
	1.0	44.0	27.8	43.2	29.0	44.0	29.8	83.0	75.4	82.8	75.8	78.2	69.0	99.2	97.6	99.2	97.8	98.4	96.0
	1.5	78.2	62.8	79.4	64.8	75.6	65.6	99.4	98.4	99.4	98.6	98.2	96.8	100.0	100.0	100.0	100.0	100.0	100.0

Notes : (i) 1000 iterations;

(ii)  $GLR$ , the generalized likelihood ratio test,  $q_n$  and  $q_n^0$ , loss function-based tests;

(iii)  $q_n$  is standardized by  $SSR_1$ , the sum of squared residuals of the nonparametric regression estimates, and  $q_n^0$  is standardized by  $SSR_0$ , the sum of squared residuals of the null linear model;

(iv) The uniform kernel is used for  $GLR$ ,  $q_n$  and  $q_n^0$ ; the bandwidth  $h = S_X n^{-2/9}$ , where  $S_X$  is the sample standard deviation of  $\{X_t\}_{t=1}^n$ ;

(v) The  $q_n$  tests are based on the linex loss function:  $d(z) = \frac{\beta}{\alpha^2} [\exp(\alpha z) - 1 - \alpha z]$ ;

(vi) DGP P.3,  $Y_t = 1 + X_t + [1 - F(X_t)\theta]X_t + \varepsilon_t$ ,  $F(X_t) = \frac{1}{1 + \exp(-X_t)}$ , where  $\{\varepsilon_i\} \sim i.i.d.N(0, 1)$ .