Testing Heteroskedasticity in Nonlinear and Nonparametric Regressions with an Application to Interest Rate Volatility

XU ZHENG*

School of Economics, Shanghai University of Finance and Economics
Revised: December 2006

ABSTRACT: This paper presents new tests for heteroskedasticity in nonlinear and nonparametric regression models. The tests have an asymptotically standard normal distribution under the null hypothesis of homoskedasticity, and are consistent against any form of heteroskedasticity and local departures from homoskedasticity converging to the null at proper rates. Monte Carlo simulations demonstrate that the tests have high power against those heteroskedastic models which previous parametric tests have low power against. When applied to term structure models using the 7-day Eurodollar deposit spot rate data, from June 1973 to February 1995, the tests reject the volatility functions commonly used in parametric interest rate models of mean or drift functions. However, with a nonparametric drift function, the tests do not reject the parametric volatility function of Aït-Sahalia (1996), who finds higher volatility at both low and high interest rates. By contrast, this work suggests that higher volatility only occurs at higher interest rates.

KEYWORDS: Heteroskedasticity, Homoskedasticity, Consistent Test, Nonlinear Regression, Nonparametric Regression, Nonparametric Estimation, Local Alternative, Term Structure Model, Interest Rate Volatility.

JEL classification: C12, C14, G13

* I thank Maxwell Stinchcombe for helpful comments that significantly improved the paper. I also thank Yacine Aït-Sahalia for kindly sharing his data used in the application and Heather Anderson for helpful comments on an earlier version of this paper. Correspondence address: School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China. Phone: 86-21-65904613. Fax: 86-21-65904198. Email: xuzheng1@gmail.com.
1. Introduction

Least squares procedures give consistent but inefficient estimators, and inconsistent covariance estimators in the presence of heteroskedasticity. This has led to the development of many parametric tests for heteroskedasticity in regression models, especially linear ones. Most of the tests can be classified into two categories as in Pagan and Pak (1993). One type of tests is the Lagrange multiplier (LM) test or score test. The tests include those of Glejser (1969), Harvey (1976), Godfrey (1978), Breusch and Pagan (1979), Cook and Weisberg (1983), Evans and King (1988), Bera and Higgins (1992), Eubank and Thomas (1993), and others. The Breusch-Pagan test includes many such tests as special cases. The other type of tests is based on the least squares residual. They include those of Anscombe (1961), Goldfeld and Quandt (1972), Bickel (1978), Szroeter (1978), White (1980), Koenker (1981), Engle (1984), Wooldridge (1991), Lee (1992), Diblasi and Bowman (1997), and others. Those test statistics can be obtained from a regression of squared least squares residual on some function of regressors. A third category of tests are based on comparing different quantile or expectile estimates. The work includes those of Koenker and Bassett (1982) and Newey and Powell (1987).

As shown by Pagan and Pak (1993), most of the tests can be recast as special cases of the conditional moment test of Newey (1985) and Tauchen (1985). However the conditional moment test is not robust against misspecification. Specifically, all heteroskedasticity tests, implicitly or explicitly, require that the conditional variance be characterized by a certain parametric family of functions. As a result, if the form of heteroskedasticity is misspecified, these tests may have low power in detecting the real heteroskedasticity. In other words, these tests are not robust against misspecification. Lee (1992) and Eubank and Thomas (1993) used nonparametric methods in estimating the regression function while still assuming a parametric functional form for either the conditional variance or the error distribution. Thus their tests are not robust against misspecification of the form of

1 See section 5 and tables 1B and 4 for two examples where a score test has low power.
heteroskedasticity either.

There are two possible solutions to circumvent the aforementioned inconsistency problem in both types of tests. One solution is to stick with the least squares and use the heteroskedasticity-corrected variance estimator proposed by Eicker (1967) and White (1980). However the least squares estimator (OLS) is not efficient if heteroskedasticity exists. The efficiency loss of OLS relative to the efficient generalized least squares (GLS) can be substantial\(^2\). Another solution is to estimate the conditional variance by nonparametric methods and use the estimator to obtain the asymptotically efficient generalized least squares. Various nonparametric estimates of the conditional variance have been proposed by Carroll (1982), Robinson (1987), Delgado (1992), and Hidalgo (1992). But if no heteroskedasticity is present, then it is not necessary to use the complicated approach. To decide which method to use, a test which can detect any form of heteroskedasticity is needed.

In this paper, we first propose a nonparametric test of heteroskedasticity for nonlinear regression models, without assuming a parametric form for heteroskedasticity as in most of the parametric tests. The proposed test is shown to be robust against all possible departures from homoskedasticity. We then extend the test to testing for heteroskedasticity in nonparametric regressions. Since no parametric form is assumed for the regression function, the test is also robust to possible misspecification of the parametric regression function. The extension to nonparametric regressions is important, both for a general and a specific reason. Generally, it is hard to imagine real data where the regression function clearly belongs to a parametric family but there is no information about the form of the conditional variance. More specifically, there is little theoretical guidance on the parametric form of term structure models. If we believe that heteroskedasticity may be important in interest rates, then a test should not be limited by potential misspecification of the term structure model.

\(^2\) See section 5 and table 1A for an example.
The plan of the paper is as follows. Section 2 motivates the nonparametric tests. Section 3 presents the test statistic for testing heteroskedasticity in nonlinear regression models. Section 4 extends the test in section 3 to nonparametric regressions. Section 5 contains Monte Carlo simulation results. The simulations demonstrate that the tests have high power against those heteroskedastic models which previous parametric tests have low power against. Section 6 applies the tests to testing for volatility functions in interest rate models. Using the 7-day Eurodollar deposit spot rate data, from June 1973 to February 1995, our tests reject the volatility functions commonly used in term structure models. However, the tests do not reject the parametric volatility function of Aït-Sahalia (1996) in the model with a nonparametric drift function. This volatility function implies that the volatility increases as the interest rate increases, in contrast to Aït-Sahalia’s (1996) result. Section 7 concludes the paper.

2. The Motivation of the Tests

Consider a nonlinear regression model or a nonparametric regression model
\[ y = g(x) + \epsilon, \quad E(\epsilon|x) = 0, \]
where \( y \) is the dependent variable, \( x \) is a \( m \times 1 \) vector of regressors, and \( \epsilon \) is the error term. In nonlinear regressions, \( g \) is assumed to belong to a parametric family of functions \( f(x, \theta) \) where \( \Theta \subset R^l \). In nonparametric regressions, \( g \) is assumed to be a smooth function, but is not specified to have a parametric form.

The null hypothesis to be tested is that the regression model is homoskedastic:
\[ H_0 : \quad \Pr[V(\epsilon|x) = V(\epsilon)] = 1, \]
while the alternative is that heteroskedasticity exists:
\[ H_1 : \quad \Pr[V(\epsilon|x) = V(\epsilon)] < 1. \]

Unlike previous tests of heteroskedasticity in the literature, we do not specify
a parametric form for heteroskedasticity. Thus the alternative encompasses all possible forms of heteroskedasticity.

We extend Zheng’s (1996) test on regression functions to this case. Other nonparametric tests of functional form have been proposed by Lee (1988), Bierens (1990), Horowitz and Härdle (1992), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Hong and White (1995), Fan and Li (1996), Stinchcombe and White (1997), and others. Denote \( \sigma^2 \equiv E(\epsilon^2) \). As in the White’s (1980) test, we examine the moment condition

\[
E[(\epsilon^2 - \sigma^2)\phi(x)] = 0, \tag{2.4}
\]

which holds under the null for any vector of functions \( \phi(x) \). In White’s test, \( \phi(x) \) is chosen to be a vector of second and cross moments of the regressors. We choose the function to be

\[
\phi(x) = E[(\epsilon^2 - \sigma^2)|x] p(x), \tag{2.5}
\]

where \( p(\cdot) \) is the density function of \( x_i \). Denote \( W \equiv E[(\epsilon^2 - \sigma^2)\phi(x)] \). If the null is true, then \( W = 0 \). But, under the alternative, since \( E[(\epsilon^2 - \sigma^2)|x] = V(\epsilon|x) - V(\epsilon) \neq 0 \), we have

\[
W = E\{\epsilon^2 - \sigma^2\}E[(\epsilon^2 - \sigma^2)|x] p(x)\}
  = E\{E[(\epsilon^2 - \sigma^2)|x] \}^2 p(x)\}
  = E\{V(\epsilon|x) - V(\epsilon) \}^2 p(x) > 0. \tag{2.6}
\]

Therefore \( W \) can detect any form of heteroskedasticity.

Given a random sample \( \{(y_i, x_i)\}_{i=1}^n \) from model (2.1), the unknown conditional variance \( E(\epsilon^2|x_i) \) can be estimated by the kernel regression method:

\[
\hat{E}(\epsilon^2|x_i) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) \epsilon_j^2 / \hat{p}(x_i), \tag{2.7}
\]

where \( \hat{p} \) is the kernel estimator of the density function \( p \),

\[
\hat{p}(x_i) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right),
\]

(2.8)

where \( K(\cdot) \) is a kernel function and \( h \) is a bandwidth parameter. Let \( \hat{g}(x) \) be a parametric or nonparametric estimator of \( g(x) \). Denote the estimated residual by \( e_i = y_i - \hat{g}(x_i) \). Then the (unconditional) variance \( \sigma^2 \) can be estimated by \( \hat{\sigma}^2 \equiv \sum_{i=1}^{n} [y_i - \hat{g}(x_i)]^2 / n \).

Substituting these estimators into \( W \), we obtain its sample analogue

\[
W_n \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right)(e_i^2 - \hat{\sigma}^2)(e_j^2 - \hat{\sigma}^2),
\]

(2.9)

where the negligible terms with \( i = j \) are dropped in order to apply \( U \)-statistic theory. Our final test statistics will be based on \( W_n \).

Note that it is not necessary for the mean function \( g(x_i) \) to have the same \( x_i \) as the conditioning variables in the variance function \( V(y_i|x_i) \). All above arguments go through if \( g(x_i) \) is replaced by \( g(z_i) \) for any random vector \( z_i \).

### 3. Testing Heteroskedasticity in Nonlinear Regressions

Consider a nonlinear regression model

\[
y = f(x, \theta) + \epsilon, \quad E(\epsilon|x) = 0,
\]

(3.1)

here \( f(x, \theta) \) is defined on \( R^m \times \Theta \) where \( \Theta \subset R^l \). Let \( \hat{\theta} \) be the nonlinear least squares estimator of \( \theta_0 \) [cf. Jennrich (1969)]. Then the residual \( \epsilon_i \) can be estimated by \( \epsilon_i = y_i - f(x_i, \hat{\theta}) \). The (unconditional) variance \( \sigma^2 \) can be estimated by \( \hat{\sigma}^2 \equiv \sum_{i=1}^{n} \epsilon_i^2 / n \).
The statistic \( W_n \) becomes
\[
W_{1n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K(\frac{x_i - x_j}{h})(e_i^2 - \hat{\sigma}^2)(e_j^2 - \hat{\sigma}^2)
\]
(3.2)

Our test statistic \( T_{1n} \) is based on \( W_{1n} \):
\[
T_{1n} \equiv \frac{nh^{m/2}W_{1n}}{\hat{s}},
\]
(3.3)
where \( \hat{s} \) is the estimated standard error of \( W_{1n} \) and is defined in Theorem 1 below.

For any matrix \( A \), let \( \|A\| \) denote its Euclidean norm, i.e., \( \|A\| = \left[ \text{tr}(AA') \right]^{1/2} \).
We impose the following regularity assumptions.

**Assumption A1:** \( \{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\} \) is a random sample from a probability distribution \( F(y, x) \) on \( \mathbb{R} \times \mathbb{R}^m \). The density function \( p(x) \) and its first-order derivatives are uniformly bounded.

**Assumption A2:** \( E(\epsilon^8|x) \) is continuously differentiable and bounded by a measurable function \( b(x) \) such that \( E[b^2(x)] < \infty \).

**Assumption A3:** The parameter space \( \Theta \) is a compact and convex subset of \( \mathbb{R}^l \). \( f(x, \theta) \) is a Borel measurable function on \( \mathbb{R}^m \) for each \( \theta \) and a twice continuously differentiable real function on \( \Theta \) for each \( x \in \mathbb{R}^m \). Moreover there exist a constant \( M < \infty \) such that
\[
E\left\{ \sup_{\theta \in \Theta} |f^2(x, \theta)| \right\} \leq M,
\]
\[
E\left\{ \sup_{\theta \in \Theta} \left\| \frac{\partial f(x, \theta)}{\partial \theta} \frac{\partial f(x, \theta)}{\partial \theta'} \right\| \right\} \leq M,
\]
\[
E\left\{ \sup_{\theta \in \Theta} \left\| [y - f(x, \theta)] \frac{2\partial f(x, \theta)}{\partial \theta} \frac{\partial f(x, \theta)}{\partial \theta'} \right\| \right\} \leq M,
\]
\[
E\left\{ \sup_{\theta \in \Theta} \left\| [y - f(x, \theta)] \frac{2\partial^2 f(x, \theta)}{\partial \theta \partial \theta'} \right\| \right\} \leq M.
\]

**Assumption A4:** \( E\left[ (y - f(x, \theta))^2 \right] \) takes a unique minimum at \( \theta_0 \in \Theta \) where \( \theta_0 \)
is an interior point of $\Theta$.

**Assumption A5:** $K(u)$ is a nonnegative, bounded, continuous, and symmetric function such that $\int K(u)du = 1$, $\int \|u\|^2 K(u)du < \infty$, and $\int \int K(u)K(u+v)du^2dv < \infty$.

The assumptions A1, A3, and A4 are essentially the standard assumptions for ensuring the consistency and asymptotic normality of nonlinear least squares estimators. Assumption A2 ensures that our test statistic has a finite limiting variance. The kernel function in assumption A5 is the one most commonly used in the kernel method.

In addition to testing fixed alternative (2.3), we also examine the power of the test against the following sequence of local alternatives

$$H_{1n} : V(\epsilon|x) = V(\epsilon) + \delta_n \cdot l(x),$$

(3.4)

where the known function $l(\cdot)$ is continuously differentiable and $E[l^2(x)] < \infty$.

The asymptotic behaviors of $W_{1n}$ and $T_{1n}$ are provided in the following theorem (all proofs are provided in the appendix).

**THEOREM 1:** Given assumptions A1-A5, if $h \to 0$ and $nh^m \to \infty$, then the following hold:

(i) under the null hypothesis (2.2),

$$nh^{m/2}W_{1n} \xrightarrow{d} \mathcal{N}(0, s^2),$$

(3.5)

where

$$s^2 = 2 \int K^2(u)du \cdot E\left\{[V(\epsilon^2|x)]^2 p(x)\right\}.$$  

(3.6)

$s^2$ can be consistently estimated by

$$\hat{s}^2 = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{hm} K^2(\frac{x_i - x_j}{h})(e_i^2 - \hat{\sigma}^2)(e_j^2 - \hat{\sigma}^2).$$

(3.7)

Thus $T_{1n} \xrightarrow{d} \mathcal{N}(0, 1)$;
(ii) under the alternative hypothesis (2.3),

\[ T_{1n} \left/ \left( nh^{m/2} \right) \right. \xrightarrow{p} E \left\{ \left[ V(\epsilon|x) - V(\epsilon) \right]^2 p(x) \right\} / s_1 > 0, \quad (3.8) \]

where

\[ s_1^2 \equiv 2 \int K^2(u) du \cdot E \left\{ \left[ V(\epsilon^2|x) + (V(\epsilon|x) - \sigma^2)^2 \right]^2 p(x) \right\} > 0; \quad (3.9) \]

(iii) under the local alternatives (3.4) with \( \delta_n = n^{-1/2}h^{-m/4} \),

\[ T_{1n} \xrightarrow{d} N(\mu, 1), \quad (3.10) \]

where

\[ \mu = E \left[ l^2(x)p(x) \right] / s. \quad (3.11) \]

This bandwidth condition is the same as that for obtaining the consistency in quadratic mean of kernel density estimators [see Prakasa Rao (1983), theorem 3.1.2]. Since \( T_n \to \infty \) in probability under the alternative, the test is consistent. It is obvious that the test is also consistent against any local alternatives converging to the null at rates slower than the parametric rate \( n^{-1/2} \) or the Pitman’s drift but has no power against the parametric rate. The convergence rate is between \( n^{-1/4} \) and \( n^{-1/2} \). This local power result is typical in nonparametric testing literature and parallel to that for testing a parametric regression function [see Eubank and Spiegelman (1990), Härdle and Mammen (1993), and Zheng (1996)].

Note that the test can be easily extended to testing for heteroskedasticity of a known form. For example, suppose that null hypothesis is \( H_0 : V(\epsilon_i|x_i) = v(x_i, \delta) \) where \( \delta \) is a vector of parameters. We can use the proposed test to test the null hypothesis by dividing the data by \( \sqrt{v(x_i, \hat{\delta})} \) where \( \hat{\delta} \) is a \( \sqrt{n} \) consistent estimator
of \( \delta \) so the null model becomes a homoskedastic model. Or we can use the following statistic similar to \( W_{1n} \),

\[
W_{1n}' \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h} K\left(\frac{x_i - x_j}{h}\right) [e_i^2 - v(x_i, \hat{\delta})][e_j^2 - v(x_j, \hat{\delta})].
\] (3.12)

We can easily show that the test based on \( W_{1n}' \) is consistent against any departures from the known conditional variance structure.

### 4. Testing Heteroskedasticity in Nonparametric Regressions

Consider a nonparametric regression model

\[
y = g(x) + \epsilon, \quad E(\epsilon|x) = 0,
\] (4.1)

where \( g(x) \) is assumed to be a smooth function, but otherwise unspecified.

We use the kernel method to estimate \( g(x) \):

\[
\hat{g}(x) = \frac{\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right) y_j}{\sum_{j=1}^{n} K\left(\frac{x - x_j}{h}\right)}.
\]

The (unconditional) variance \( \sigma^2 \) can be estimated by \( \hat{\sigma}^2 \equiv \sum_{i=1}^{n} [y_i - \hat{g}(x_i)]^2 / n. \)

Let \( \hat{u}_i = [y_i - \hat{g}(x_i)]^2 - \hat{\sigma}^2 \). The statistic \( W_n \) becomes

\[
W_{2n} \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h} K\left(\frac{x_i - x_j}{h}\right) \hat{u}_i \hat{u}_j.
\] (4.2)

Our test statistic \( T_{2n} \) is defined as

\[
T_{2n} \equiv \frac{nh^{m/2} W_{2n}}{s},
\] (4.3)

where \( s \) is defined in Theorem 2 below.
We impose the following assumptions.

**Assumption B1:** \(\{(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)\}\) is a random sample from a probability distribution \(F(y, x)\) on \(\mathbb{R} \times \mathbb{R}^m\). \(E(e^\delta|x)\) are continuously differentiable and bounded by a measurable function \(b(x)\) such that \(E[b^2(x)] < \infty\).

**Assumption B2:** The density function \(p(x)\) has a compact support \(C\), is positive on \(C\), and has a continuous and bounded second derivative on \(C\). All derivatives of \(g(x) \equiv E(Y|X = x)\) of order \(r\) exit and the \(r\)th order derivative of \(g\) is bounded.

**Assumption B3:** \(K(u)\) is a nonnegative, bounded, continuous, symmetric, and satisfies \(\int K(u)du = 1\), \(\int u_1^{l_1} \cdots u_m^{l_m} K(u)du = 0\) for all \(1 \leq l_1 + \cdots + l_m < r\), and \(\int u_1^{l_1} \cdots u_m^{l_m} K(u)du \neq 0\) for \(l_1 + \cdots + l_m = r\), where \(l_1, \ldots, l_m\) are nonnegative integers.

The assumption B1 is similar to assumption A2. Assumption B2 ensures that the test statistic is well-behaved. Since the density estimator \(\hat{p}\) appears in the denominator of \(\hat{g}(x)\), extreme small values of \(\hat{p}\) may cause the estimator \(g\) and the test statistic to be ill-behaved. One solution to this problem is to introduce a trimming parameter to trim out those small values (e.g., Härdle and Stoker (1989)). Here we follow the approach of Collomb and Härdle (1986) to assume that \(p\) has a compact support and thus avoid introducing the extra trimming parameter. The compact support is also needed for obtaining uniform convergences of \(\hat{p}\) and \(\hat{g}\) (see Collomb and Härdle (1986)). In practice, if the support of \(p\) is not compact, a suitable transformation can be used to transform it to a compact support. The existence of \(r\)th order derivative of \(g\) in B2 and the use of a higher order kernel in B3 ensure that the bias of \(\hat{g}\) is sufficiently small and of order \(O(h^r)\) (e.g., Powell et al. (1989)). \(r\) will be determined in Theorem 2.

The following theorem provides the asymptotic behaviors of \(W_{2n}\) and \(T_{2n}\).

**THEOREM 2:** Given assumptions B1-B3, if \(h \to 0\), \(nh^m \to \infty\), and \(nh^{m+4r} \to 0\), then the following hold:
(i) under the null hypothesis (2.2),

\[ n h^{m/2} W_{2n} \xrightarrow{d} \mathcal{N}(0, s^2), \]  

where

\[ s^2 = 2 \int K^2(u) du \cdot E \left\{ [V(\epsilon^2|x)]^2 p(x) \right\}. \]  

\[ (4.4) \]

\[ (4.5) \]

\[ s^2 \text{ can be consistently estimated by} \]

\[ s^2 \equiv \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K^2(x_i - x_j) h \hat{u}_i^2 \hat{u}_j^2. \]  

\[ (4.6) \]

Thus \( T_{2n} \xrightarrow{d} \mathcal{N}(0, 1); \)

(ii) under the alternative hypothesis (2.3),

\[ T_{2n}/(nh^{m/2}) \xrightarrow{P} E \left\{ [V(\epsilon|x) - V(\epsilon)]^2 p(x) \right\} / s_1 > 0, \]  

where

\[ s_1^2 \equiv 2 \int K^2(u) du \cdot E \left\{ [V(\epsilon^2|x) + (V(\epsilon|x) - \sigma^2)^2]^2 p(x) \right\} > 0; \]  

\[ (4.7) \]

\[ (4.8) \]

(iii) under the local alternatives (3.4) with \( \delta_n = n^{-1/2} h^{-m/4}, \)

\[ T_{2n} \xrightarrow{d} \mathcal{N}(\mu, 1), \]  

where

\[ \mu = E \left[ l^2(x)p(x) \right] / s. \]  

\[ (4.9) \]

\[ (4.10) \]
The asymptotic behaviors of $T_{2n}$ are similar to that of $T_{1n}$. Similarly to the case of nonlinear regressions, we can also extend the test to testing for a parametric conditional variance function, $v(x_i, \delta)$, based on the statistic

$$W_{2n}^2 \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K(\frac{x_i - x_j}{h})$$

$$\cdot [(y_i - \hat{g}(x_i))^2 - v(x_i, \hat{\delta})][(y_j - \hat{g}(x_j))^2 - v(x_j, \hat{\delta})], \quad (4.11)$$

where $\delta$ is a $\sqrt{n}$-consistent estimator of $\delta$. We can easily show that the test has limiting normal distribution under the null and is robust against any misspecification of the conditional variance.

5. Monte Carlo Simulations

In this section, we conduct some Monte Carlo simulations of the size and power of the tests proposed in sections 3 and 4. The random numbers are generated using the normal or uniform pseudo number generators, RNDN, RNDU, on GAUSS. The seeds used are the default seeds generated by the system clock on Gauss.

First we present an example where the heteroskedasticity is severe and efficiency loss of OLS relative to GLS is large, but a parametric test has low power to detect the heteroskedasticity.

In model 1, the dependent variable $y$ is generated as follows:

$$y_i = \beta_0 + \beta_1 x_i + x_i \epsilon_i, \quad i = 1, \ldots, n, \quad (5.1)$$

where true $\beta_0 = 1$ and $\beta_1 = 1$, $x_i$ has the uniform distribution on $(-1, 1)$, $\epsilon_i \sim \mathcal{N}(0, 1)$, and $x_i$ and $\epsilon_i$ are independent.

The null hypothesis $H_0$ we want to test is that the model is homoskedastic. We measure the degree or severity of heteroskedasticity by the coefficient of variation.
(CV) of the conditional variance $V(\epsilon_i|x_i)$:

$$CV = \{V[V(\epsilon_i|x_i)]\}^{1/2} \frac{E[V(\epsilon_i|x_i)]}{E[V(\epsilon_i|x_i)]}.$$  \tag{5.2}

Heteroskedasticity with $CV$ less 0.5 is usually regarded as moderate and greater than 0.5 as severe (see Evans and King (1988)). For model 1, $CV = 0.911$.

Table 1A provides the bias, standard deviation (SD), square root of mean square error (RMSE) for OLS and GLS estimators and the efficiency of OLS relative to GLS. Each calculation is based on 100,000 simulated samples. The simulation is conducted to sample sizes 100, . . . , 500. We can see that the efficiency loss of OLS relative to GLS are substantial, especially for $\beta_0$. Note that the GLS based on an incorrect specification of $V(\epsilon_i|x_i)$ is biased and inconsistent and will result in even larger efficiency loss than OLS.

We compare our tests with the parametric score test of Cook and Weisberg (1983). This test is similar to the test of Breusch and Pagan (1979). In all of the following models, to form the score test, we assume that the conditional variance be $V(\epsilon_i|x_i) = w(\lambda x_i)$ for some function $w$ with $w'(0) \neq 0$ and $\lambda = 0$ corresponds to homoskedasticity. Let $\gamma(x_i) = \frac{\partial V(\epsilon_i|x_i)}{\partial \lambda}|_{\lambda=0} = x_i w'(0)$. The score test is defined as

$$S_n = \frac{\sum_{i=1}^n (x_i - \bar{x})(\epsilon_i^2/\hat{\sigma}^2 - 1))^2}{2\sum_{i=1}^n (x_i - \bar{x})^2}. \tag{5.3}$$

Under the null of homoskedasticity, $S_n \rightarrow_d \chi^2(1)$.

As in Pagan and Pak (1993), this score test can be interpreted as a conditional moment test with the moment condition $E[x_i(\epsilon_i^2 - \sigma^2)] = 0$ under the null of homoskedasticity. Model 1 is constructed so that $E[x_i(\epsilon_i^2 - \sigma^2)] = 0$ also holds under the heteroskedasticity. Thus the score test should have low power. We choose significance levels $\alpha = 1\%, 5\%$ and $10\%$. For the score test, the null is rejected if $S_n > \chi^2_{1-\alpha}$ where $\chi^2_{1-\alpha}$ is the $(1 - \alpha)$ quantile of $\chi^2(1)$. Each experiment is based on 1,000 replications.
The results in Tables 1B show that the score test incurs very large type II error even for large sample size 1,000. The power does not improve as the sample size gets larger. This clearly demonstrates that this score test is not robust against misspecification of the form of heteroskedasticity.

To conduct our two nonparametric tests, we choose the kernel function $K$ to be the Epanechnikov kernel: $K(u) = 0.75(1 - u^2)$ if $|u| \leq 1$ and $K(u) = 0$ otherwise. We choose bandwidth to be $h = s_{x^*}n^{-1/5}$ where $s_x$ is the sample standard deviation of $x$. For test statistic $T_{1n}$, the parametric regression function is estimated by the nonlinear least squares. For $T_{2n}$, the regression function is estimated by the kernel method with the same bandwidth chosen as above. Each experiment is based on 1,000 replications.

The simulation results in tables 1B show that both nonparametric tests have very high power against the heteroskedastic model. Though $T_{2n}$ assumes no knowledge about the regression function, its power is comparable with that of $T_{1n}$ which assumes that the parametric regression function is correctly specified.

We next compare our tests with the score test in two models where the form of heteroskedasticity is correctly specified by the score test. Since the score test is a locally most powerful test under correct specification, we expect that the score test will do better than our nonparametric tests, which is designed to have power against misspecifications from all directions.

In model 2, the dependent variable $y$ is generated as follows:

$$y_i = \beta_0 + \beta_1 x_i + e^{\lambda x_i} \epsilon_i, \quad i = 1, \ldots, n, \quad (5.4)$$

where true $\beta_0 = 1$ and $\beta_1 = 1$, $x_i \sim U(-1, 1)$, $\epsilon_i \sim N(0, 1)$, and $x_i$ and $\epsilon_i$ are independent. Corresponding to $CV = 0$, 0.5, and 1, $\lambda = 0$, 0.447, and 0.961 respectively.

In model 3,

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + (1 + \lambda x_i) \epsilon_i, \quad i = 1, \ldots, n, \quad (5.5)$$
where true $\beta_0 = 1, \beta_1 = 1, \beta_2 = 1$, $x_i \sim U(-1, 1)$, $\epsilon_i \sim N(0, 1)$, and $x_i$ and $\epsilon_i$ are independent. Corresponding to $CV = 0, 0.5$, and $1$, $\lambda = 0, 0.460$, and $1.331$ respectively.

The sizes of the tests are shown in Tables 2 and 3 for $\lambda = 0$. The results show that the nonparametric test are conservative in rejecting a true model. The power of the nonparametric tests always increases as the sample size gets larger or as heteroskedasticity gets more severe. The score test, as expected, has higher power than our nonparametric tests, particularly for small samples or moderate degree of heteroskedasticity.

In the last model, we consider a case where the regression function is nonlinear and the conditional variance is not a monotonic function of $x$ as in models 2 and 3. In model 4,

$$y_i = \exp(\beta_0 + \beta_1 x_i) + (1 + \lambda x_i^2)\epsilon_i, \quad i = 1, \ldots, n,$$

(5.6)

where true $\beta_0 = 1$ and $\beta_1 = 1$, $x_i \sim U(-1, 1)$, $\epsilon_i \sim N(0, 1)$, and $x_i$ and $\epsilon_i$ are independent. Corresponding to $CV = 0, 0.5$, and $1$, $\lambda = 0, 1.151$, and $4.530$ respectively.

The size and power of the three tests for model 4 are provided in Table 4. Our tests have much higher power than the score test. As the sample size increases, the power of the score test does not improve as the nonparametric tests. Again, the worse performance of the score test is due to its misspecification of the form of heteroskedasticity.

To summarize, from the simulations, we may draw the following conclusions: First, the power of the parametric score or LM test crucially depends on correct specification of the form of heteroskedasticity. If the form is misspecified, the test may have low power. Secondly, if the form of heteroskedasticity is correctly specified, the score test has higher power than our consistent tests, especially for small samples and moderate heteroskedasticity. But for moderate samples or severe
heteroskedasticity, the powers of the score test and the nonparametric tests are comparable. Thirdly, the proposed nonparametric tests have power against all possible departures from homoskedasticity. The nonparametric tests can detect some heteroskedasticity that the parametric tests can not. Thus the nonparametric tests are most useful when the form of heteroskedasticity can not be specified with certainty or when the form is too complicated to be specified parametrically.

6. An Application to Interest Rate Volatility

In this section, we apply our tests to testing for parametric volatility functions of short-term interest rate.

Correct specifications of interest rate models, especially the volatility function, are crucial to pricing and hedging interest rate sensitive contingent claims. Mis-specifying the underlying term structure can lead to serious pricing and hedging errors [Canabarro (1995)]. Many models of the term structure have been proposed to price interest rate derivatives (see Chan et al. (1992), Aït-Sahalia (1996), and Stanton (1997) for more complete references).

However, little guidance can be found in theory as to how to choose a parametric specification for the term structure model. As a result, most of the interest rate models rely on ad hoc assumptions about the specifications of the regression function and the volatility function. Empirical tests of these models yield mixed results. For example, Chan et al. (1992) rejected a variety of parametric interest rate models using Hansen’s (1982) test and proposed a model with linear mean and constant elasticity of variance diffusion that could not be rejected by their test. But their model is rejected by the test of Aït-Sahalia (1996). The mixed results in empirical tests may be due to low power against false models or over-rejection of true models by a test. For instance, Hansen’s (1982) test is known to have low power rejecting certain false models. As shown in the simulations in section 5, our nonparametric tests are conservative in rejecting a true model and have high power against any false model, especially in large samples. Thus our
tests are ideally suited for the interest rate models, where a large amount of data is available.

We apply our test to a real data of short-term interest rate. The data used is the 7-day Eurodollar deposit rate, bid-ask midpoint, previously analyzed by Aıt-Sahalia (1996). The data set consists of \( n = 5505 \) daily observations from June 1, 1973 to February 25, 1995. The yields are converted to annualized interest rates. The annualized yields are plotted in Figure 1.

We have data on the interest rate observed in discrete times \( t_1, t_2, \ldots, t_n \) with \( 0 = t_1 < t_2 < \ldots < t_n = T \) over a time period \([0, T]\). Denote \( r_i \equiv r_{t_i} \). We have the data \( \{r_i\}_{i=1}^{n} \). To estimate the model, we use the following discrete-time model of Chan et al. (1992)

\[
r_{i+1} - r_i = \mu(r_i, \theta_0) + u_{i+1},
\]

\[
E(u_{i+1}|r_i) = 0, \quad E(u_{i+1}^2|r_i) = \sigma^2(r_i, \theta_0),
\]

(6.1)

where \( \mu(r, \theta) \) and \( \sigma^2(r, \theta) \) are the drift function and diffusion function respectively. In our framework, they are the conditional mean and variance. Though the above discrete-time model is an approximation to the underlying continuous-time process, Stanton (1997) shows that the approximation errors due to discretization are extremely small, especially compared with parameters estimation error and possible model misspecification errors. Thus we adopt the discrete model of Chan et al. (1992).

The daily changes in interest rates \( r_{i+1} - r_i \) are plotted against the previous day’s interest rate \( r_i \) in Figure 2. It clearly shows evidence of heteroskedasticity since the variance of interest rate tends to increase as the level of interest rate increases.

We first test the constant elasticity of variance model of Chan et al. (1992) (CKLS model), which contains many commonly used term structure models as special cases. The mean and variance are specified as \( \mu(r, \theta) = \theta_1 + \theta_2 r \) and
σ²(r, θ) = θ₁² r²θ₂ for θ = (θ₁, θ₂, θ₃, θ₄)' ∈ R⁴. The null hypothesis is that σ²(r) is correctly specified. As in Chan et al. (1992), the parameters θ are estimated by the Generalized Method of Moments (GMM) of Hansen (1982) based on the following four moments

\[ E(ε_{i+1}) = 0, \quad E(ε_{i+1} r_i) = 0, \]
\[ E(ε_{i+1}^2 - β_3^2 r_i^2 r_{i+1}) = 0, \quad E[(ε_{i+1}^2 - β_3^2 r_i^2 r_{i+1}) r_i] = 0. \] (6.2)

The weighting matrix in the GMM is chosen to be the identity matrix. The GMM estimates of drift and diffusion functions are

\[ \hat{μ}(r) = 0.00308 - 0.03687 r, \quad \hat{σ}²(r) = 0.00018 r^{0.29}. \] (6.3)

The estimates are similar to those obtained in Aït-Sahalia (1996). The estimated drift and diffusion functions are plotted in Figure 3 and Figure 4 respectively.

To conduct our nonparametric tests, we choose the same kernel as that used in the simulation and bandwidth \( h = s_r \cdot n^{-1/5} \) where \( s_r \) is the sample standard deviation of \( \{r_i\}_{i=1}^n \). The bandwidth is calculated to be \( h = 0.00641 \). The test statistic \( T_{1n} \) is calculated to be 520.89. Thus the CKLS diffusion specification is rejected at 5% level.

However, Figure 3 suggests that the mean function might be misspecified and the rejection may be due to the mean misspecification. Thus we also conduct the robust test \( T_{2n} \). To calculate \( T_{2n} \), we first estimate \( β_3 \) and \( β_4 \) based on minimizing the nonparametric residual sum of squares \( \sum_{i=1}^n [(r_{i+1} - r_i - \hat{μ}(r_i))^2 - β_3^2 r_i^{2β_4}] \) where \( \hat{μ}(r) \) is the kernel estimator of \( μ(r) \):

\[ \hat{μ}(r) = \frac{\sum_{j=1}^{n-1} K(\frac{r-r_j}{h}) (r_{j+1} - r_j)}{\sum_{j=1}^{n-1} K(\frac{r-r_j}{h})}. \] (6.4)

The estimates are \( \hat{β}_3 = 0.00609 \) and \( \hat{β}_4 = 0.145 \). \( T_{2n} \) is 22.47. The CKLS diffusion
function is still rejected at 5% level 4.

To see why the CKLS model is rejected by our tests, we compare the parametric CKLS estimates of drift and diffusion with their nonparametric kernel estimates. The kernel estimate of the drift function is given in (6.4). The diffusion function can be estimated as

\[ \hat{\sigma}^2(r) = \frac{\sum_{j=1}^{n-1} K\left(\frac{r-r_j}{h}\right) (r_{j+1} - r_j)^2}{\sum_{j=1}^{n-1} K\left(\frac{r-r_j}{h}\right)} - \left[\hat{\mu}(r)\right]^2. \]  

(6.5)

From kernel estimation theory (see Härdle (1990)), \( \sqrt{n\hat{h}}[\hat{\mu}(r) - \mu(r)] \to_d N(0, V_{\mu}) \) where \( V_{\mu} = \int K^2(u)du \cdot \sigma^2(r)/\hat{p}(r) \) where \( \hat{p} \) is the usual kernel density estimator of \( p \). Then a 95% confidence bands for \( \mu(r) \) can be constructed as \([\hat{\mu} - 1.96\sqrt{\hat{V}_\mu}/\sqrt{n\hat{h}}, \hat{\mu} + 1.96\sqrt{\hat{V}_\mu}/\sqrt{n\hat{h}}]\). Similarly, \( \sqrt{n\hat{h}}[\hat{\sigma}^2(r) - \sigma^2(r)] \to_d N(0, V_{\sigma^2}) \) where \( V_{\sigma^2} = \int K^2(u)du \cdot \hat{V}(u_i^2|r_i = r)/\hat{p}(r) \). \( V_{\sigma^2} \) can be estimated by the kernel estimator \( \hat{V}_{\sigma^2} = \int K^2(u)du \cdot \hat{V}(u_{i+1}^2|r_i = r)/\hat{p}(r) \) where

\[ \hat{V}(u_{i+1}^2|r_i = r) = \frac{\sum_{j=1}^{n-1} K\left(\frac{r-r_j}{h}\right) (r_{j+1} - r_j - \hat{\mu}(r_j))^4}{\sum_{j=1}^{n-1} K\left(\frac{r-r_j}{h}\right)} - \left[\hat{\sigma}^2(r)\right]^2. \]  

(6.6)

A 95% confidence bands for \( \sigma^2(r) \) can be constructed in a similar way as that for \( \mu(r) \). From Figure 4, one can clearly see that the CKLS diffusion function is also misspecified.

Note that for both kernel estimators of the drift and diffusion, the confidence bands are much tighter for lower interest rates, below about 12%, than for higher interest rates. As can be seen from Figure 2, this is due to the fact that data are

4 Since our focus here is the volatility, to save space, we do not report the results of tests of the drift. Using the test of Zheng (1996), the CKLS drift specification is also strongly rejected. The drift misspecification can be seen from Figures 3, by comparing it with the kernel estimate.
sparser when interest rates get higher than 12%. Due to even sparser data around 13%, 17%, and 21%, there are spikes around these interest rates in the volatility estimates.

Next we test the interest model proposed by Aït-Sahalia (1996). The drift and diffusion are specified as

\[
\mu(r, \beta) = \beta_1 + \beta_2 r + \beta_3 r^2 + \beta_4 / r,
\]

\[
\sigma^2(r, \beta) = \beta_5 + \beta_6 r + \beta_7 r \beta_8.
\] (6.7)

To estimate the eight parameters, \(\beta_1, \ldots, \beta_8\), using the GMM, we use the following four extra moments in addition to the four moments in (6.2)

\[
E(\epsilon_{i+1}r_i^2) = 0, \quad E(\epsilon_{i+1}r_i^3) = 0,
\]

\[
E[(\epsilon_{i+1}^2 - \beta_3^2 r_i^{2\beta_1})r_i^2] = 0, \quad E[(\epsilon_{i+1}^2 - \beta_3^2 r_i^{2\beta_1})r_i^3] = 0. \quad (6.8)
\]

Again the identity matrix is used as the weighting matrix. The GMM estimates of drift and diffusion functions are

\[
\hat{\mu}(r) = -0.004643 - 0.0433r - 0.1143r^2 + 0.0001360/r,
\]

\[
\hat{\sigma}^2(r) = 0.0001108 - 0.001883r + 0.9681r^{2.073}. \quad (6.9)
\]

The estimates are virtually identical to that obtained in Aït-Sahalia (1996).

The test statistics are calculated to be \(T_{1n} = 85.70\) and \(T_{2n} = 1.37\). Thus the parametric diffusion function with the parametric drift is rejected at 5% level. From Figure 5, it is clear that the drift is misspecified. Since, in the parametric model, consistency of the diffusion function estimator depends on correct specification of both drift and diffusion, the parametric diffusion estimator shown in Figure 6
is biased. But the parametric diffusion with the nonparametric drift can not be rejected at 5% level. This semiparametric estimator of diffusion function is

\[ \hat{\sigma}^2(r) = 0.00000317 - 0.000196r + 0.00412r^2.073. \]

(6.10)

The semiparametric estimator eliminates the bias due to misspecification of the drift and is a good estimator, as shown in Figure 7 labeled as the semiparametric estimator.

Thus we conclude that the Aït-Sahalia’s (1996) parametric diffusion or volatility specification does provide a good approximation, if not the true function, to the underlying diffusion of interest rates. The diffusion function suggests that the volatility is an increasing function of the interest rate. This accepted volatility function can then be used in pricing bond and other derivatives. This application, however, is beyond the scope of this paper.

7. Concluding Remarks

This paper has provided two tests of heteroskedasticity, one for nonlinear regression models and one for nonparametric regression models. The tests are based on nonparametric estimation techniques and are robust against any departure from homoskedasticity. The test for nonparametric regressions is also robust to misspecification of parametric regression functions. Monte Carlo simulations show that, when the parametric tests misspecify the form of heteroskedasticity, the nonparametric tests have much higher power against heteroskedasticity than the parametric tests.

We apply the tests to testing for correct specifications of the parametric diffusion or volatility functions in interest rate models. Using the 7-day Eurodollar deposit spot rate data, from June 1973 to February 1995, the tests reject the parametric volatility functions commonly used in parametric term structure models with parametric mean or drift functions. Though our test for nonlinear regressions
rejects the term structure model of Aït-Sahalia (1996) with both parametric drift
and volatility functions, the robust test for nonparametric regressions do not reject
his volatility function. The more robust estimator of the volatility function with a
nonparametrically estimated drift suggests that the volatility is an increasing func-
tion of the level of interest rate. This finding is contrary to the result of Aït-Sahalia
(1996) that the volatility function has a parabolic shape or that interest rates are
more volatile for both low and high interest rates. It may be that the inconsistency
of Aït-Sahalia’s (1996) estimated volatility function is due to the misspecification
of the drift function, not the volatility function itself.

The proposed tests can be extended to finding a good specification of the
volatility function for other data, such as exchange rates and stock returns. These
extensions are left for future research.

APPENDIX: Proofs

Proof of Theorem 1:

(i) The idea of the proof is similar to the one used in Zheng (1996). The
statistic $W_{1n}$ is decomposed into three parts:

$$W_{1n} = \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K(\frac{x_i - x_j}{h})(e_i^2 - \sigma^2)(e_j^2 - \sigma^2) \right\}$$

$$- 2\left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K(\frac{x_i - x_j}{h})(e_i^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2) \right\}$$

$$+ \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K(\frac{x_i - x_j}{h})(\hat{\sigma}^2 - \sigma^2)^2 \right\}$$

$$\equiv U_{1n} - 2U_{2n} + U_{3n}.$$ (A.1)

Under the null, we will show that $nh^{m/2}U_{1n} \rightarrow_d N(0, s^2)$, $nh^{m/2}U_{2n} = o_p(1)$, and
$nh^{m/2}U_{3n} = o_p(1)$.
First we show that $nh^{m/2}U_{1n} \xrightarrow{d} \mathcal{N}(0, s^2)$. $U_{1n}$ is decomposed into three parts:

$$U_{1n} = \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) (\epsilon_i^2 - \sigma^2)(\epsilon_j^2 - \sigma^2) \right\}$$

$$+ 2\left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) (\epsilon_i^2 - \sigma^2)(\epsilon_j^2 - \epsilon_j^2) \right\}$$

$$+ \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) (\epsilon_i^2 - \epsilon_i^2)(\epsilon_j^2 - \epsilon_j^2) \right\}$$

$$\equiv Q_{1n} + 2Q_{2n} + Q_{3n}.$$

$Q_{1n}$ is a $U$-statistic with kernel

$$H_n(z_i, z_j) = \frac{1}{h^m} K\left(\frac{x_i - x_j}{h}\right) (\epsilon_i^2 - \sigma^2)(\epsilon_j^2 - \sigma^2),$$

where $z_i = (y_i, x_i)$. Under the null hypothesis, since $E(\epsilon_i^2|x_i) = \sigma^2$, $Q_{1n}$ is a degenerate $U$-statistic.

We apply theorem 1 of Hall (1984) to obtain the asymptotic distribution of $Q_{1n}$. Similarly to Zheng (1996), the condition of the Hall’s theorem can be easily verified. Since

$$E\left[H_n^2(z_1, z_2)\right] = \int \int \frac{1}{h^2m} K^2\left(\frac{x_1 - x_2}{h}\right) V(\epsilon_1^2|x_1)V(\epsilon_2^2|x_2)p(x_1)p(x_2)dx_1dx_2$$

$$= \frac{1}{h^2m} \int \int K^2(u)V(\epsilon_1^2|x_1)V(\epsilon_2^2|x_1 - hu)p(x_1)p(x_1 - hu)dx_1h^mdu$$

$$= \frac{1}{h^m} \int K^2(u)du \cdot \int \left[V(\epsilon_1^2|x_1)^2\right] p^2(x_1)dx_1 + O(1/h^m)$$

$$= s^2/(2h^m) + o(1/h^m).$$

By the Hall’s theorem, we have $nh^{m/2}Q_{1n} \xrightarrow{d} \mathcal{N}(0, s^2)$.

Similarly to Zheng (1996), we can easily shown that $nh^{m/2}Q_{2n} = o_P(1)$, $nh^{m/2}Q_{3n} = o_P(1)$, $nh^{m/2}W_{2n} = o_P(1)$, and $nh^{m/2}W_{3n} = o_P(1)$. Thus we have proved that $nh^{m/2}W_{1n} \xrightarrow{d} \mathcal{N}(0, s^2)$. 

24
Next we show that \( \hat{s}^2 \) is a consistent estimator of \( s^2 \). Similarly to the above proof, we can show that

\[
\hat{s}^2 = \left\{ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{h^m} K^2 \left( \frac{x_i - x_j}{h} \right) \left( \epsilon_i^2 - \sigma^2 \right) \left( \epsilon_j^2 - \sigma^2 \right) \right\} + o_p(1)
\]

\( \equiv V_n + o_p(1) \),

where \( V_n \) is a \( U \)-statistic with kernel

\[
H_n(z_i, z_j) = \frac{2}{h^m} K^2 \left( \frac{x_i - x_j}{h} \right) \left( \epsilon_i^2 - \sigma^2 \right) \left( \epsilon_j^2 - \sigma^2 \right).
\]

By lemma 3.1 of Zheng (1996) and (A.4), \( V_n = E[H_n(z_i, z_j)] + o_p(1) = \hat{s}^2 + o_p(1) \). Thus \( \hat{s}^2 \rightarrow_p s^2 \).

(ii). First we can show that \( W_{1n} = U_{1n} + o_p(1) \) where \( U_{1n} \) is defined in (A.1) and is a \( U \)-statistic with kernel \( H_n \) given in (A.3). Under the alternative, since \( E[(\epsilon_i^2 - \sigma^2)|x_i] = V(\epsilon_i|x_i) - V(\epsilon_i) \), we have

\[
E[H_n(z_i, z_j)] = \frac{1}{h^m} \int K \left( \frac{x_i - x_j}{h} \right) \left[ V(\epsilon_i|x_i) - V(\epsilon_i) \right] \left[ V(\epsilon_j|x_j) - V(\epsilon_j) \right] p(x_i)p(x_j)dx_idx_j
\]

\[
= \frac{1}{h^m} \int K(u) \left[ V(\epsilon_i|x_i + hu) - V(\epsilon_i) \right] \left[ V(\epsilon_j|x_j) - V(\epsilon_j) \right] p(x_i + hu)p(x_j)h^mudu dx_j
\]

\[
= E \left\{ \left[ V(\epsilon|x - V(\epsilon) \right]^2 p(x) \right\} + o(1).
\]

By lemma 3.1 of Zheng (1996), \( W_{1n} = U_{1n} + o_p(1) = E[H_n(z_i, z_j)] + o_p(1) = E \left\{ \left[ V(\epsilon|x - V(\epsilon) \right]^2 p(x) \right\} + o_p(1) \). Similarly we can show that \( \hat{s}^2 \rightarrow_p s_1^2 \). Thus (ii) is proved.

(iii). Similarly to (i), we can show that \( W_{1n} = U_{1n} + o_p((nh^m)^{-1}) \). Let \( u_i^2 = \epsilon_i^2 - \delta_n \cdot g(x_i) \). Then \( E(u_i^2|x_i) = \sigma^2 \) under the local alternative (3.4). \( W_{1n} \) is
decomposed as

\[
W_{1n} = \left\{ \frac{1}{n(n - 1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K\left( \frac{x_i - x_j}{h} \right) \left( u_i^2 - \sigma^2 \right) \left( u_j^2 - \sigma^2 \right) \right\}
\]

\[+ \delta_n \cdot \left\{ \frac{1}{n(n - 1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K\left( \frac{x_i - x_j}{h} \right) \left( u_i^2 - \sigma^2 \right) g(x_j) \right\} \]

\[+ \delta_n^2 \cdot \left\{ \frac{1}{n(n - 1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K\left( \frac{x_i - x_j}{h} \right) g(x_i) g(x_j) \right\} \]

\[\equiv R_{1n} + \delta_n \cdot R_{2n} + \delta_n^2 \cdot R_{3n}. \]

Similarly to the proof of theorem 1, we can show that \(nh^{m/2}R_{1n} \rightarrow_d \mathcal{N}(0, s^2)\). By lemma 3.1 of Zheng (1996), we can easily show that \(\sqrt{n}R_{2n} = O_p(1)\). Similarly to (ii), we can show that \(R_{3n} \rightarrow_p E[g^2(x)p(x)]\).

If \(\delta_n = n^{-1/2}h^{-m/4}\), then \(nh^{m/2}\delta_n R_{2n} = h^m/4 \cdot \sqrt{n}R_{2n} \rightarrow_p 0\), and \(nh^{m/2}\delta_n^2 \cdot R_{3n} = R_{3n} \rightarrow_p E[g^2(x)p(x)]\). Thus \(nh^{m/2}W_{1n} \rightarrow_d \mathcal{N}\left(E[g^2(x)p(x)], s^2\right)\). Similarly to (ii), we can easily show that \(s^2 \rightarrow_p s^2\). Thus \(T_{1n} = nh^{m/2}W_{1n}/\hat{s} \rightarrow_d \mathcal{N}(\mu, 1)\) where \(\mu = E[g^2(x_i)p(x_i)]/s\).

Q.E.D.

**Proof of Theorem 2:**

(i) For notational simplicity, denote \(u_i \equiv (y_i - g(x_i))^2 - \sigma^2\), \(\epsilon_i \equiv y_i - g(x_i)\), \(g_i \equiv g(x_i)\), \(\hat{g}_i \equiv \hat{g}(x_i)\), and \(K_{ij} = K(\frac{x_i - x_j}{h})\). Then \(\hat{u}_i = (y_i - \hat{g}_i)^2 - \hat{\sigma}^2 = u_i - 2\epsilon_i(\hat{g}_i - g_i) + (\hat{g}_i - g_i)^2 - (\hat{\sigma}^2 - \sigma^2)\). Accordingly, \(W_{2n}\) is decomposed into ten parts:
\[ W_{2n} = \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} u_i u_j \right\} \\
+ 4 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} \epsilon_i \epsilon_j (\hat{g}_i - g_i)(\hat{g}_j - g_j) \right\} \\
+ \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} (\hat{g}_i - g_i)^2 (\hat{g}_j - g_j)^2 \right\} \\
+ \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} (\hat{\sigma}^2 - \sigma^2)^2 \right\} \\
- 4 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} u_i \epsilon_j (\hat{g}_j - g_j) \right\} \\
+ 2 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} u_i (\hat{g}_j - g_j)^2 \right\} \\
- 2 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} u_i (\hat{\sigma}^2 - \sigma^2) \right\} \\
- 4 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} \epsilon_i (\hat{g}_i - g_i)(\hat{g}_j - g_j)^2 \right\} \\
+ 4 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} \epsilon_i (\hat{g}_i - g_i)(\hat{\sigma}^2 - \sigma^2) \right\} \\
- 2 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^m} K_{ij} (\hat{g}_i - g_i)^2 (\hat{\sigma}^2 - \sigma^2) \right\} \\
\equiv U_{1n} + 4U_{2n} + U_{3n} + U_{4n} - 4U_{5n} + 2U_{6n} - 2U_{7n} - 4U_{8n} + 4U_{9n} - 2U_{10n}. \ (A.9) \]

Under the null, from theorem 1, \( nh^{m/2} U_{1n} \rightarrow_d N(0, s^2) \). To prove theorem 2, it remains to show that \( nh^{m/2} U_{2n} = o_p(1), \ldots, nh^{m/2} U_{10n} = o_p(1) \).

By theorem 3 of Collomb and Härdle (1986), we have

\[
\sup_{x \in C} |\hat{p}(x) - p(x)| = O_p(\sqrt{\frac{\ln n}{nh^m}}),
\]

27
Noting that

We shall show that

Thus

\[
E = \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h^m} K_{ij} \epsilon_i \epsilon_j (\hat{g}_i - g_i)(\hat{g}_j - g_j) \cdot \frac{\hat{p}_i}{p_i} \cdot \frac{\hat{p}_j}{p_j} \right\} \\
+ \left\{ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h^m} K_{ij} \epsilon_i \epsilon_j (\hat{g}_i - g_i)(\hat{g}_j - g_j) \cdot \frac{\hat{p}_i}{p_i} \cdot \frac{\hat{p}_j}{p_j} - 2 \left( \frac{\hat{p}_i}{p_i} \cdot \frac{\hat{p}_j}{p_j} \right) \right\}
\]

\[
\equiv Q_{1n} + o_p(1).
\]

We shall show that \(Q_{1n} = o_p((nh^{m/2})^{-1})\). Substituting kernel estimates \(\hat{g}\) and \(\hat{p}\) into \(Q_{1n}\), we have

\[
E[Q_{1n}^2] = E\left[ \frac{1}{n^3(n-1)h^6m} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{1}{p_ip_j} K_{ij} K_{jk} (y_k - g_i)(y_l - g_j) \epsilon_i \epsilon_j \epsilon_k \epsilon_l \right] \]

\[
= E\left[ \frac{1}{n^6(n-1)^2h^6m} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i,j}^{n} \sum_{l=1}^{n} \sum_{i'}^{n} \sum_{k'}^{n} \sum_{l'=1}^{n} \frac{1}{p_ip_jp_ip_j'} K_{ij} K_{ik} K_{jk'} K_{j'k'} (y_k - g_i)(y_l - g_j) \epsilon_i \epsilon_j \epsilon_k \epsilon_l \epsilon_{i'} \epsilon_{k'} \epsilon_{j'} \epsilon_{k'} \epsilon_{j'} \right] \\
+ o((n^2h^{-m})^{-1})
\]  

(A.12)

Noting that \(E[\epsilon_i \epsilon_j \epsilon_{i'} \epsilon_{j'}] \neq 0\) only if \(i = i', j = j'\) or \(i = j', j = i'\), we have

\[
E[Q_{1n}^2] = 2 \frac{2}{n^6(n-1)^2h^6m} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \frac{1}{p_ip_j^2} K_{ij}^2 K_{ik} K_{jk} K_{jk'} K_{j'k'} (y_k - g_i)(y_l - g_j) (g_{k'} - g_i)(g_{l'} - g_j) \sigma^4 \]

\[
\equiv o((n^2h^m)^{-1}).
\]  

(A.13)
By changing variables as \( \frac{x_i - x_j}{h} = u_1, \frac{x_i - x_k}{h} = u_2, \frac{x_i - x_l}{h} = u_3, \frac{x_i - x_m}{h} = u_4, \) and \( \frac{x_i - x_r}{h} = u_5 \), we have

\[
E[Q_{1n}^2] = \frac{2\sigma^4(n - 2)^2(n - 3)^2}{n^5(n - 1)h^{6m}} \int \frac{1}{p^2_p \rho^2_p} \frac{1}{K^2_{ij}K_{ik}K_{jl}K_{ik'j}K_{ij'}} \cdot (g_k - g_i)(g_l - g_j)(g_{k'} - g_i)(g_{l'} - g_j) + o((n^2 h^m)^{-1})
\]

\[
= \frac{2\sigma^4(n - 2)^2(n - 3)^2}{n^5(n - 1)h^{6m}} \int \int \int \int \int \int \frac{1}{p^2(x_i)p^2(x_i - hu_1)} \cdot K^2(u_1)K(u_2)K(u_3)K(u_4)K(u_5)
\cdot [g(x_i - hu_2) - g(x_i)]\{[g(x_i - hu_1 - hu_3) - g(x_i - hu_1)]
\cdot [g(x_i - hu_4) - g(x_i)]\{[g(x_i - hu_1 - hu_5) - g(x_i - hu_1)]
\cdot p(x_i)p(x_i - hu_2)p(x_i - hu_1 - hu_3)p(x_i - hu_4)p(x_i - hu_1 - hu_5)
\cdot h^{5m}dx_idu_1du_2du_3du_4du_5 + o((n^2 h^m)^{-1}).
\]  

(A.14)

By taking Taylor expansions of \( g(x_i - hu_1) - g(x_i) \) and similar terms around \( x_i \) and using assumptions 2B and 3B, we have

\[
E[Q_{1n}^2] = O(h^{5m}h^{4r}) = O(h^{4r}/n^2 h^{6m}).
\]  

(A.15)

By Chebyshev’s inequality, \( Q_{1n} = O_p(h^{2r}/n^{h^{m}/2}) \). Thus \( nh^{m/2}Q_{1n} = O_p(h^{2r}) = o_p(1) \). Thus we have proved \( nh^{m/2}U_{2n} = o_p(1) \).

For \( U_{3n} \), since

\[
|U_{3n}| \leq \frac{1}{n(n - 1)} \sum_{i=1}^{n} \sum_{j \neq i} h^{m} K_{ij} [\sup |\hat{g}(x) - g(x)|]^4
\]

\[
= O_p(1) \cdot O_p(\frac{(\ln n)^2}{n^2 h^{2m}}) = o_p(\frac{1}{nh^{m/2}}),
\]  

(A.16)

\[
h^{m/2}U_{3n} = o_p(1).
\]
Similarly to Hall and Marron (1990), we can easily show that $\sigma^2 = \sigma^2 + O_p(h^{2r})$. Thus $U_{4n} = O_p(h^4r) = o_p((nh^{m/2})^{-1})$, since $nh^{m/2}h^{4r} \to 0$.

$n h^{m/2} U_{5n} = o_p(1), \ldots, n h^{m/2} U_{10n} = o_p(1)$ can be proved in similar ways as above. Thus we have proved (i).

(ii) Following similar steps as (i) and (ii) of Theorem 1, (ii) is proved.

(iii) Following similar steps as (i) and (iii) of Theorem 1, (iii) is proved.

Q.E.D.

REFERENCES


### TABLE 1A
Efficiency of OLS Relative to GLS in Model 1: \( y = \beta_0 + \beta_1 x + x\epsilon, \ x \sim U(-1,1) \)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( n )</th>
<th>Bias( \times 10^4 )</th>
<th>SD( \times 10^2 )</th>
<th>RMSE( \times 10^2 )</th>
<th>RMSE(_{OLS} )</th>
<th>RMSE(_{GLS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>100</td>
<td>1.91</td>
<td>5.7231</td>
<td>5.7231</td>
<td>7.18</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>100</td>
<td>0.24</td>
<td>0.7975</td>
<td>0.7975</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>200</td>
<td>0.12</td>
<td>4.0830</td>
<td>4.0830</td>
<td>10.59</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>200</td>
<td>-0.99</td>
<td>0.3855</td>
<td>0.3856</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>300</td>
<td>-0.67</td>
<td>3.3277</td>
<td>3.3278</td>
<td>12.52</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>300</td>
<td>-0.10</td>
<td>0.2658</td>
<td>0.2658</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>400</td>
<td>3.87</td>
<td>2.8832</td>
<td>2.8834</td>
<td>14.22</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>400</td>
<td>-0.12</td>
<td>0.2028</td>
<td>0.2028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>500</td>
<td>-0.14</td>
<td>2.5991</td>
<td>2.5991</td>
<td>16.16</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>500</td>
<td>0.29</td>
<td>0.1608</td>
<td>0.1608</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimations of \( \beta_0 = 1 \)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( n )</th>
<th>Bias( \times 10^4 )</th>
<th>SD( \times 10^2 )</th>
<th>RMSE( \times 10^2 )</th>
<th>RMSE(_{OLS} )</th>
<th>RMSE(_{GLS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>100</td>
<td>-30.90</td>
<td>13.3859</td>
<td>13.3895</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>100</td>
<td>-25.42</td>
<td>10.0886</td>
<td>10.0918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>200</td>
<td>-14.16</td>
<td>9.4851</td>
<td>9.4862</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>200</td>
<td>-4.87</td>
<td>7.0759</td>
<td>7.0760</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>300</td>
<td>-3.26</td>
<td>7.7816</td>
<td>7.7817</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>300</td>
<td>-0.49</td>
<td>5.8270</td>
<td>5.8270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>400</td>
<td>4.94</td>
<td>6.7069</td>
<td>6.7071</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>400</td>
<td>11.51</td>
<td>5.0190</td>
<td>5.0203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>500</td>
<td>6.67</td>
<td>5.9681</td>
<td>5.9684</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>GLS</td>
<td>500</td>
<td>0.33</td>
<td>4.4702</td>
<td>4.4702</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimations of \( \beta_1 = 1 \)
TABLE 1B
Percentage of Rejections in Model 1: $y = 1 + x + x\epsilon$

<table>
<thead>
<tr>
<th>n</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>21.2</td>
<td>36.1</td>
<td>44.6</td>
<td>90.4</td>
<td>99.2</td>
<td>99.7</td>
<td>93.3</td>
<td>99.0</td>
<td>99.6</td>
</tr>
<tr>
<td>200</td>
<td>21.8</td>
<td>36.2</td>
<td>44.8</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>22.1</td>
<td>35.7</td>
<td>42.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>400</td>
<td>21.9</td>
<td>35.3</td>
<td>44.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>20.9</td>
<td>34.5</td>
<td>42.8</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>600</td>
<td>21.6</td>
<td>34.4</td>
<td>42.0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>700</td>
<td>20.7</td>
<td>36.5</td>
<td>44.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>800</td>
<td>19.2</td>
<td>33.3</td>
<td>41.6</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>900</td>
<td>23.0</td>
<td>35.5</td>
<td>44.1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1000</td>
<td>22.2</td>
<td>33.5</td>
<td>40.8</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

TABLE 2
Percentage of Rejections in Model 2: $y = 1 + x + e^{\lambda x}\epsilon$

<table>
<thead>
<tr>
<th>CV</th>
<th>\lambda</th>
<th>n</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>200</td>
<td>0.7</td>
<td>3.5</td>
<td>9.4</td>
<td>0.5</td>
<td>3.4</td>
<td>6.8</td>
<td>0.8</td>
<td>2.8</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>0.9</td>
<td>4.9</td>
<td>10.1</td>
<td>1.3</td>
<td>3.5</td>
<td>6.1</td>
<td>0.4</td>
<td>3.5</td>
<td>6.0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.447</td>
<td>200</td>
<td>99.2</td>
<td>99.5</td>
<td>99.8</td>
<td>81.8</td>
<td>90.7</td>
<td>94.5</td>
<td>81.1</td>
<td>91.5</td>
<td>95.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>96.8</td>
<td>99.1</td>
<td>99.5</td>
<td>97.3</td>
<td>99.5</td>
<td>99.8</td>
</tr>
<tr>
<td>1.0</td>
<td>0.961</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>90.6</td>
<td>98.5</td>
<td>99.4</td>
<td>92.2</td>
<td>99.1</td>
<td>99.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
## TABLE 3
Percentage of Rejections in Model 3: \( y = 1 + x + x^2 + (1 + \lambda x) \epsilon \)

<table>
<thead>
<tr>
<th>CV</th>
<th>( \lambda )</th>
<th>( n )</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>200</td>
<td>0.7</td>
<td>5.6</td>
<td>11.2</td>
<td>1.0</td>
<td>3.8</td>
<td>6.9</td>
<td>1.1</td>
<td>4.0</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>1.0</td>
<td>4.9</td>
<td>9.1</td>
<td>1.0</td>
<td>3.5</td>
<td>5.5</td>
<td>1.1</td>
<td>2.9</td>
<td>6.3</td>
</tr>
<tr>
<td>0.5</td>
<td>0.460</td>
<td>200</td>
<td>99.1</td>
<td>99.8</td>
<td>100</td>
<td>84.5</td>
<td>93.2</td>
<td>96.3</td>
<td>85.6</td>
<td>94.3</td>
<td>97.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98.8</td>
<td>99.7</td>
<td>99.9</td>
<td>97.8</td>
<td>99.5</td>
<td>99.8</td>
</tr>
<tr>
<td>1.0</td>
<td>1.331</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>96.8</td>
<td>99.8</td>
<td>100</td>
<td>97.5</td>
<td>99.3</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

## TABLE 4
Percentage of Rejections in Model 4: \( y = \exp(1 + x) + (1 + \lambda x^2) \epsilon \)

<table>
<thead>
<tr>
<th>CV</th>
<th>( \lambda )</th>
<th>( n )</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>200</td>
<td>1.4</td>
<td>4.8</td>
<td>9.1</td>
<td>1.1</td>
<td>4.0</td>
<td>5.9</td>
<td>0.9</td>
<td>3.3</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>1.0</td>
<td>5.1</td>
<td>11.0</td>
<td>1.0</td>
<td>3.5</td>
<td>6.1</td>
<td>1.7</td>
<td>3.8</td>
<td>6.1</td>
</tr>
<tr>
<td>0.5</td>
<td>1.151</td>
<td>200</td>
<td>10.2</td>
<td>21.7</td>
<td>29.6</td>
<td>74.4</td>
<td>88.6</td>
<td>93.8</td>
<td>76.0</td>
<td>89.1</td>
<td>93.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>11.0</td>
<td>21.2</td>
<td>29.6</td>
<td>95.8</td>
<td>98.6</td>
<td>99.3</td>
<td>95.2</td>
<td>97.8</td>
<td>99.0</td>
</tr>
<tr>
<td>1.0</td>
<td>4.530</td>
<td>200</td>
<td>26.2</td>
<td>41.6</td>
<td>50.1</td>
<td>99.9</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>26.7</td>
<td>39.2</td>
<td>47.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 1
Daily 7-day Eurodollar Rate
Figure 3
Estimated Drift Functions
Figure 4
Estimated Diffusion Functions

- CKLS Estimator
- Kernel Estimator
- Upper 95% Conf. Band
- Lower 95% Conf. Band

Diffusion Function vs. Spot Rate
Figure 5
Estimated Drift Functions

Drift Function vs. Spot Rate

- - - - Ait-Sahalia's Estimator

Kernel Estimator

Upper 95% Conf. Band

Lower 95% Conf. Band
Figure 6
Estimated Diffusion Functions
Figure 7
Estimated Diffusion Functions

- - - Semiparametric Estimator
- - - Kernel Estimator
- - - Upper 95\% Conf. Band
- - - Lower 95\% Conf. Band