A Specification Test for Time Series Models by a Normality Transformation

Jin-Chuan Duan*

(First Draft: May 2003)
(This Draft: Oct 2003)

Abstract

A correctly specified time series model can be used to transform data set into an *i.i.d.* sequence of standard normal random variables, assuming that the true parameter values are known. In reality, however, one only has an estimated model and must therefore address the sampling errors associated with the parameter estimates. This paper presents a new test that does not rely on specifying any specific alternative model. The test explores both normality and independence of the transformed sequence. Specifically, we utilize the theoretical properties of the transformed residuals to construct a set of four test statistics, and for which the sampling errors associated with any root-T consistent parameter estimates are eliminated. The size and power of this new test are examined. We find the size of this test to be accurate for dynamic models such as AR, GARCH and diffusion. The power of this test is also good in the sense that it can reject a mis-specified model using a reasonable sample size. The test is then applied to real data series of stock returns and interest rates.

Key Words: Size, Power, GARCH, Diffusion, Asymptotic distribution

---

*Duan is with Joseph L. Rotman School of Management, University of Toronto and is affiliated with CIRANO. E-mail: jcduan@rotman.utoronto.ca; Tel: 416-946 5653; Fax: 416-971 3048. The author acknowledges support received as the Manulife Chair in Financial Services and research funding from both the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. The author thanks Andras Fulop for his programming assistance and benefits from the valuable comments by Haitao Li, Atsushi Inoue and the participants of the CIRANO-CIREQ financial econometrics conference in May, 2003 and the NBER/NSF Time Series Conference in September 2003.

1 Introduction

Time series models are used in econometrics and statistics to describe data set recorded over a period of time. The most pressing question is arguably about whether the given time series model is a suitable specification for the data set. This paper provides a new test for addressing this question. A time series model is a complete specification of the law governing the evolution of a stochastic system that generates a data set recorded over time. It is quite common that the time series model uses a dynamic location-scale specification, meaning that the conditional mean and variance are specified as functions of the previous variables but the distribution of the standardized random variable is not. Examples of the dynamic location-scale model abound. The ARMA and GARCH models are such examples. There are also models that have a dependence structure beyond location and scale. A good example of this type is the mean-reverting square-root diffusion process (Feller process) that was adopted by Cox, *et al* (1985) to model interest rates. The conditional distribution over a discrete time period under this model is a non-central chi-square that cannot be reduced to a dynamic location-scale specification. For either case though, the assumed model has a common feature that can be exploited in designing a specification test. As opposed to relying on nesting the assumed model as is typically done, we propose in this paper a test that utilizes the model specification as a transformation devise and examines the logical consequences of such a transformation.

The classical way of testing a distribution assumption without relying on specific alternatives is the Kolmogorov-Smirnov test or its variants such as the Andersen-Darling test. Such a test measures the distance between the theoretical distribution under the assumed model and the empirical distribution function. Different tests have different ways of measuring the distance with an intent to detect the departure from the assumed distribution in certain dimensions. For time series models, the distribution assumption does not completely characterize the system simply due to the presence of dependence structure. The similar idea, nevertheless, applies. Diebold, *et al* (1998), Diebold, *et al* (1999), Bai (2002) and Hong and Li (2002) have proposed to transform the dependent data series using the conditional distribution function so as to obtain an *i.i.d.* sequence of uniformly distributed transformed residuals. One can then proceed with the Kolmogorov type test on the transformed data set. Diebold, *et al* (1998) and Diebold, *et al* (1999) have not dealt with the complex issue of parameter estimation uncertainty that inevitably accompanies the parameter estimate used in the conditional distribution function.[1] Bai (2002) and Hong and Li (2002), on the other hand, differ in their ways of dealing with parameter estimation uncertainty. Bai (2002) employs the Khmaladze (1981) martingale transformation to rid off the parameter estimation uncertainty. Hong and Li (2002) rely instead on a distance between a bivariate nonparametric kernel density estimate and the bivariate uniform density, which after normalization is not subject to the parameter estimation uncertainty.[2]

[1]The approach developed in Diebold, *et al* (1998) and Diebold, *et al* (1999) is more constructive in nature because the approach is meant to identify the correct conditional density function. Parameter estimation uncertainty is left unaddressed, however.

[2]Bai's (2002) test has not truly utilized the *i.i.d.* property of the transformed data set. In a way, it is similar to the Kolmogorov test that only tests the marginal distribution as opposed to the *i.i.d.* assumption. Hong and Li (2002), however, use the bivarate uniform distribution in constructing their test, which specifically takes into account

In the context of diffusion models, testing a specification without committing to a specific class of alternatives has also been studied. The first paper is, to our knowledge, Ait-Sahalia (1996a), in which the parametric density implied by the assumed diffusion model is compared to a density estimated nonparametrically. Such an approach is innovative in the sense that the parameter values are estimated by minimizing the distance between two densities where only the parametric density depends on the model parameters. The minimum distance itself (after normalization) serve as a test statistic to determine the adequacy of the assumed model. In other words, the parameter estimation and the test statistic are performed in one step. In contrast, the tests proposed by Bai (2002) and Hong and Li (2002) are of the two-step nature. First, some $\sqrt{T}$-consistent parameter estimate is used to transform the data set. Then, one creates a test statistic that specifically explores the properties of the transformed residual.[3] Although circumventing the parameter estimation uncertainty is desirable, it is not without costs; for example, it is not clear as to how one can move from measuring the distance between two marginal densities as in Ait-Sahalia (1996a) to that between two conditional densities without having obtained some parameter estimate first.[4] Restricting to marginal densities will clearly weaken the power of test because the method does not separate an *i.i.d.* data series from a dependent stationary series. As reported in Pritsker (1998), the Ait-Sahalia (1996a) test has a slow convergence rate for interest rate data, which results in significant over-rejections even using a fairly large data sample, a result mainly attributable to the high persistence of the interest rate process. Relying on marginal densities (or distributions) has another drawback. As noted in Corradi and Swanson (2002), such a test cannot distinguish two different models that share the same marginal density (or distribution).

The test proposed in this paper differs from the papers discussed thus far. Our test is constructed in three steps. First, we use the conditional distribution function under the assumed model to transform the time series of size T into an *i.i.d.* sequence of standard normal random variables. In this regard, the transformation is in spirit similar to Diebold, *et al* (1998), Diebold, *et al* (1999), Bai (2002) and Hong and Li (2002). Second, we partition the transformed residuals into many independent blocks of size n. Within each block of size n, we take advantage of normality to create a vector of n random variables, which is done by sequentially adding one variable at a time to a group of variables and then performing a nonlinear transformation. Third, the resulting *i.i.d.* sequence of n-dimensional random vectors is used to create a chi-square test statistic free of the root-T consistent parameter estimation error. The creation of the n-dimensional random vector via the block structure is for the purpose of eliminating the parameter estimation error by linear transformations. At the same time, the block structure allows the test statistic to exploit the

the *i.i.d.* feature of the transformed data set.

[3]Thompson (2002) proposes a two-step method as well. The parameter estimation uncertainty is dealt with by simulating the test statistic's distribution using the asymptotic distribution of the $\sqrt{T}$-consistent parameter estimate. Corradi and Swanson (2002) is another such kind of test being proposed in the literature. They use a shorter transformed series to construct the test statistic while relying on the parameter estimate from the longer data series as a way of "eliminating" the parameter estimation error. The cut-off value for the proposed test is then obtained by the bootstrap technique.

[4]It should be noted that Ait-Sahalia (1996a) has also developed a way of utilizing conditional density information that measures "transition discrepancy" albeit such a test has not been empirically implemented in that paper. The conditional version of the Ait-Sahalia (1996a) test is based on a theoretical observation that the time-derivative gap between the forward and backward equations for stationary diffusions should be zero.

independence nature of the transformed residuals under the assumed model. In this regard, the test shares the spirit of Hong and Li (2002) in utilizing the independence property. The test is, however, completely parametric and relies on asymptotic inference. In this regard, it shares the same feature of Bai's (2002) test. Similar to the Bai (2002) and Hong and Li (2002), the cut-off value for testing is obtained from a known distribution. Thus, there is no need to perform computer intensive simulation or bootstrapping to determine the cut-off value as required by Thompson (2002) or Corradi and Swanson (2002).

It is worth noting that the proposed test does not require specifying an alternative class of models to nest the assumed model. In this sense, the proposed testing method along with the ones discussed thus far extends the line of the Kolmogorov test. Thus, they are subject to the same weakness and enjoy the similar benefit. A test without specifying a specific class of alternatives is expected to sacrifice some power of rejection, but it does not suffer the type-II error associated with specifying a wrong class of alternative models. Different ways of constructing specification tests without alternatives are likely to have different powers of rejection. Intuitively though, the test utilizing the independence structure of the transformed residuals can be expected to have a higher power. In this sense, our proposed test is intuitively appealing.

This new test actually consists of four test statistics with an intent to examine four aspects of the transformed residual. We conduct a power analysis using two classes of models - AR-GARCH and diffusion. We use several special cases of the AR-GARCH model to generate data sets and test the constant mean-variance and normality specification. The sizes of the four test statistics are all found to be accurate for a sample size as small as 200. The power of the test is also quite satisfactory, and rejection occurs on different elements of the four test statistics as expected and depending on which aspect of the generated data deviates from the assumed model. We also generate three versions of the mean-reverting constant-elasticity-of-variance diffusion models to simulate the interest rate data. We then study the size and power of testing the Vasicek (1977) specification for interest rates, which is one of the three versions used to generate the data sets. The sizes of the four test statistics are accurate for 500 data points (daily interest rates over two years), which is the smallest sample size studied for the diffusion model. For the sample size of $2,500$ (daily interest rates over five years), the test has a moderate power of rejecting the Vasicek (1977) specification if the data set is generated by the Cox, *et al* (1985) model. In the case of the Chan, *et al* (1992) specification, the power of rejecting the Vasicek (1977) model is very high.

We also apply the test to the real data of the S&P500 index returns (daily). The results indicate that the GARCH(1,1) model with conditional normality fails to pass the test for each of two subsamples but the GARCH(1,1) model with a conditional t-distribution only fails to pass for one of the two subsamples. When the test is applied to the Eurodollar deposit rate data (daily), we find that both the Vasicek (1977) and Cox, *et al* (1985) specifications are rejected resoundingly for all subsamples.

2 Specification test by a normality transformation

Consider a time series $\{X_t : t = 1, 2, \cdots\}$ and define $\mathcal{F}_{t-1}$ to be the σ-field generated by $\{X_\tau : \tau \leq t-1\}$ and all exogenous stochastic variables observable up to time t. Let $G_{t-1}(X_t; \theta)$ be the distribution function of X_t, conditional on $\mathcal{F}_{t-1}$, where θ denotes the model parameter(s). We maintain an assumption that X_t has a continuous conditional distribution function and is differentiable almost everywhere.

For typical dynamic location-scale models, such as ARMA and GARCH, $G_{t-1}(X_t;\theta)$ is simplified to some distribution function, say for example normal or student-t, and only the conditional mean and standard deviation are used to reflect the dependence structure. For such cases, the model is more commonly expressed as $X_t = f_t(\mu) + v_t(\sigma)\varepsilon_t$, where $f_t(\mu)$ and $v_t(\sigma)$ are measurable with respect to $\mathcal{F}_{t-1}$ and ε_t's are random variables with mean 0, variance 1 and a common distribution function with parameter η (or no additional parameter at all). Thus, $\theta = (\mu, \sigma, \eta)$. This dynamic location-scale setup clearly encompasses typical time-series regression models with exogenous variables.

Our specification also includes the data series being sampled discretely from a diffusion model. For some diffusion model, such as the Ornstein-Uhlenbeck process that is commonly used to model interest rates, the discretely sampled data series is exactly governed by a dynamic location-scale model. For the mean-reverting square-root process (Feller process), on the other hand, the discretely sampled data series can no longer be described as a dynamic location-scale model. Nevertheless, the conditional distribution exists, and in particular, $G_{t-1}(X_t;\theta)$ is a non-central chi-square with the non-centrality parameter depending on X_{t-1}. For more complex diffusion models, one may need to use a method such as the expansion idea of Ait-Sahalia (2002) to obtain an approximate closed-form expression for $G_{t-1}(X_t;\theta)$.

Let k_θ be the number of parameters in θ. Denote by $\hat{\theta}_T$ some $\sqrt{T}$-consistent estimator for the true parameter value θ_0. The standard normal distributional function is denoted by $\Phi(z)$. Define

$$\xi_t(\theta) = \Phi^{-1}\left[G_{t-1}\left(X_t;\theta\right)\right]. \tag{1}$$

It is clear that $\xi_t(\theta_0)$ forms an *i.i.d.* sequence of standard normal random variables, but $\xi_t(\theta)$ does not in general. We further transform $\xi_t(\theta)$ with the intention of eventually utilizing the fact that $\xi_t(\theta_0)$'s are *i.i.d.* standard normal random variables.[5] Let

$$q_{m,i}^{(p)}(\theta) = \sum_{j=1}^{m} \xi_{(i-1)*m+j}^{p}(\theta) \quad \text{for } p = 1,2 \tag{2}$$

$$q_{m,i}^{(3)}(\theta) = \frac{1}{m}\left(\sum_{j=1}^{m} \xi_{(i-1)*m+j}(\theta)\right)^2 \tag{3}$$

$$q_{m,i}^{(4)}(\theta) = \frac{1}{m^2}\left(\sum_{j=1}^{m} \xi_{(i-1)*m+j}^{2}(\theta) - m\right)^2. \tag{4}$$

Use $R_m^{(p)}(\cdot)$ to denote the distribution function for $q_{m,i}^{(p)}(\theta_0)$. It is clear that $R_m^{(1)}(\cdot)$ is a normal distribution function with mean 0 and variance m. It is also clear that $R_m^{(2)}(\cdot)$ is the chi-square

[5] Transforming the observed variable into a standard normal random variable is not a theoretical necessity. Without this normality transformation, however, we cannot take advantage of the analytical convenience associated with combining normal random variables on which the test construction heavily depends.

distribution with m degrees of freedom. Similarly, $R_m^{(3)}(\cdot)$ is the chi-square distribution with 1 degree of freedom and $R_m^{(4)}(x) = R_m^{(2)}\left[m\left(1+\sqrt{x}\right)\right] - R_m^{(2)}\left[m\left(1-\sqrt{x}\right)\right]$. Note that $R_m^{(p)}(\cdot)$ is not model-specific.

For an integer $m \geq 1$ and $i = 1, 2, \cdots, [T/m]$, we define

$$Y_{m,i}^{(p)}(\theta) = R_m^{(p)}\left(q_{m,i}^{(p)}(\theta)\right) - \frac{1}{2} \quad \text{for } p = 1, 2, 3, 4. \tag{5}$$

Thus, $Y_{m,i}^{(p)}(\theta_0)$ (for $i = 1, 2, \cdots$) forms an *i.i.d.* sequence of uniform (over $\left[-\frac{1}{2}, \frac{1}{2}\right]$) random variables for any $m \geq 1$ and $p \in \{1, 2, 3, 4\}$. Our set of test statistics is based on the following constructed variables:

$$Z_{m,T}^{(p)}(\theta) = \frac{1}{\sqrt{m}[T/m]} \sum_{i=1}^{[T/m]} Y_{m,i}^{(p)}(\theta), \text{ for } p = 1, 2, 3, 4. \tag{6}$$

It is clear that $\sqrt{T} Z_{m,T}^{(p)}(\theta_0)$ converges to a normal random variable with mean 0 by the Central Limit Theorem and the asymptotic variance can also be easily computed. Unfortunately, we do not know θ_0 and need to evaluate $Z_{m,T}^{(p)}(\theta)$ at some parameter estimate $\hat{\theta}_T$. Later, we devise the test statistic corresponding to any given p, which relies on using some linear combinations of $Z_{m,T}^{(p)}(\hat{\theta}_T)$ for different values of m to remove the sampling error associated with $\hat{\theta}_T$.

We now make a usual assumption about a $\sqrt{T}$-consistent parameter estimator.

Assumption 1. The parameter estimator $\hat{\theta}_T$ for the time series $\{X_t : t = 1, 2, \cdots\}$ governed by the conditional distribution function $G_{t-1}(X_t; \theta_0)$ satisfies:

1. $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) = O_p(1)$.
2. $\frac{\partial Z_{m,T}^{(p)}(\hat{\theta}_T)}{\partial \theta'}$ converge in probability to a constant $1 \times k_\theta$ vector for any finite $m \geq 1$ and $p = 1, 2, 3, 4$.

By Assumption 1, we can apply the Taylor expansion to $Z_{m,T}^{(p)}(\hat{\theta}_T)$ to obtain

$$\sqrt{T} Z_{m,T}^{(p)}(\hat{\theta}_T) = \sqrt{T} Z_{m,T}^{(p)}(\theta_0) + \frac{\partial Z_{m,T}^{(p)}(\theta_0)}{\partial \theta'} \sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1). \tag{7}$$

The term $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)$ does not vanish because it converges to a proper random vector as T approaches infinity. This $\sqrt{T}$-consistent estimator carries with it a parameter estimation uncertainty. In order to have a test that has a correct size, one must address this parameter estimation uncertainty. The above Taylor expansion serves as the basis for us to construct a test statistic without subjecting to the asymptotic distribution of the parameter estimate.

We now define a variance-covariance matrix of the limiting random variables, $\lim_{T\to\infty} \sqrt{T} Z_{m,T}^{(p)}(\theta_0)$ for different m's, which will be proved to be the case later in Theorem 1. Denote this matrix corresponding to $\{m = 1, 2, \cdots, n\}$ by $\mathbf{A}^{(p)}_{n\times n}$ and its (i,j)-element of this matrix is

$$a^{(p)}_{ij} = \frac{\sqrt{ij}}{\kappa(i,j)} E\left(\sum_{k=1}^{\kappa(i,j)/i} W^{(p)}_{i,k} \sum_{l=1}^{\kappa(i,j)/j} W^{(p)}_{j,l}\right) \tag{8}$$

where

$$W^{(p)}_{m,l} = R^{(p)}_m\left(\sum_{j=1}^{m} \epsilon^{p}_{(i-1)*m+j}\right) - \frac{1}{2} \text{ for } p = 1, 2 \tag{9}$$

$$W^{(3)}_{m,l} = R^{(3)}_m\left(\frac{1}{m}\left(\sum_{j=1}^{m} \epsilon_{(i-1)*m+j}\right)^2\right) - \frac{1}{2} \tag{10}$$

$$W^{(4)}_{m,l} = R^{(4)}_m\left(\frac{1}{m^2}\left(\sum_{j=1}^{m} \epsilon^{2}_{(i-1)*m+j} - m\right)^2\right) - \frac{1}{2} \tag{11}$$

$\kappa(i,j)$ is the lowest common multiple of i and j, and $\{\epsilon_t; t = 1, 2, \cdots\}$ is an *i.i.d.* sequence of standard normal random variables. Note that $\mathbf{A}^{(p)}$ does not depend on parameter value. The diagonal elements of $\mathbf{A}^{(p)}$ are always the same and can be computed analytically to yield $a^{(p)}_{ii} = \frac{1}{12}$ for all i's. The off-diagonal elements need to be assessed numerically and can, for example, be computed by Monte Carlo simulation. Matrices $\mathbf{A}^{(p)}_{10\times 10}$ for $p = 1, 2, 3$ and 4 are presented in Appendix B for which one million simulation paths were used. These calculations need not be repeated because these matrices do not depend on a specific model.

The main result for the test statistic is stated in the following theorem.

Theorem 1. Assume that $\mathbf{A}^{(p)}_{n\times n}$ is invertible and maintain Assumption 1. Let $\mathbf{A}^{(p)1/2}_{n\times n}$ denote its Cholesky decomposition (defined as a lower triangular matrix in this paper) and $\|\cdot\|$ be the Euclidean norm. Let $\mathbf{B}^{(p)}_{n\times k_\theta}(\theta_0)$ consist of the limiting vectors $\frac{\partial Z^{(p)}_{m,T}(\hat{\theta}_T)}{\partial \theta'}$ for $m = 1, 2, \cdots, n$, and r denote the column rank of $\mathbf{B}^{(p)}_{n\times k_\theta}(\theta_0)$. Then, for $p \in \{1, 2, 3, 4\}$ and $n > r$, there exists $\alpha^{(p)}_{k\times n}$ that solves

$$\alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\mathbf{B}^{(p)}_{n\times k_\theta}(\theta_0) = \mathbf{0}_{k\times k_\theta} \tag{12}$$

$$\alpha^{(p)}_{k\times n}\alpha^{(p)\prime}_{k\times n} = \mathbf{I}_{k\times k}, \tag{13}$$

and

$$J^{(p)}_T(\hat{\theta}_T) = T\left\|\alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\begin{bmatrix} Z^{(p)}_{1,T}(\hat{\theta}_T) \\ \vdots \\ Z^{(p)}_{n,T}(\hat{\theta}_T)\end{bmatrix}\right\|^2 \xrightarrow{D} \chi^2(k). \tag{14}$$

where $k = n - r$.

Proof: see Appendix A

Although $\alpha^{(p)}$ always exists, the solution is not unique. Non-uniqueness is due to rotations. A given solution can always be rotated by pre-multiplying $\alpha^{(p)}$ with a unitary matrix. This non-uniqueness is inconsequential because the test statistic $J_T^{(p)}(\hat{\theta}_T)$ is invariant to the operation of pre-multiplying by a unitary matrix. In other words, all solutions produce the same value for the test statistic. To solve the system defined by equations (12) and (13), one can use the standard computer routine to find the orthonormal basis $\alpha^{(p)}$ for the null space defined by equation (12); for example, Matlab offers a procedure "Null" for this task.

Matrix $\mathbf{B}_{n\times k_\theta}^{(p)}(\theta_0)$ can be computed analytically in some cases. If, for example, the assumed model is a constant mean μ and variance σ^2 with a normal distribution, then one can show that for $p=1$, $\lim_{T\to\infty}\frac{\partial Z_{m,T}^{(1)}(\mu_0,\sigma_0)}{\partial\mu}=-\frac{1}{2\sigma_0\sqrt{\pi}}$ and $\lim_{T\to\infty}\frac{\partial Z_{m,T}^{(1)}(\mu_0,\sigma_0)}{\partial\sigma}=0$. Similarly for $p=2$, we have $\lim_{T\to\infty}\frac{\partial Z_{m,T}^{(2)}(\mu_0,\sigma_0)}{\partial\mu}=0$ and $\lim_{T\to\infty}\frac{\partial Z_{m,T}^{(2)}(\mu_0,\sigma_0)}{\partial\sigma}=-\frac{2}{\sqrt{m}\sigma_0}\int_0^\infty z\left[h\left(z;m\right)\right]^2dz$ where $h\left(\cdot;m\right)$ is the chi-square density function with m degrees of freedom. Unlike $\mathbf{A}^{(p)}$, $\mathbf{B}_{n\times k_\theta}^{(p)}(\theta_0)$ is model-specific and is a function of the parameter. Since we do not know the true value, we need to use $\hat{\theta}_T$ instead. In other words, we use $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)$ in place of $\mathbf{B}_{n\times k_\theta}^{(p)}(\theta_0)$. In general, the analytical expression for $\mathbf{B}_{n\times k_\theta}^{(p)}(\theta_0)$ may be too complex to derive, but can always be approximated by generating a simulated sample under the assumed model.

Before proceeding with finding the null space's orthonormal basis, we need to recognize a problem posed by the sampling error of $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)$. If a particular column of $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)$ is theoretically linearly dependent on others, it might not be so numerically simply due to sampling errors. In other words, we may have a theoretical rank that is less than k_θ, but numerically it appears to have a rank equal to k_θ. When the "genuine" rank is less than the rank in appearance, the null space will have a dimension less than its true dimension. Failing to find the "genuine" rank does not affect the size of test but may reduce its power. This is easily understood by an example. Consider the case that $n=4$, $k_\theta=3$ and the true column rank equals 2, i.e., there is one linearly dependent column in $\mathbf{B}_{4\times 3}^{(p)}(\theta_0)$. Due to sampling errors, its rank appears to be 3 numerically, and thus the resulting null space has the dimension equal to 1. The orthonormal basis $\alpha^{(p)}$ thus has only one row, but it could have two rows in the absence of sampling errors. The degrees of freedom of the test statistic rightly reflects the number of rows in $\alpha^{(p)}$. The sampling error has, however, unduly restricted the test statistic to be constructed in the one-dimensional subspace of the "genuine" two-dimensional null space. As a result, the power of the test may be adversely affected.

To address this issue, it is preferable to set a reasonable tolerance level so as to uncover the "genuine" rank by factoring in the approximate linear dependency. A reasonable tolerance level must take into account the sampling error for $\hat{\theta}_T$. Consider, for example, a parameter with a small sampling error. A small deviation from linear dependency may actually indicate a true linear independence. In order to set a uniform tolerance level, we consider a system that is equivalent to equations (12) and (13):

$$\alpha_{k\times n}^{(p)}\mathbf{A}_{n\times n}^{(p)-1/2}\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)\mathbf{V}_{k_\theta\times k_\theta}^{1/2}\left(\hat{\theta}_T\right) = \mathbf{0}_{k\times k_\theta} \quad (15)$$

$$\alpha_{k\times n}^{(p)}\alpha_{k\times n}^{(p)\prime} = \mathbf{I}_{k\times k}, \quad (16)$$

where $\mathbf{V}_{k_\theta\times k_\theta}\left(\hat{\theta}_T\right)$ is the asymptotic variance-covariance matrix for $\sqrt{T}\left(\hat{\theta}_T-\theta_0\right)$ under the assumed model and the superscript 1/2 denotes its Cholesky decomposition (defined as a lower triangular matrix). The above system is motivated by the fact that the asymptotic variance-covariance matrix for $\sqrt{T}\mathbf{V}_{k_\theta\times k_\theta}^{-1/2}\left(\hat{\theta}_T\right)\left(\hat{\theta}_T-\theta_0\right)$ is an identity matrix. Note that $\mathbf{A}_{n\times n}^{(p)-1/2}\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)\mathbf{V}_{k_\theta\times k_\theta}^{1/2}\left(\hat{\theta}_T\right)$ is basically the multiplying matrix used to determine how much sampling errors of $\sqrt{T}\mathbf{V}_{k_\theta\times k_\theta}^{-1/2}\left(\hat{\theta}_T\right)\left(\hat{\theta}_T-\theta_0\right)$ gets transmitted to the test statistic. Setting a tolerance level for $\mathbf{A}_{n\times n}^{(p)-1/2}\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)\mathbf{V}_{k_\theta\times k_\theta}^{1/2}\left(\hat{\theta}_T\right)$ amounts to applying a uniform maximum allowance level for the sampling errors to impact the test statistic.

Since we do not know whether the assumed model is the one that generates the data set, we need to compute $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)$ and $\mathbf{V}_{k_\theta\times k_\theta}\left(\hat{\theta}_T\right)$ directly by theory or approximate them numerically using a simulated sample under the assumed model and the estimated parameter value $\hat{\theta}_T$. The results reported in this paper are all based on using the simulated data to compute these two matrices. The size of the simulated sample is set equal to the size of the original data sample to be tested. Alternatively, one can avoid simulation by sticking to the original data set. Using the original data set to compute these two matrices will not affect the size of the test, but it can affect the power when the true model deviates from the assumed model.

We perform the singular-value decomposition on $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)\mathbf{V}_{k_\theta\times k_\theta}^{1/2}\left(\hat{\theta}_T\right)$ to determine the "genuine" rank of $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)$. We set to zero all singular values that are smaller than 0.01 and then reconstitute the matrix. The reconstituted matrix, denoted by $\mathbf{P}_{n\times k_\theta}^{(p)}$, is approximately equal to $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\hat{\theta}_T\right)\mathbf{V}_{k_\theta\times k_\theta}^{1/2}\left(\hat{\theta}_T\right)$, and presumably carries with it the true rank of $\mathbf{B}_{n\times k_\theta}^{(p)}\left(\theta_0\right)$. We then proceed to find the orthonormal basis $\alpha^{(p)}$ for the null space defined by equations

$$\alpha_{k\times n}^{(p)}\mathbf{A}_{n\times n}^{(p)-1/2}\mathbf{P}_{n\times k_\theta}^{(p)} = \mathbf{0}_{k\times k_\theta} \quad (17)$$

$$\alpha_{k\times n}^{(p)}\alpha_{k\times n}^{(p)\prime} = \mathbf{I}_{k\times k}. \quad (18)$$

For the results presented later, we have chosen to fix the degrees of freedom to a particular value of k as opposed to setting n. This can be easily accomplished using the following procedure. First we tentatively set $n=k_\theta+k$, and then check the dimension of the null space. If it equals k, then stop. Otherwise, we reduce n by 1 and repeat the same check. It is not difficult to see that this procedure guarantees the dimension of the final null space equal to k.

3 Power analysis

The four statistics ($p = 1, 2, 3$ and 4) test different dimensions of the model specification. For $p = 1$ and 2, we examine the transformed residuals (*i.i.d.* standard normally distributed residuals under the assumed model) to see whether their mean and variance are correctly specified. The test statistic for $p = 3$ checks to see whether the transformed residuals are autocorrelated. In the case of $p = 4$, we test whether the squared transformed residuals are autocorrelated. In an intuitive way, the four test statistics offer more than just a formal statistical statement of rejection/no rejection. They actually reveal the nature of model misspecification. For example, when a data set generated by a symmetric fat-tailed distribution is mistakenly modeled as a normal distribution, the transformed residual is effectively "stretched" so that its variance becomes larger than the predicted value under the assumed model. Similarly, if a data set generated by a symmetric distribution is erroneously modeled as an asymmetric distribution, the transformed residual will have a mean distorted away from the value predicted by the assumed model.

In this section, we conduct a power analysis using two popular classes of models: AR-GARCH and diffusion models. Two specific null hypotheses are tested with various alternative models used to generate the data sets. In all cases, we set the degrees of freedom of the test statistic to 2 and tabulate the rejection rate using 500 simulations.

3.1 AR-GARCH models

Consider the following model:

$$X_t = \mu + \gamma X_{t-1} + \sigma_t \varepsilon_t \tag{19}$$

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \varepsilon_{t-1}^2 \tag{20}$$

where ε_t's are *i.i.d.* mean 0 and variance 1 continuous random variables with a t-distribution function of η degrees of freedom. For the power analysis, we test special cases of this model without assuming the knowledge that the above model is used to generate the data set. A model to be tested has the relevant parameter θ that is a subset of $(\mu, \gamma, \beta_0, \beta_1, \beta_2, \eta)$. In order to compute the test statistic, we only need to estimate the restricted model. Specifically, we use the maximum likelihood estimator $\hat{\theta}_T$ for the restricted model.

Case 1: The assumed model: constant mean and variance with normality, i.e., $\gamma = 0$, $\beta_1 = 0$, $\beta_2 = 0$ and $\eta = \infty$.

Under the assumed model, we have $\theta = \{\mu, \sigma\}$ where $\sigma = \sqrt{\beta_0}$. We can use the sample mean and standard deviation of $\{X_1, X_2, \cdots, X_T\}$. We consider all four test statistics: $J_T^{(p)}(\hat{\theta}_T)$, $p = 1, 2, 3, 4$ and study the size and power for each of them.

We set the stationary mean and variance to 0 and 1, respectively, in generating the data sets. For the power analysis we alter the value for γ, β_2 and η individually. When $\gamma \neq 0$, it is the constant mean assumption being violated. Similarly, $1/\eta \neq 0$ implies a violation of normality. In the power analysis with respect to γ, we control for both the level and volatility of the process. Specifically, we keep $\mu/(1-\gamma)$ and $\sigma^2/(1-\gamma^2)$ constant where the two formulas are well known results for the AR(1) model. For the power analysis of stochastic volatility, i.e., varying β_2, we set out to maintain the same overall level of volatility when generating the ARCH(1) model. If the value of β_0 is fixed, an increase in β_2 will not only generate stochastic volatility but also cause the overall level of volatility to rise. To control for the volatility level, we set $\bar{\sigma}^2 = 1$ and maintain $\beta_0 = \bar{\sigma}^2 (1 - \beta_1 - \beta_2)$ when the value β_2 is varied.[6] In generating data, we also set $\sigma_1^2 = \bar{\sigma}^2$, meaning that the initial data point has the average volatility.

Tables 1.a-1.c present the rejection rates (5% test, 2 degrees of freedom) for various parameter values and three sample sizes (200, 500 and 1000). The size in all cases are fairly accurate. When the data generating model is the AR(1) process, the assumed model (normality with constant mean and variance) is rejected mainly for $p = 3$ (Table 1.a). This result is hardly surprising because the transformed residuals should be autocorrelated. If the data is generated by a t-distribution, we have strong rejection for $p = 2$ (Table 1.b). As discussed in the beginning of Section 3, we expect rejection under $p = 2$ for this case. This is true because the transformed residual is effectively "stretched" so that its variance becomes larger than the predicted value under the assumed model. Finally in Table 1.c, we see rejection under $p = 2$ and 4 when the data set is generated using the ARCH(1) model. Having rejection under $p = 4$ is completely expected because the squared transformed residuals are autocorrelated. As to rejection under $p = 2$, the reason is similar to the case of t-distributed data. The ARCH(1) model with conditional normality makes the marginal distribution exhibit a fat-tailed feature, and as a result, produces the "stretching" effect.

3.2 Diffusion models

We limit our power study to the class of mean-reverting constant-elasticity-of-variance diffusion models:

$$dX_t = \kappa(\mu - X_t)dt + \sigma X_t^{\delta} dW_t \tag{21}$$

where W_t is a Wiener process. We simulate the data set by approximating the above diffusion model with 20 subintervals within one recording data interval. In other words, we record once every 20 simulated data points to mimic a discretely sampled data set under the diffusion assumption. The Milstein scheme is used to generate the data set.

The assumed model is the Ornstein and Uhlenbeck process, a special case of equation (21) by setting $\delta = 0$. The parameters of this model can be estimated by maximum likelihood or other $\sqrt{T}$-consistent estimation methods like GMM. The results presented below are based on the maximum likelihood estimates.

Three special versions of the diffusion model in equation (21) are used to generate the data for testing. The first model is the Ornstein-Uhlenbeck process, i.e., $\delta = 0$. Its use for modeling interest rates was popularized by Vasicek (1977) in which this process plays a pivotal role in deriving the Vasicek bond pricing model. For ease of discussion, we refer to it as the Vasicek model. Next, we use the mean-reversion square-root diffusion (Feller process), i.e., $\delta = 1/2$. This process was used by Cox, Ingersoll and Ross (1985) to obtain the well-known CIR bond pricing model. Again for ease of discussion, we refer to it as the CIR model. Finally, the version established in Chan, *et al* (1992) is referred to as the CKLS model, which reports an estimate of δ equal to 1.4999.

[6] This is due to the fact that the stationary volatility equals $\bar{\sigma}^2 = \beta_0 / (1 - \beta_1 - \beta_2)$

Case 2: The assumed model: the Vasicek model, i.e., $\delta = 0$.

The data sets used in the power analysis are generated by three different models. First, we generate the Vasicek model using the following set of parameter values: $(\kappa, \mu, \sigma^2) = (0.85837, 0.089102, 0.002185)$, which was used in Pritsker (1998) and was in turn taken from Ait-Sahalia (1996b). The second model is the CIR model with the following set of parameter values: $(\kappa, \mu, \sigma^2) = (0.89218, 0.090495, 0.032742)$. Again they were used in Pritsker (1998) and Ait-Sahalia (1996b). These two sets of parameter values are based on the unit of time being one year. For the daily frequency, we thus divide both κ and σ^2 by 252. The final model is the CKLS specification with the other parameter values reported in Chan, *et al* (1992): $(\kappa, \mu, \sigma^2) = (0.5921, 0.0689, 1.6704)$. These parameter values correspond to one month as the basic unit of time. Thus, we divide both κ and σ^2 by 21.

Table 2 indicates that the size of test (for $p = 1, 2, 3$ and 4) is correct for the sample size as small as 500, a sample size that is roughly equal to daily observations over two years. This result stands in sharp contrast to that of the Ait-Sahalia test as discussed in Pristker (1998), where he argues that the Ait-Sahalia test needs to have an exceedingly large sample of interest rate data in order to have a right size.

The power of rejecting the Vasicek model mainly resides with the test under $p = 4$, which checks to see whether the squared transformed residuals are autocorrelated. Assuming the Vasicek model while the data is generated by the CIR model induces autocorrelation in the conditional variance because these two models only differ in how the diffusion term is specified. A similar argument applies to the data set generated by the CKLS model.[7] In order to have a moderate power (50%) of rejecting the Vasicek model when the data is generated by the CIR model, one needs to have about 10 years worth of daily interest rate data. It is, however, much easier to reject the Vasicek model (in excess of 90%) if the data is generated by the CKLS model, a result that is hardly surprising.

4 Application to real data

Two real data series are now considered. The first series is the daily S&P500 index returns (total return index from April 16, 1993 to April 16, 2003, continuously compounded) extracted from Datastream whereas the second series is the 7-day Eurodollar deposit spot rates on a daily frequency from June 1, 1973 to February 25, 1995, a data set used in Ait-Sahalia (1996a). We present the parameter estimates under the assumed models as well as the testing results using the four test statistics for each of the assumed models. Similar to the power analysis, we set the degrees of freedom of the test statistic to 2.

[7] When the data set is generated by the CKLS model, the power of test ($p = 4$) does not necessarily increase with the sample size. For example, it drops from 76.8% to 41.4% when the sample size is increased from 500 to $1,000$. This result is due to sometimes obtaining the AR(1) coefficient estimate greater than 1 when the sample size equals 500. It leads to an explosive $\mathbf{B}^{(p)}_{n \times k_\theta}\left(\hat{\theta}_T\right)$ and thus a rejection.

4.1 S&P500 index returns

The equity market index return is commonly modeled by the GARCH model in the empirical finance literature. For the S&P500 index return series, we consider two cases under the GARCH model. The results for the whole sample (2520 data points) and two subsamples (938 and 1582 data points) are provided. The whole sample is divided on December 31, 1996, because inspection of the return time series plot reveals a clear structural break. The maximum likelihood parameter estimates are used to compute the test statistics. The parameter estimates obtained for this model, reported in the bottom panel of Table 3.a, are similar to the typical results reported in the literature.

The first model applied to the S&P500 index return series is the linear GARCH(1,1) model with a conditional normal distribution. The results are presented in Table 3.a. The model is not rejected (at the 5% significance level) using the whole data sample but rejected for either subsample. Rejection takes place either with $p = 2$ or 4, meaning that the transformed residual has a variance different from the one predicted by the assumed model or the squared transformed residuals are autocorrelated. Given the extensive evidence supporting conditional leptokurtosis in the empirical finance literature, rejection ($p = 2$) for the second subsample is expected. For the first subsample, rejection ($p = 4$) suggests that the GARCH(1,1) variance dynamic is likely misspecified. What is surprising is our failure to reject the GARCH(1,1)-normality model using the whole sample. Perhaps, the structural break has mixed together two different kinds of misspecification present in the two subsamples.

The second model considered for the S&P500 index return series is the GARCH(1,1) model with a conditional t-distribution and the results are presented in Table 3.b. One can view this model as a natural progression from the GARCH model with conditional normality. We can still reject the model at the 5% level for the second subsample, and rejection continues to occur with $p = 2$, indicating the transformed residual still has a variance different from that predicted by the assumed model. Since the assumed model has allowed for leptokurtosis, rejection may indicate a problem with using t-distribution to model the data in the second subsample.

4.2 Eurodollar deposit rates

Arguably the most popular models used for describing interest rates are two specific diffusion models that we considered in Section 3.2: the Vasicek and CIR models. Their popularity has much to do with the fact that both models yield a bond pricing model that is exponential-affine and analytically convenient. We now study their performance in describing the Eurodollar deposit rate series. We run the four tests for the whole series with $5,505$ data points as well as five subsamples of $1,000$ data points for both models.

The results for the Vasicek model are presented in Table 4.a. Since the Vasicek model for the discretely sampled data set is effectively an AR(1) model, we simply use the lagged regression to obtain the simplest $\sqrt{T}$-consistent estimates. Each of the parameter values can be converted to the parameter values under the Vasicek model. Note that κ and σ are stated in terms of the unit of time equal to one year. The test results ($p = 2$ and 4) clearly indicate that the Vasicek model performs poorly. These resounding rejections suggest that the transformed residual does not have the right variance and the squared transformed residuals are autocorrelated. The fact that rejection consistently falls on $p = 2$ and 4 for the whole sample as well as for the five subsamples indicates a persistent pattern of misspecification.

Table 4.b present the results of testing the CIR model. Contrary to a common belief, this model performs no better than the Vasicek model. It has been resoundingly rejected for all $p = 2$ and 4 except for the most recent $1,000$ data points. In addition, the mean of the transformed residual also differs from the model prediction. This result is interesting given that the mean-reverting specification presumably allows for the mean of the interest rate series to be freely located and should thus exhibit no level bias. The conditional density function under the assumed model has, however, distorted the mean of the transformed residual. The result clearly points to the wrong skewed distribution (non-central chi-square) implied by the CIR model.

A Proof of Theorem 1

First, we argue that there exists a solution $\alpha^{(p)}_{k\times n}$. Because the column rank of $\mathbf{A}^{(p)-1/2}_{n\times n}\mathbf{B}^{(p)}_{n\times k_\theta}(\theta_0)$ is r, the null space has a dimension equal to $k = n - r$. Therefore, there will always be a solution $\alpha^{(p)}_{k\times n}$ to equations (12) and (13).

Next, we use Assumption 1 and the Taylor expansion in equation (7) to compute

$$
\begin{aligned}
&\sqrt{T}\alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\begin{bmatrix} Z^{(p)}_{1,T}(\hat{\theta}_T) \\ \vdots \\ Z^{(p)}_{n,T}(\hat{\theta}_T) \end{bmatrix} \\
= \; & \alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\begin{bmatrix} \sqrt{T}Z^{(p)}_{1,T}(\theta_0) \\ \vdots \\ \sqrt{T}Z^{(p)}_{n,T}(\theta_0) \end{bmatrix} + \alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\begin{bmatrix} \frac{\partial Z^{(p)}_{1,T}(\theta_0)}{\partial\theta'} \\ \vdots \\ \frac{\partial Z^{(p)}_{n,T}(\theta_0)}{\partial\theta'} \end{bmatrix}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1) \\
= \; & \alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\begin{bmatrix} \sqrt{T}Z^{(p)}_{1,T}(\theta_0) \\ \vdots \\ \sqrt{T}Z^{(p)}_{n,T}(\theta_0) \end{bmatrix} + \alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\mathbf{B}^{(p)}_{n\times k_\theta}(\theta_0)\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) + o_p(1) \\
= \; & \alpha^{(p)}_{k\times n}\mathbf{A}^{(p)-1/2}_{n\times n}\begin{bmatrix} \sqrt{T}Z^{(p)}_{1,T}(\theta_0) \\ \vdots \\ \sqrt{T}Z^{(p)}_{n,T}(\theta_0) \end{bmatrix} + o_p(1).
\end{aligned}
$$

Let n^* be the lowest common multiple of $\{1, 2\cdots, n\}$. For any $1 \leq m \leq n$,

$$
\begin{aligned}
\sqrt{T} Z_{m,T}^{(p)}(\theta_0) &= \frac{\sqrt{T/m}}{[T/m]} \sum_{i=1}^{[T/m]} Y_{m,i}^{(p)}(\theta_0) \\
&= \frac{1}{\sqrt{T/m}} \sum_{l=1}^{[T/n^*]} \sum_{j=1}^{n^*/m} Y_{m,(l-1)\frac{n^*}{m}+j}^{(p)}(\theta_0) + o_p(1) \\
&= \frac{1}{\sqrt{T/n^*}} \sum_{l=1}^{[T/n^*]} \left(\sqrt{\frac{m}{n^*}} \sum_{j=1}^{n^*/m} Y_{m,(l-1)\frac{n^*}{m}+j}^{(p)}(\theta_0) \right) + o_p(1) \\
&= \frac{1}{\sqrt{T/n^*}} \sum_{l=1}^{[T/n^*]} H_l(p,m) + o_p(1)
\end{aligned}
$$

where

$$
H_l(p,m) = \sqrt{\frac{m}{n^*}} \sum_{j=1}^{n^*/m} Y_{m,(l-1)\frac{n^*}{m}+j}^{(p)}(\theta_0).
$$

Thus,

$$
\begin{aligned}
\alpha_{k\times n}^{(p)} \mathbf{A}_{n\times n}^{(p)-1/2} \begin{bmatrix} \sqrt{T} Z_{1,T}^{(p)}(\theta_0) \\ \vdots \\ \sqrt{T} Z_{n,T}^{(p)}(\theta_0) \end{bmatrix} &= \frac{1}{\sqrt{T/n^*}} \sum_{l=1}^{[T/n^*]} \left(\alpha_{k\times n}^{(p)} \mathbf{A}_{n\times n}^{(p)-1/2} \begin{bmatrix} H_l(p,1) \\ \vdots \\ H_l(p,n) \end{bmatrix} \right) + o_p(1) \\
&= \frac{1}{\sqrt{T/n^*}} \sum_{l=1}^{[T/n^*]} Q_l(p) + o_p(1)
\end{aligned}
$$

where

$$
Q_l(p) = \alpha_{k\times n}^{(p)} \mathbf{A}_{n\times n}^{(p)-1/2} \begin{bmatrix} H_l(p,1) \\ \vdots \\ H_l(p,n) \end{bmatrix}.
$$

Note that $Q_l(p)$'s form an *i.i.d.* sequence of k-dimensional random vectors with mean $\mathbf{0}_{k\times 1}$ and variance covariance matrix $\mathbf{I}_{k\times k}$. This is true because (1) $Y_{m,j}^{(p)}(\theta_0)$ has a zero mean, and (2) $E\left(\sum_{k=1}^{n^*/i} Y_{i,k}^{(p)}(\theta_0) \sum_{l=1}^{n^*/j} Y_{j,l}^{(p)}(\theta_0)\right) = a_{ij}^{(p)}$ with $a_{ij}^{(p)}$ being defined in equation (8). The second fact can be established by decomposing the n^*-element block into $n^*/\kappa(i,j)$ independent blocks with $\kappa(i,j)$ elements each so that

$$
\begin{aligned}
E\left(\sum_{k=1}^{n^*/i} Y_{i,k}^{(p)}(\theta_0) \sum_{l=1}^{n^*/j} Y_{j,l}^{(p)}(\theta_0)\right) &= E\left(\sum_{k=1}^{n^*/i} W_{i,k}^{(p)} \sum_{l=1}^{n^*/j} W_{j,l}^{(p)}\right) \\
&= \frac{\sqrt{ij}}{\kappa(i,j)} E\left(\sum_{k=1}^{\kappa(i,j)/i} W_{i,k}^{(p)} \sum_{l=1}^{\kappa(i,j)/j} W_{j,l}^{(p)}\right).
\end{aligned}
$$

Applying the Central Limit Theorem yields

$$\alpha_{k\times n}^{(p)}\mathbf{A}_{n\times n}^{(p)-1/2}\begin{bmatrix}\sqrt{T}Z_{1,T}^{(p)}(\theta_0)\\ \vdots \\ \sqrt{T}Z_{n,T}^{(p)}(\theta_0)\end{bmatrix}\xrightarrow{D} N\left(0,\mathbf{I}_{k\times k}\right).$$

Thus,

$$J_T^{(p)}(\hat{\theta}_T)\xrightarrow{D}\chi^2\left(k\right).$$

The proof is thus complete.

B Expression for matrix $\mathbf{A}^{(p)}$

The following 10×10 matrices are computed using Monte Carlo simulation with one million repetitions.

$$\mathbf{A}_{10\times 10}^{(1)}=\begin{pmatrix}
0.0833 & 0.0813 & 0.0808 & 0.0806 & 0.0803 & 0.0801 & 0.0801 & 0.0799 & 0.0800 & 0.0800\\
0.0813 & 0.0834 & 0.0814 & 0.0813 & 0.0808 & 0.0808 & 0.0806 & 0.0804 & 0.0802 & 0.0803\\
0.0808 & 0.0814 & 0.0833 & 0.0812 & 0.0810 & 0.0814 & 0.0808 & 0.0806 & 0.0807 & 0.0806\\
0.0806 & 0.0813 & 0.0812 & 0.0834 & 0.0814 & 0.0815 & 0.0810 & 0.0812 & 0.0806 & 0.0808\\
0.0803 & 0.0808 & 0.0810 & 0.0814 & 0.0833 & 0.0814 & 0.0816 & 0.0810 & 0.0808 & 0.0814\\
0.0801 & 0.0808 & 0.0814 & 0.0815 & 0.0814 & 0.0834 & 0.0814 & 0.0813 & 0.0813 & 0.0812\\
0.0801 & 0.0806 & 0.0808 & 0.0810 & 0.0816 & 0.0814 & 0.0834 & 0.0816 & 0.0813 & 0.0812\\
0.0799 & 0.0804 & 0.0806 & 0.0812 & 0.0810 & 0.0813 & 0.0816 & 0.0834 & 0.0816 & 0.0815\\
0.0800 & 0.0802 & 0.0807 & 0.0806 & 0.0808 & 0.0813 & 0.0813 & 0.0816 & 0.0833 & 0.0813\\
0.0800 & 0.0803 & 0.0806 & 0.0808 & 0.0814 & 0.0812 & 0.0812 & 0.0815 & 0.0813 & 0.0834
\end{pmatrix}$$

$$\mathbf{A}_{10\times 10}^{(2)}=\begin{pmatrix}
0.0833 & 0.0766 & 0.0730 & 0.0707 & 0.0692 & 0.0683 & 0.0676 & 0.0670 & 0.0666 & 0.0664\\
0.0766 & 0.0833 & 0.0792 & 0.0785 & 0.0767 & 0.0762 & 0.0751 & 0.0749 & 0.0744 & 0.0740\\
0.0730 & 0.0792 & 0.0834 & 0.0799 & 0.0793 & 0.0794 & 0.0782 & 0.0778 & 0.0776 & 0.0768\\
0.0707 & 0.0785 & 0.0799 & 0.0832 & 0.0802 & 0.0802 & 0.0796 & 0.0800 & 0.0788 & 0.0786\\
0.0692 & 0.0767 & 0.0793 & 0.0802 & 0.0833 & 0.0803 & 0.0803 & 0.0800 & 0.0797 & 0.0801\\
0.0683 & 0.0762 & 0.0794 & 0.0802 & 0.0803 & 0.0834 & 0.0807 & 0.0805 & 0.0806 & 0.0799\\
0.0676 & 0.0751 & 0.0782 & 0.0796 & 0.0803 & 0.0807 & 0.0833 & 0.0807 & 0.0806 & 0.0806\\
0.0670 & 0.0749 & 0.0778 & 0.0800 & 0.0800 & 0.0805 & 0.0807 & 0.0834 & 0.0806 & 0.0809\\
0.0666 & 0.0744 & 0.0776 & 0.0788 & 0.0797 & 0.0806 & 0.0806 & 0.0806 & 0.0832 & 0.0808\\
0.0664 & 0.0740 & 0.0768 & 0.0786 & 0.0801 & 0.0799 & 0.0806 & 0.0809 & 0.0808 & 0.0833
\end{pmatrix}$$

$$\mathbf{A}^{(3)}_{10\times 10} = \begin{pmatrix} 0.0833 & 0.0434 & 0.0331 & 0.0278 & 0.0243 & 0.0218 & 0.0202 & 0.0188 & 0.0175 & 0.0166 \\ 0.0434 & 0.0833 & 0.0436 & 0.0434 & 0.0330 & 0.0330 & 0.0277 & 0.0277 & 0.0244 & 0.0244 \\ 0.0331 & 0.0436 & 0.0833 & 0.0437 & 0.0393 & 0.0434 & 0.0333 & 0.0313 & 0.0330 & 0.0280 \\ 0.0278 & 0.0434 & 0.0437 & 0.0833 & 0.0437 & 0.0433 & 0.0380 & 0.0435 & 0.0337 & 0.0331 \\ 0.0243 & 0.0330 & 0.0393 & 0.0437 & 0.0833 & 0.0438 & 0.0414 & 0.0389 & 0.0373 & 0.0435 \\ 0.0218 & 0.0330 & 0.0434 & 0.0433 & 0.0438 & 0.0834 & 0.0440 & 0.0437 & 0.0435 & 0.0395 \\ 0.0202 & 0.0277 & 0.0333 & 0.0380 & 0.0414 & 0.0440 & 0.0833 & 0.0437 & 0.0422 & 0.0407 \\ 0.0188 & 0.0277 & 0.0313 & 0.0435 & 0.0389 & 0.0437 & 0.0437 & 0.0832 & 0.0441 & 0.0437 \\ 0.0175 & 0.0244 & 0.0330 & 0.0337 & 0.0373 & 0.0435 & 0.0422 & 0.0441 & 0.0834 & 0.0439 \\ 0.0166 & 0.0244 & 0.0280 & 0.0331 & 0.0435 & 0.0395 & 0.0407 & 0.0437 & 0.0439 & 0.0834 \end{pmatrix}$$

$$\mathbf{A}^{(4)}_{10\times 10} = \begin{pmatrix} 0.0841 & 0.0485 & 0.0311 & 0.0229 & 0.0186 & 0.0160 & 0.0141 & 0.0127 & 0.0120 & 0.0109 \\ 0.0485 & 0.0833 & 0.0538 & 0.0461 & 0.0351 & 0.0316 & 0.0267 & 0.0249 & 0.0222 & 0.0210 \\ 0.0311 & 0.0538 & 0.0831 & 0.0536 & 0.0462 & 0.0455 & 0.0359 & 0.0326 & 0.0321 & 0.0277 \\ 0.0229 & 0.0461 & 0.0536 & 0.0832 & 0.0528 & 0.0499 & 0.0427 & 0.0449 & 0.0359 & 0.0340 \\ 0.0186 & 0.0351 & 0.0462 & 0.0528 & 0.0839 & 0.0521 & 0.0481 & 0.0443 & 0.0412 & 0.0445 \\ 0.0160 & 0.0316 & 0.0455 & 0.0499 & 0.0521 & 0.0835 & 0.0508 & 0.0495 & 0.0481 & 0.0431 \\ 0.0141 & 0.0267 & 0.0359 & 0.0427 & 0.0481 & 0.0508 & 0.0826 & 0.0500 & 0.0479 & 0.0452 \\ 0.0127 & 0.0249 & 0.0326 & 0.0449 & 0.0443 & 0.0495 & 0.0500 & 0.0827 & 0.0500 & 0.0486 \\ 0.0120 & 0.0222 & 0.0321 & 0.0359 & 0.0412 & 0.0481 & 0.0479 & 0.0500 & 0.0832 & 0.0490 \\ 0.0109 & 0.0210 & 0.0277 & 0.0340 & 0.0445 & 0.0431 & 0.0452 & 0.0486 & 0.0490 & 0.0823 \end{pmatrix}$$

References

[1] Ait-Sahalia, Y., 1996a, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* 9, 385-426.

[2] Ait-Sahalia, Y., 1996b, Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica* 64, 527-560.

[3] Ait-Sahalia, Y., 2002, Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach, *Econometrica* 70, 223-262.

[4] Bai, J., 2002, Testing Parametric Conditional Distributions of Dynamic Models, unpublished manuscript, Boston College.

[5] Bontemps, C. and N. Meddahi, 2002, Testing Normality: A GMM Approach, unpublished manuscript, CIRANO.

[6] Chan, K., A. Karolyi, F. Longstaff and A. Sanders, 1992, An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance* 47, 1209-1227.

[7] Corradi, V. and N. Swanson, 2002, A Bootstrap Specification Test for Diffusion Processes, unpublished manuscript, University of Exeter.

[8] Cox, J., J. Ingersoll and S. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.

[9] Diebold, F., T. Gunther and A. Tay, 1998, Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review* 39, 863-883.

[10] Diebold, F., J. Hahn and A. Tay, 1999, Multivariate Density Forcast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange, *Review of Economics and Statistics* 81, 661-673.

[11] Hong, Y. and H. Li, 2002, Nonparametric Specification Testing for Continuous-Time Models with Applications to Spot Interest Rates, unpublished manuscript, Cornell University.

[12] Khmaladze, E., 1981, Martingale Approach in the Theory of Goodness-of-Fit Tests, Theory of Probability and its Applications 26, 240-257.

[13] Thompson, S., 2002, Specification Tests for Continuous Time Models, unpublished manuscript, Harvard University.

[14] Vasicek, O., 1977, An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics* 5, 177-188.

Table 1.a The assumed model is constant mean and variance with normality. The data set is generated using AR(1) with γ as the AR coefficient. The rejection rates (500 simulations, 5% test, 2 degrees of freedom) for four test statistics (p = 1,2,3 and 4) and three sample sizes are reported in this table.

Rejection rate				
p	1	2	3	4
γ	Sample size = 200			
-0.5	0.022	0.168	**0.864**	0.042
-0.2	0.054	0.044	**0.250**	0.038
-0.1	0.046	0.052	**0.106**	0.056
0	**0.050**	**0.056**	**0.038**	**0.044**
0.1	0.072	0.052	**0.078**	0.048
0.2	0.038	0.066	**0.182**	0.048
0.5	0.044	0.132	**0.680**	0.072
γ	Sample size = 500			
-0.5	0.012	0.324	**0.978**	0.144
-0.2	0.044	0.050	**0.520**	0.048
-0.1	0.064	0.046	**0.168**	0.054
0	**0.054**	**0.062**	**0.050**	**0.062**
0.1	0.072	0.056	**0.138**	0.048
0.2	0.056	0.080	**0.378**	0.036
0.5	0.030	0.308	**0.906**	0.120
γ	Sample size = 1000			
-0.5	0.012	0.610	**0.984**	0.284
-0.2	0.050	0.054	**0.762**	0.048
-0.1	0.060	0.054	**0.290**	0.054
0	**0.058**	**0.064**	**0.040**	**0.038**
0.1	0.048	0.052	**0.236**	0.044
0.2	0.080	0.074	**0.676**	0.046
0.5	0.024	0.548	**0.964**	0.248

Table 1.b The assumed model is constant mean and variance with normality. The data set is generated using t-distribution with η as the degrees of freedom. The rejection rates (500 simulations, 5% test, 2 degrees of freedom) for four test statistics (p = 1,2,3 and 4) and three sample sizes are reported in this table.

Rejection rate				
p	1	2	3	4
η	Sample size = 200			
∞	**0.068**	**0.056**	**0.044**	**0.050**
20	0.060	**0.090**	0.052	0.032
10	0.066	**0.088**	0.048	0.066
7	0.094	**0.140**	0.062	0.064
5	0.070	**0.308**	0.050	0.150
4	0.076	**0.438**	0.094	0.228
η	Sample size = 500			
∞	**0.052**	**0.050**	**0.046**	**0.052**
20	0.054	**0.088**	0.060	0.060
10	0.062	**0.166**	0.064	0.096
7	0.088	**0.306**	0.076	0.146
5	0.070	**0.524**	0.150	0.256
4	0.070	**0.708**	0.192	0.450
η	Sample size = 1000			
∞	**0.062**	**0.044**	**0.038**	**0.042**
20	0.070	**0.110**	0.066	0.060
10	0.058	**0.212**	0.070	0.114
7	0.064	**0.490**	0.146	0.204
5	0.054	**0.750**	0.264	0.404
4	0.054	**0.874**	0.402	0.574

Table 1.c The assumed model is constant mean and variance with normality. The data set is generated using ARCH(1) with β_2 as the ARCH coefficient. The rejection rates (500 simulations, 5% test, 2 degrees of freedom) for four test statistics (p = 1,2,3 and 4) and three sample sizes are reported in this table.

Rejection rate				
p	1	2	3	4
β_2	Sample size = 200			
0	**0.068**	**0.070**	**0.038**	**0.060**
0.1	0.096	**0.070**	0.052	**0.034**
0.2	0.046	**0.118**	0.056	**0.062**
0.3	0.072	**0.134**	0.098	**0.104**
0.5	0.070	**0.162**	0.070	**0.144**
0.7	0.094	**0.190**	0.094	**0.224**
β_2	Sample size = 500			
0	**0.044**	**0.052**	**0.046**	**0.050**
0.1	0.058	**0.088**	0.062	**0.074**
0.2	0.066	**0.156**	0.076	**0.108**
0.3	0.082	**0.214**	0.112	**0.184**
0.5	0.064	**0.318**	0.138	**0.306**
0.7	0.106	**0.374**	0.182	**0.496**
β_2	Sample size = 1000			
0	**0.040**	**0.052**	**0.040**	**0.046**
0.1	0.064	**0.104**	0.070	**0.078**
0.2	0.060	**0.260**	0.130	**0.160**
0.3	0.062	**0.442**	0.190	**0.334**
0.5	0.068	**0.528**	0.252	**0.532**
0.7	0.064	**0.720**	0.346	**0.782**

Table 2 The assumed model is that of Vasicek (1977). The data set are generated using the Vasicek, CIR and CKLS specifications. The rejection rates (500 simulations, 5% test, 2 degrees of freedom) for four test statistics (p = 1,2,3 and 4) and four sample sizes are reported in this table.

Rejection rate				
p	1	2	3	4
Generating model	Sample size = 500			
Vacicek	**0.068**	**0.064**	**0.056**	**0.054**
CIR	0.068	0.068	0.052	**0.086**
CKLS	0.084	0.080	0.028	**0.768**
	Sample size = 1000			
Vacicek	**0.054**	**0.058**	**0.058**	**0.056**
CIR	0.058	0.056	0.050	**0.106**
CKLS	0.070	0.372	0.030	**0.414**
	Sample size = 2500			
Vacicek	**0.048**	**0.042**	**0.046**	**0.052**
CIR	0.044	0.080	0.050	**0.510**
CKLS	0.046	0.846	0.022	**0.938**
	Sample size = 5500			
Vacicek	**0.062**	**0.054**	**0.060**	**0.040**
CIR	0.072	0.088	0.054	**0.910**
CKLS	0.054	0.844	0.020	**1.000**

Table 3.a The assumed model is constant-mean GARCH(1,1) with conditional normality. The four test statistics (p = 1,2,3 and 4) along with the corresponding tail probabilities for chi-square of 2 degrees of freedom are given for the entire data sample (2520 S&P500 daily index returns) and two subsamples (938 and 1582). The bottom panel provides the parameter estimates with the standard errors.

Testing results				
p	1	2	3	4
Whole sample				
Test stat	0.8966	3.7121	1.7936	4.4100
Tail prob	0.6387	0.1563	0.4079	0.1102
First subsample (before December 31, 1996)				
Test stat	0.6285	4.3924	0.5948	7.6292
Tail prob	0.7303	0.1112	0.7427	**0.0220**
Second subsample (after December 31, 1996)				
Test stat	0.7196	8.3280	4.9996	1.2520
Tail prob	0.6978	**0.0155**	0.0821	0.5347
Estimation results				
	$\mu \times 10^4$	$\beta_0 \times 10^6$	β_1	β_2
Whole sample				
Estimate	6.74237	0.62628	0.91879	0.08044
Std err	1.64537	0.16021	0.00684	0.00679
First subsample (before December 31, 1996)				
Estimate	7.26237	1.20327	0.91697	0.05122
Std err	1.92170	0.42218	0.01940	0.01143
Second subsample (after December 31, 1996)				
Estimate	5.06697	9.47883	0.84799	0.10085
Std err	3.18199	2.14413	0.02140	0.01379

Table 3.b The assumed model is constant-mean GARCH(1,1) with conditional t distribution with η degrees of freedom. The four test statistics (p = 1,2,3 and 4) along with the corresponding tail probabilities for chi-square of 2 degrees of freedom are given for the entire data sample (2520 S&P500 daily index returns) and two subsamples (938 and 1582). The bottom panel provides the parameter estimates with the standard errors.

Testing results					
p	1	2	3	4	
Whole sample					
Test stat	1.4703	1.6775	1.0675	3.0521	
Tail prob	0.4794	0.4322	0.5864	0.2174	
First subsample (before December 31, 1996)					
Test stat	1.3317	1.9642	0.7506	4.9519	
Tail prob	0.5138	0.3745	0.6871	0.0841	
Second subsample (after December 31, 1996)					
Test stat	0.5886	9.3469	3.8627	3.2162	
Tail prob	0.7450	**0.0093**	0.1450	0.2003	
Estimation results					
	$\mu \times 10^4$	$\beta_0 \times 10^6$	β_1	β_2	η
Whole sample					
Estimate	7.25576	0.43187	0.93093	0.06907	7.20783
Std err	1.54516	0.19303	0.00908	0.00957	0.93936
First subsample (before December 31, 1996)					
Estimate	7.97966	1.37482	0.91306	0.05175	5.74627
Std err	1.76610	0.79606	0.03360	0.01865	1.22183
Second subsample (after December 31, 1996)					
Estimate	4.78140	6.78283	0.88077	0.08218	9.38635
Std err	3.00429	2.46545	0.02629	0.01769	1.76006

Table 4.a The assumed model is that of Vasicek (1977), i.e., $\delta = 0$. The four test statistics (p = 1,2,3 and 4) along with the corresponding tail probabilities for chi-square of 2 degrees of freedom are given for the entire data sample (5505 daily observations of 7-day Eurodollar deposit rates) and five subsamples dividing the last 5,000 data points (with 5 being the most recent 1,000). The bottom panel provides the parameter estimates.

Testing results				
p	1	2	3	4
5505				
Test stat	11.2164	3648.892	1.7415	203.5810
Tail prob	**0.0037**	**0.0000**	0.4186	**0.0000**
1000 (5)				
Test stat	1.6580	9.8752	2.3462	30.2330
Tail prob	0.4365	**0.0072**	0.3094	**0.0000**
1000 (4)				
Test stat	0.3615	88.0039	10.5642	67.2995
Tail prob	0.8346	**0.0000**	**0.0051**	**0.0000**
1000 (3)				
Test stat	1.0054	17.8354	6.1298	44.7609
Tail prob	0.6049	**0.0001**	**0.0467**	**0.0000**
1000 (2)				
Test stat	2.1791	186.2623	3.0478	7.2676
Tail prob	0.3364	**0.0000**	0.2179	**0.0264**
1000 (1)				
Test stat	0.8758	201.8716	3.3437	26.1881
Tail prob	0.6454	**0.0000**	0.1879	**0.0000**
Estimation results				
	κ	μ	σ	
5505	1.60884	0.08308	0.06460	
1000 (5)	1.38018	0.04047	0.01596	
1000 (4)	2.82658	0.08123	0.02310	
1000 (3)	1.15538	0.08053	0.02709	
1000 (2)	5.73857	0.13570	0.11729	
1000 (1)	3.52208	0.07110	0.05287	

Table 4.b The assumed model is that of CIR (1985), i.e., $\delta = 1/2$. The four test statistics (*p* = 1,2,3 and 4) along with the corresponding tail probabilities for chi-square of 2 degrees of freedom are given for the entire data sample (5505 daily observations of 7-day Eurodollar deposit rates) and five subsamples dividing the last 5,000 data points (with 5 being the most recent 1,000). The bottom panel provides the parameter estimates.

Testing results				
p	1	2	3	4
5505				
Test stat	12.3392	1708.326	4.3769	41.6322
Tail prob	**0.0021**	**0.0000**	0.1121	**0.0000**
1000 (5)				
Test stat	11.0817	2.9258	5.9018	3.1848
Tail prob	**0.0039**	0.2316	0.0523	0.2034
1000 (4)				
Test stat	40.8427	83.4566	2.7771	119.2926
Tail prob	**0.0000**	**0.0000**	0.2494	**0.0000**
1000 (3)				
Test stat	0.4294	32.0258	7.6149	10.0785
Tail prob	0.8068	**0.0000**	**0.0222**	**0.0065**
1000 (2)				
Test stat	14.1412	27.8218	3.3855	121.5066
Tail prob	**0.0008**	**0.0000**	0.1840	**0.0000**
1000 (1)				
Test stat	9.3204	144.7785	4.3413	180.1325
Tail prob	**0.0095**	**0.0000**	0.1141	**0.0000**
Estimation results				
	κ	μ	σ	
5505	1.61852	0.07543	0.19805	
1000 (5)	1.38018	0.04047	0.07831	
1000 (4)	2.65367	0.08125	0.07800	
1000 (3)	1.15599	0.08050	0.09032	
1000 (2)	6.06670	0.13640	0.32317	
1000 (1)	3.60625	0.07212	0.19717	