JOURNAL OF
Econometrics

Journal of Econometrics 103 (2001) 225–258
www.elsevier.com/locate/econbase

A consistent test for conditional symmetry in time series models

Jushan Bai *, Serena Ng

Department of Economics, Boston College, Chestnut Hill, MA 92467, USA

Abstract

The assumption of conditional symmetry is often invoked to validate adaptive estimation and consistent estimation of ARCH/GARCH models by quasi-maximum likelihood. Imposing conditional symmetry can increase the efficiency of bootstraps if the symmetry assumption is valid. This paper proposes a procedure for testing conditional symmetry. The proposed test does not require the data to be stationary or i.i.d., and the dimension of the conditional variables can be infinite. The proposed test is consistent and is asymptotically distribution free. In addition, the test is shown to have nontrivial power against root-T local alternatives. The size and power of the test are satisfactory even for small samples. Applying the test to various time series, we reject conditional symmetry in inflation, exchange rate and stock returns. Among the nonfinancial time series considered, we find that investment, the consumption of durables, and manufacturing employment also reject conditional symmetry. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: C12; C22

Keywords: Skewness; Conditional symmetry; Empirical distribution function; Brownian motion; ARCH/GARCH; Nonlinear time series

1. Introduction

The objective of this paper is to construct a consistent test for conditional symmetry using time series data. Given a sequence of stochastic variables $\{Y_t, X_t\}$, conditional symmetry is said to hold if the distribution of Y_t,

* Corresponding author.
E-mail addresses: jushan.bai@bc.edu (J. Bai), serena.ng@bc.edu (S. Ng).

0304-4076/01/$ - see front matter
PII: S0304-4076(01)00044-6

conditional on X_t, is symmetric with respect to its conditional mean. That is, $F_t(y+\mu_t|X_t)=1-F_t(-y+\mu_t|X_t)$, where F_t is the (conditional) cumulative distribution function (cdf) of Y_t conditional on X_t, and $\mu_t=E(Y_t|X_t)$ is the conditional mean. In this paper, we consider the following nonlinear time series regression model:

$$Y_t=h(X_t,\beta)+\sigma(X_t,\lambda)e_t,\quad t=1,\ldots,T \tag{1}$$

where $h(X_t,\beta)$ is the conditional mean, $\sigma^2(X_t,\lambda)$ is the conditional variance of Y_t, and e_t is a zero mean disturbance with unit variance and is independent of current and past X_t's. Under (1), conditional symmetry is equivalent to the symmetry of e_t about zero. That is, $f(e)=f(-e)$, or $1-F(e)-F(-e)=0$ for all e, where f and F are the density and the cdf of e_t, respectively.

The above framework encompasses linear and finite order autoregressive models with and without exogenous variables, as well as nonlinear models such as the self-exciting threshold autoregressive (SETAR) model. However, for many time series models, the conditioning information set may consist of an infinite number of variables. To accommodate this situation, we denote by $\Omega_t=\{Y_{t-1},Y_{t-2},\ldots;X_t,X_{t-1},\ldots\}$ the information set at time t, and test conditional symmetry using the following model:

$$Y_t=h(\Omega_t,\beta)+\sigma(\Omega_t,\lambda)e_t. \tag{2}$$

This framework is very general. For example, an MA(1) process $Y_t=e_t+\theta e_{t-1}$ can be written as

$$Y_t=\sum_{j=1}^{\infty}(-\theta)^jY_{t-j}+e_t. \tag{3}$$

This corresponds to $h(\Omega_t,\theta)=\sum_{j=1}^{\infty}(-\theta)^jY_{t-j}$ with $\Omega_t=\{Y_{t-1},Y_{t-2},\ldots\}$. A regression model with GARCH disturbances

$$Y_t=X_t'\beta+\sigma_te_t,$$

where $\sigma_t^2=\alpha+\delta\sigma_{t-1}^2+\gamma\sigma_{t-1}^2e_{t-1}^2$ can be written as (2) with $\Omega_t=\{Y_{t-1},Y_{t-2},\ldots,X_{t-1},X_{t-2},\ldots\}$ and

$$\sigma(\Omega_t,\lambda)=\left(\alpha\Big/(1-\delta)+\gamma\sum_{j=1}^{\infty}\delta^{j-1}(Y_{t-j}-X_{t-j}'\beta)^2\right)^{1/2}, \tag{4}$$

where $\lambda=(\alpha,\delta,\gamma)$. By allowing β to include elements of λ, Eq. (2) also allows for ARCH-M processes. The test statistics to be developed in this article can still be applied even though the dimension of the conditioning variables in all these cases is infinite.

The existence (or lack thereof) of conditional symmetry is important in a number of situations. A widely popular approach to modeling time varying conditional variances is the family of ARCH and GARCH models developed in Engle (1982) and Bollerslev (1986). Lee and Hansen (1994) and

Lumsdaine (1996) showed that when the conditional mean and the conditional variance in the Gaussian likelihood are correctly specified, the quasi-maximum likelihood estimator (QMLE) is consistent for the parameters of the GARCH(1,1) model even when the assumption of normality is false. However, motivated by the fact that the innovations in financial time series usually have fat tails and are sometimes asymmetric, many applications no longer specify a Gaussian likelihood.[1] In a recent paper, Newey and Steigerwald (1997) showed that when the likelihood is nonGaussian, consistent estimation of the GARCH parameters can be obtained by QMLE if both the true and the assumed innovation density are symmetric around zero and unimodal. However, if conditional symmetry does not hold, an additional parameter will be necessary to identify the location of the innovation distribution. Our proposed test can be used to determine if estimation of this additional parameter is necessary.

The assumption of conditional symmetry is commonly used in adaptive estimation. In a linear regression setting, Bickel (1982) showed that if conditional symmetry holds, adaptive estimation of the parameters can achieve the same information bound as the maximum likelihood estimator whether or not the error density is known. The results of Bickel have been extended to homoskedastic ARMA models by Kriess (1987), ARCH processes by Linton (1993), and error-correction models by Hodgson (1998). Newey (1988) showed that the parameters of a linear regression model can be estimated adaptively by generalized methods of moments, also under the maintained assumption of conditional symmetry.[2]

Knowledge about the properties of e_t also has efficiency implications for bootstrapping. The general bootstrap procedure for nonparametric and semi-parametric estimators is based on resampling from the (unrestricted) empirical distribution. As discussed in Brown and Newey (1998), a more efficient procedure is to bootstrap from the restricted (parametric) distribution. The intuition is simply that imposing a restriction (when it is true) increases statistical efficiency. One such restriction is the symmetry of $\hat{e}_t$, the estimated residuals.

Whether or not conditional symmetry holds is also an issue of macroeconomic interest. Symmetry of e_t implies that positive shocks to the conditional mean are as likely as negative shocks. If this is not the case, our forecasts should adjust to the possibility that positive and negative forecast errors are not equally likely. There are instances when economic behavior naturally gives rise to conditional asymmetry. A specific example is given by the 'No

[1] Bollerslev (1987), for example, used a t distribution to model exchange rates and stock returns, while Nelson (1991) used the exponential power distribution to model stock prices.

[2] It is noted that conditional symmetry is only needed for estimating the intercepts in the conditional mean and the conditional variance of regression models.

news is good news' model of Campbell and Hentschel (1992) in which the residuals in a model of log returns conditional on volatility are asymmetrically distributed. This is precisely our notion of conditional asymmetry. It should be noted, however, that except in special cases such as those to be discussed below, conditional symmetry does not, in general, imply unconditional symmetry.

The rest of this paper is organized as follows. In Section 2, we propose test statistics for conditional symmetry and analyze their asymptotic properties. The power of the tests is analyzed in Section 3. Results from simulations and empirical applications are provided in Section 4. Proofs are given in Appendix A.

2. The test statistics

2.1. The skewness coefficient

Skewness, or the third moment, is the statistic that naturally comes to mind when the object of interest is symmetry of a distribution. For this reason, the skewness coefficient is often used to describe evidence on asymmetry in the economics literature. Hsieh (1988), for example, performed diagnostics on the standardized estimated residuals using the coefficient of skewness. It is therefore important to understand the strengths and limitations of the skewness coefficient as a test for symmetry.

Consider the simple regression model $Y_t = X_t'\beta + e_t$, with e_t being i.i.d. $(0, \sigma^2)$ and independent of X_t. Conditional symmetry of Y_t is equivalent to the symmetry of e_t. Let $\hat{\beta}$ be the OLS estimator of β with residuals $\hat{e}_t = e_t - X_t(\hat{\beta} - \beta)$. We assume an intercept is included in the regression so that $\sum_t \hat{e}_t = 0$. The skewness coefficient is defined as $\tau = \mu_3/\sigma^3$, where $\mu_3 = Ee_t^3$. When applied to $\hat{e}_t$, the sample skewness coefficient is

$$\hat{\tau} = \hat{\mu}_3/\hat{\sigma}^3,$$

where $\hat{\mu}_3 = (1/T)\sum_{t=1}^{T} \hat{e}_t^3$, and $\hat{\sigma} = \sqrt{T^{-1}\sum_{t=1}^{T} \hat{e}_t^2}$. To derive the limiting distribution of $\sqrt{T}(\hat{\tau} - \tau)$ under the null hypothesis $\tau = 0$, notice that because $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, it is sufficient to obtain the limiting distribution of $T^{-1/2}\sum_{t=1}^{T} \hat{e}_t^3$. From $\hat{e}_t^3 = e_t^3 - 3e_t^2 X_t'(\hat{\beta} - \beta) + 3e_t[X_t'(\hat{\beta} - \beta)]^2 - [X_t'(\hat{\beta} - \beta)]^3$, it is easy to show that

$$T^{-1/2}\sum_{t=1}^{T} \hat{e}_t^3 = T^{-1/2}\sum_{t=1}^{T} e_t^3 - 3\left(\frac{1}{T}\sum_{t=1}^{T} e_t^2 X_t'\right)\sqrt{T}(\hat{\beta} - \beta) + o_p(1).$$

The first term on the right-hand side above converges to a normal random variable with zero mean (under the null) and variance $\mu_6 = Ee_t^6$. The second term depends on the distribution of $\sqrt{T}(\hat{\beta}-\beta)$. Different estimation methods (e.g., LAD or M-estimation) will yield different limiting distributions. With OLS, we have $(\hat{\beta}-\beta)=(X'X)^{-1}X'e$. Let $\alpha'=3\operatorname{plim}(T^{-1}\sum_{t=1}^{T}e_t^2X_t)(X'X/T)^{-1}$. Then

$$T^{-1/2}\sum_{t=1}^{T}\hat{e}_t^3=T^{-1/2}\sum_{t=1}^{T}e_t^3-\alpha' T^{-1/2}\sum_{t=1}^{T}X_te_t+o_\mathrm{p}(1).$$

Now

$$\begin{pmatrix} T^{-1/2}\sum_{t=1}^{T}e_t^3 \\ T^{-1/2}\sum_{t=1}^{T}X_te_t \end{pmatrix} \xrightarrow{\mathrm{d}} N(0,\Sigma), \qquad \Sigma=\begin{pmatrix} u_6 & \gamma' \\ \gamma & \Omega \end{pmatrix}$$

where $\gamma=\operatorname{plim}(T^{-1}\sum_{t=1}^{T}X_te_t^4)$ and $\Omega=\operatorname{plim}\sigma^2(X'X/T)$. We thus have $T^{-1/2}\sum_{t=1}^{T}\hat{e}_t^3\xrightarrow{\mathrm{d}}N(0,\delta'\Sigma\delta)$, where $\delta'=(1,-\alpha')$, and $\sqrt{T}\hat{\tau}\xrightarrow{\mathrm{d}}N(0,\delta'\Sigma\delta/\sigma^6)$. Let $\hat{\delta}$ and $\hat{\Sigma}$ be consistent estimators of δ and Σ, respectively. Define the test statistic $\hat{\pi}$ as the normalized sample skewness coefficient:

$$\hat{\pi}=\hat{\sigma}^3(\hat{\delta}'\hat{\Sigma}\hat{\delta})^{-1/2}\sqrt{T}\hat{\tau}.$$

The properties of $\hat{\pi}$ can be summarized as follows:

Theorem 1. Assume that $(X'X/T)$ converges to a positive definite matrix and that the errors e_t are i.i.d. with zero mean, variance σ^2, finite sixth moment, and are independent of the past and current regressors. With least squares estimation and under the hypothesis that $\tau=0$, we have

$$\hat{\pi}\xrightarrow{\mathrm{d}}N(0,1).$$

The test statistic based on the skewness coefficient is asymptotically normal. The advantage of the skewness coefficient is that it is intuitive and easy to construct. This will still be the case for estimators other than ordinary least squares, provided the limiting distribution of the estimator is known and has a simple form. Furthermore, when the data are not i.i.d., $\hat{\Sigma}$ can be replaced by an estimate of the long-run variance, and the construction of the test is still fairly simple. However, the skewness coefficient also has a number of limitations. First, the standard error of $\hat{\tau}$ depends on how the model is estimated, but not every estimator has a simple limiting distribution, particularly for more complicated models. [3] Second, the statistic requires the existence of the sixth moment, which is not satisfied by many useful distributions (such

[3] For example, deriving the skewness coefficient test for model (2) is not as simple as the linear model just considered. In addition to its dependence on the limiting distribution of $\hat{\beta}$ and $\hat{\lambda}$, the asymptotic distribution of $\hat{\tau}$ also depends on the functional forms of $h(\Omega_t,\beta)$ and $\sigma(\Omega_t,\lambda)$ as well as their derivatives.

as the t_5). GARCH processes will also fail this requirement. Third, the test is not consistent against alternatives which are asymmetric and yet have a skewness coefficient of zero. [4]

In Section 2.2, we propose a new test that can overcome these limitations. In particular, the new test only requires stochastic boundedness for $\sqrt{T}(\hat{\beta}-\beta)$ and $\sqrt{T}(\hat{\lambda}-\lambda)$, as opposed to the knowledge of their limiting distributions. The test is also consistent.

2.2. A new test

Recently, some consistent tests have been developed to test the null hypothesis of conditional symmetry. Most of these tests are based on the estimated residuals of a linear model and only require consistent estimates of the parameters β. Fan and Gencay (1995) proposed a test based on the idea that under symmetry, $2\int f(x)f(-x)\,\mathrm{d}x = \int f^2(x)\,\mathrm{d}x + \int f^2(-x)\,\mathrm{d}x$. Ahmad and Li's (1996) test is based on $\int_{-\infty}^{\infty}[f(x)-f(-x)]^2\,\mathrm{d}F=0$. The unknown density function $f(x)$ in both cases is estimated by the kernel smoothing method. Zheng (1998), on the other hand, constructed a test on the basis that under symmetry, $1-F(-x)-F(x)=0$. Building on the fact that if $\beta(\tau)$ is the quantile regression estimator for $\tau \in (0,1)$, then $\beta(\tau)+\beta(1-\tau)=2\beta(1/2)$ under symmetry, Newey and Powell (1988) suggested a test based on estimation of β by the method of 'asymmetric least squares'.

The test proposed in this paper differs from the aforementioned tests in several respects. First, previous tests are developed for i.i.d. data. Our test is more general and can be used even when X_t and/or Y_t are weakly dependent, and the data may even be nonstationary. Use of time series data raises issues that are otherwise irrelevant. For example, the conditioning variables in ARMA and GARCH models are infinite dimensional in theory, an issue that must be taken into account. Second, our test is based on the empirical distribution function (as is Zheng's test for i.i.d. data) and has nontrivial power against root-T local alternatives. In contrast, tests based on comparisons of nonparametrically estimated density functions may not have local power since local departures from symmetry will not be preserved by nonparametric smoothing.

Our test is based on empirical distribution functions and has nontrivial power against root-T local alternatives. Suppose e_t is i.i.d. with density $f(e)$ and cdf $F(e)$, and $\sigma_e=1$. Let $I(A)$ be an indicator which equals 1 when event A is true and 0 otherwise. Note that under symmetry, e_t and $-e_t$ have the same distribution. The idea of our test is to compare the empirical distribution

[4] The distribution $\mathrm{d}F = \frac{1}{48}[1-\operatorname{sgn} x \sin(|x|^{1/4})]\exp(-|x|^{1/4}), -\infty < x < \infty$ has all odd-order moments zero, but it is not symmetric. See Kendall and Ord (1994).

function of e_t $(t=1,\ldots,T)$ and that of $-e_t$ $(t=1,\ldots,T)$. Define the empirical process, $U_T^+(x)$, based on e_t, as

$$U_T^+(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(e_t \leqslant x) - F(x)].$$

It is well known that $U_T^+(x)$ converges to a Brownian bridge process, $\bar{B}(x)$, with $E[\bar{B}(x)] = 0$ and $E[\bar{B}(x)\bar{B}(y)] = F(x)(1 - F(y))$ for $x < y$. Likewise, an empirical process based upon $-e_t$, defined by

$$U_T^-(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(-e_t \leqslant x) - F(x)]$$

also converges to a Brownian bridge if e_t has a symmetric distribution. Although U_T^+ and U_T^- both depend on (the unobserved) F, their difference,

$$W_T(x) = U_T^+(x) - U_T^-(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(e_t \leqslant x) - I(-e_t \leqslant x)], \tag{5}$$

does not depend on F. For each point x, $W_T(x)$ is the difference between the number of e_t and the number of $-e_t$ less than or equal to x, then divided by the square root of T. Thus, $W_T(0)$ gives the scaled difference between the number of negative and positive values of e_t. Under symmetry, $W_T(x)$ should be small at all values of x. In view of the mathematical identity

$$W_T(x) = W_T(-x) \quad \text{(a.s.)}$$

one can consider either the positive or the negative values of x in the construction of W_T.

Lemma 1. Suppose $\{e_t, t=1,\ldots,T\}$ *is i.i.d. Let* $B(z)$ *be a standard Brownian motion on* $[0,1]$. *Then under the null hypothesis that* e_t *has a symmetric density function about zero, we have*

- *If* $x < 0$, $W_T(x) \Rightarrow B(2F(x))$, *and* $\max_{x\leqslant 0}|W_T(x)| \Rightarrow \max_{0\leqslant s\leqslant 1}|B(s)|$.
- *If* $x > 0$, $W_T(x) \Rightarrow B(2[1 - F(x)])$, *and* $\max_{x>0}|W_T(x)| \Rightarrow \max_{0\leqslant s\leqslant 1}|B(s)|$.

Note that although U_T^+ and U_T^- each converges to a Brownian bridge, their difference converges to a Brownian motion. Furthermore, because $2[1–F(\infty)]= 0$ and $2[1 - F(0)] = 1$ under symmetry, $B(2[1 - F(x)])$ $(x \geqslant 0)$ is a time-reversed Brownian motion on $[0,1]$.

If e_t was observed for every t, then $\max|W_T(x)|$ would have been the natural test for conditional symmetry. But $\{e_t\}$ is the sequence of innovations of a (possibly) nonlinear time series model, which we do not observe. Therefore, we consider feasible statistics based upon the estimated residuals, $\hat{e}_t$, and then use martingale transformation methods to obtain tests that are asymptotically distribution free. The transformation method was first studied by Khmaladze (1981) and has recently been extended in several directions by Bai (2000).

Let $\tilde{\Omega}_t = \{Y_{t-1},\ldots,Y_1,X_t,X_{t-1},\ldots,X_1,0,0,\ldots\}$ denote the feasible information set at time t. Then

$$\hat{e}_t = \frac{Y_t - h(\tilde{\Omega}_t,\hat{\beta})}{\sigma(\tilde{\Omega}_t,\hat{\lambda})}$$

and define $\hat{W}_T(x)$ by replacing e_t in $W_T(x)$ with $\hat{e}_t$. That is,

$$\hat{W}_T(x) = \hat{U}_T^+(x) - \hat{U}_T^-(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(\hat{e}_t \leqslant x) - I(-\hat{e}_t \leqslant x)].$$

The consequence of replacing e_t by the estimated residuals is that the process $\hat{W}_T(x)$ no longer converges to a Brownian motion. In fact, as shown in the appendix,

$$\hat{W}_T(x) = W_T(x) + 2f(x)\xi_{1T} + o_{\mathrm{p}}(1),$$

where $f(x)$ is the density of e_t and ξ_{1T}, given in (A.4) in Appendix A, is a stochastically bounded random variable (that does not depend on x). The presence of $f(x)$ in $\hat{W}_T(x)$ is a direct consequence of estimation of the conditional mean (see Appendix A). Estimation of the conditional variance does not, however, affect the test statistic. [5]

Since the limiting distribution of $\hat{W}_T(x)$ depends on f as well as the estimated parameters, the limiting distribution (and hence the critical values) will not be asymptotically distribution free. To circumvent this problem, we use martingale transformation methods to obtain an asymptotically distribution free test. Let $g=\dot{f}/f$, where f is the density of e_t and $\dot{f}$ is the derivative of f. Let f_T and g_T be estimates of f and g, respectively. For $x \leqslant 0$, define

$$S_T(x) = \hat{W}_T(x) - \hat{W}_T(0) + \int_x^0 h_T^-(y)\,\mathrm{d}y, \tag{6}$$

where

$$h_T^-(y) = g_T(y)f_T(y)\left[\int_{-\infty}^{y} g_T(z)^2 f_T(z)\,\mathrm{d}z\right]^{-1}\int_{-\infty}^{y} g_T(z)\,\mathrm{d}\hat{W}_T(z).$$

For $x > 0$, define

$$S_T(x) = \hat{W}_T(x) - \hat{W}_T(0) - \int_0^x h_T^+(y)\,\mathrm{d}y, \tag{7}$$

where

$$h_T^+(y) = g_T(y)f_T(y)\left[\int_{y}^{\infty} g_T(z)^2 f_T(z)\,\mathrm{d}z\right]^{-1}\int_{y}^{\infty} g_T(z)\,\mathrm{d}\hat{W}_T(z).$$

[5] Estimation of the variance introduces $xf(x)$ in Taylor series expansions. It does not, however, show up in $\hat{W}_T(x)$ because terms involving $xf(x)$ drop out when the difference of two empirical processes is considered. The underlying intuition is that if ε is symmetric about zero, then $c\varepsilon$ is also symmetric about zero for an arbitrary c, and hence estimation of the variance does not enter the empirical process in question.

The process S_T is the martingale transformation of $\hat{W}_T(x)$. Note that two separate transformations are performed: one for $x \leqslant 0$ and the other for $x > 0$. Define

$$CS^+ = \max_{x \geqslant 0} |S_T(x)|,$$

$$CS^- = \max_{x \leqslant 0} |S_T(x)|.$$

These are the test statistics for conditional symmetry. Appendix B discusses the computation of the test statistics.

The dependence of CS^+ and CS^- on T is understood and the notation is suppressed. We now state the assumptions under which Theorem 2 will be proved.

Assumption A1. The errors e_t are i.i.d. random variables with $E(e_t) = 0$, a cdf $F(x)$, and a continuous density function $f(x)$. The density function $f(x)$ has a finite Fisher information number, that is, $I_f = \int_{-\infty}^{\infty} (\dot{f}/f)^2 f \, \mathrm{d}x < \infty$. In addition, $\sup_x |xf(x)| < M$ for some $M > 0$.

Assumption A2. Let B_0 be a neighborhood of β_0. Then $\sup_{\beta \in B_0} (1/T) \sum_{t=1}^{T} ||\partial h_t(\Omega_t, \beta)/\partial \beta||^2 = O_\mathrm{p}(1)$, and $\max_{1 \leqslant t \leqslant T} T^{-1/2} ||\partial h_t(\Omega_t, \beta_0)/\partial \beta|| = o_\mathrm{p}(1)$.

Assumption A3. Let L_0 be a neighborhood of λ_0. Then $\sup_{\lambda \in L_0} (1/T) \sum_{t=1}^{T} ||\partial \sigma_t(\Omega_t, \lambda)/\partial \lambda||^2 = O_\mathrm{p}(1)$, and $\max_{1 \leqslant t \leqslant T} T^{-1/2} ||\partial \sigma_t(\Omega_t, \lambda_0)/\partial \lambda|| = o_\mathrm{p}(1)$. In addition, $\sigma(\Omega_t, \lambda_0) > c > 0$ for some c and for all t.

Assumption A4. The estimators satisfy $\sqrt{T}(\hat{\beta} - \beta_0) = O_\mathrm{p}(1)$, and $\sqrt{T}(\hat{\lambda} - \lambda_0) = O_\mathrm{p}(1)$.

Assumption A5. The effect of information truncation satisfies

$$\text{(i)} \quad T^{-1/2} \sum_{t=1}^{T} |h(\tilde{\Omega}_t, \beta_0) - h(\Omega_t, \beta_0)| = o_\mathrm{p}(1),$$

$$\text{(ii)} \quad T^{-1/2} \sum_{t=1}^{T} |\sigma(\tilde{\Omega}_t, \lambda_0) - \sigma(\Omega_t, \lambda_0)| = o_\mathrm{p}(1).$$

Assumption A6. The nonparametric estimators f_T for f and g_T for g satisfy

$$f_T(x) = f(x) + o_\mathrm{p}(1), \quad \text{and} \quad \int_{-\infty}^{\infty} (g_T - g)^2 \, \mathrm{d}F = o_\mathrm{p}(1),$$

where the first $o_\mathrm{p}(1)$ is uniform in x.

Remark 1. Assumption A1 is a technical condition for studying residual empirical processes. It is also used in nonparametric density estimations, e.g., Bickel (1982). Assumption A1 rules out certain distributions, e.g., uniform distribution as well as exponential distribution. However, we include the exponential distribution in our simulations to compare the results with those in the existing literature. Assumptions A2 and A3 are standard conditions for nonlinear estimations. Assumption A4 assumes $\sqrt{T}$ convergence of the estimated parameters and is satisfied by most estimators. Assumption A6 is a high level assumption about nonparametric estimators. If the densities are estimated nonparametrically (as we do), Assumption A6 is satisfied if the bandwidth is chosen appropriately. Parzen (1962) and Bickel (1982) offer specific conditions to ensure Assumption A6 for kernel estimations. A high level condition is used because there are many nonparametric methods, and any estimation method (e.g., kernel or series estimators) satisfying Assumption A6 is sufficient to ensure the validity of Theorem 2. In our simulations, we use the kernel method and choose the bandwidth according to Silverman (1986). Silverman's bandwidth choice satisfies the more general conditions in Parzen (1962) and Bickel (1982), and thus Assumption A6. Finally note that the first part of Assumption A6 implies $\int_{-\infty}^{\infty}(f_T(x)-f(x))^2\,\mathrm{d}x=o_{\mathrm{p}}(1)$ because $\int_{-\infty}^{\infty}(f_T(x)-f(x))^2\,\mathrm{d}x \leqslant \sup_x|f_T(x)-f(x)|\int|f_T(x)-f(x)|\,\mathrm{d}x \leqslant 2\sup_x|f_T(x)-f(x)|=o_{\mathrm{p}}(1)$.

Remark 2. Under Assumption A5, the effect of information truncation is small. It is satisfied trivially when there is no information truncation (e.g., cross-section regression models), but ARMA and GARCH models also satisfy Assumption A5. To see this, consider first an MA(1) model, (3), with $|\theta|<1$ and $h(\tilde{\Omega}_t,\theta)=\sum_{j=1}^{t-1}(-\theta)^j Y_{t-j}$. For some constant M,

$$
\begin{aligned}
E\sum_{t=1}^{T}|h(\tilde{\Omega}_t,\theta)-h(\Omega_t,\theta)| &\leqslant \sum_{t=1}^{T}\sum_{j=t}^{\infty}|\theta|^j E|Y_{t-j}| \\
&\leqslant M\sum_{t=1}^{\infty}\sum_{j=t}^{\infty}|\theta|^j = O(1).
\end{aligned}
$$

This implies that the right-hand side of Assumption A5(i) is $O_{\mathrm{p}}(T^{-1/2})$. Next consider a GARCH(1,1) process with $0<\delta<1$, see (4). From $|a-b|=|a^2-b^2|/(a+b)$, we have $|\sigma(\tilde{\Omega}_t,\lambda_0)-\sigma(\Omega_t,\lambda_0)| \leqslant |\sigma^2(\tilde{\Omega}_t,\lambda_0)-\sigma^2(\Omega_t,\lambda_0)|/c$ because the conditional variance is no smaller than c by Assumption A3. Using (4),

$$
E\sum_{t=1}^{T}|\sigma^2(\tilde{\Omega}_t,\lambda_0)-\sigma^2(\Omega_t,\lambda_0)| \leqslant \sum_{t=1}^{T}\sum_{j=t}^{\infty}\delta^j E(\sigma_t^2 e_t^2).
$$

If $E\sigma_t^2<M$, then the right-hand side above is uniformly bounded in T. This implies that the right-hand side of Assumption A5(ii) is $O_{\mathrm{p}}(T^{-1/2})$.

Remark 3. Assumptions A2–A5 are designed for nontrending regressors. Actually, our results also apply to trending regressors. Consider the simple cointegrating model $Y_t = X_t'\beta + e_t$, where e_t is i.i.d. and $X_t = X_{t-1} + \varepsilon_t$ with $X_0 = 0$. Then $T(\hat{\beta} - \beta) = O_p(1)$, violating A4. However, $\hat{e}_t = e_t - (1/\sqrt{T})(X_t/\sqrt{T})T(\hat{\beta} - \beta) = e_t + (1/\sqrt{T})O_p(1)$, where $O_p(1)$ is uniform in $t \leqslant T$. The fact that $\hat{e}_t = e_t + O_p(T^{-1/2})$, together with A1, is essentially all that is required for the proposed approach to work. There is an alternative way to see why our results extend to trending regressors. Consider a model with a linear trend as a regressor, $Y_t = a + bt + e_t$. For this model, the OLS estimator $\hat{b}$ converges at the rate of $T^{3/2}$ (violating A4 again, but clearly, $\hat{e}_t = e_t + O_p(T^{-1/2})$). However, we can also treat t/T as a regressor and rewrite the model as $Y_t = \alpha + \beta(t/T) + e_t$. By treating β as a fixed parameter independent of T, the estimated residuals are identical to the original model and yet all assumptions designed for standard regressors are met. Similarly, for the cointegrating case, one may think of $X_t/\sqrt{T}$ as the regressor, albeit somewhat unconventional.

Theorem 2. Under Assumptions A1–A6 *and the assumption of conditional symmetry, we have*

$$S_T(x) \Rightarrow B(1 - 2F(x)), \quad x \leqslant 0,$$

$$S_T(x) \Rightarrow B(2F(x) - 1), \quad x > 0,$$

$$CS^- \xrightarrow{\mathrm{d}} \max_{0 \leqslant s \leqslant 1} |B(s)|,$$

$$CS^+ \xrightarrow{\mathrm{d}} \max_{0 \leqslant s \leqslant 1} |B(s)|.$$

where $B(r)$ is a standard Brownian motion on $[0, 1]$.

The proof of the theorem assumes very general specifications for the conditional mean $h(\Omega_t, \beta)$ and conditional variance $\sigma(\Omega_t, \lambda)$, and includes conditional homoskedasticity as the special case. The asymptotic critical values at the 1%, 5%, and 10% levels of significance are 2.78, 2.21, and 1.91, respectively.

The theorem suggests that one can use either CS^- or CS^+ to test for conditional symmetry. This result arises because f is an even function and g is an odd function under the null hypothesis. Thus, if we had used f and g in the transformations, we would have $S_T(x) = S_T(-x)$ for all x, and thus $CS^+ = CS^-$ (exactly). Because f_T and g_T are consistent for f and g, transformations based on f_T and g_T are asymptotically equivalent to those based on f and g (see Lemma A.4 in Appendix A). This implies that $S_T(x) = S_T(-x) + o_p(1)$, where o_p is uniform over x. Therefore, $CS^- = CS^+ + o_p(1)$. Since CS^- and CS^+ have the same asymptotic distribution and are asymptotically equivalent, the test:

$$CS = \max\{CS^-, CS^+\} = \max_x |S_T(x)|$$

also has the same distribution as CS^+ and CS^-. We state this result as a corollary.

Corollary 1. Under the null hypothesis of conditional symmetry and the conditions of Theorem 2,

$$CS \xrightarrow{\mathrm{d}} \sup_{0 \leqslant s \leqslant 1} |B(s)|.$$

The CS statistic has three advantages and is our preferred statistic. First, it has better power since under the alternative, the equivalence of CS^- and CS^+ breaks down. Second, even if the null hypothesis was true, in finite samples, f_T may not be exactly even and g_T may not be exactly odd, and CS^- and CS^+ will not be the same. Third, use of CS makes it unnecessary to choose one test over the other.

It is instructive to examine graphically how the untransformed process $\hat{W}_T(x)$ and the transformed process $S_T(x)$ differ from $W_T(x)$. The former two processes are based on estimated residuals, while the latter is based on the true residuals. To this end, we use two samples of a normal variable, that is, $Y_i \sim N(0,1)$ to evaluate the three processes. Each sample consists of 100 observations of a standardized random variable. The residuals are defined as $\hat{e}_i = (Y_i - \bar{Y})/s_Y$, where $\bar{Y}$ is the sample mean and s_Y is the sample standard deviation. Fig. 1 plots the three processes evaluated at 200 points, x_k $(k = 1, 2, \ldots, 200)$ of which half are positive and half are negative, and thus these points are located symmetrically around zero. The dashed line and the dotted line represent, respectively, $\hat{W}_T$ and W_T. The solid line is the transformed process S_T, upon which the test statistics are based. Under the null hypothesis of symmetry, the theory says that $S_T(x)$ and $W_T(x)$ are both Brownian motion processes whereas $\hat{W}_T(x)$ is not. The departure of $\hat{W}_T(x)$ from $W_T(x)$ in all cases is apparent and indicates the effects of parameter estimation. However, $S_T(x)$ and $W_T(x)$ are quite close to each other, showing the effectiveness of the martingale transformations. In particular, the test statistic $CS = \max|S_T(x)|$ is close to $\max|W_T(x)|$. Furthermore, since $W_T(x)$ and $\hat{W}_T(x)$ are symmetric about zero, their graphs should be symmetric about the middle point x_k for $k = 100$. If the null hypothesis is true, the process S_T should also be roughly symmetric about the middle point. These features are all confirmed.

To examine the properties of the three empirical processes under the alternative of asymmetry, we consider two samples of $\chi^2_{(2)}$ observations, that is, $Y_i \sim ((\chi^2_{(2)} - 2)/2)$. The 95% confidence interval is given by $[-2.2, 2.2]$ and is also shown in the graphs. From Fig. 2, we see that both $W_T(x)$ and

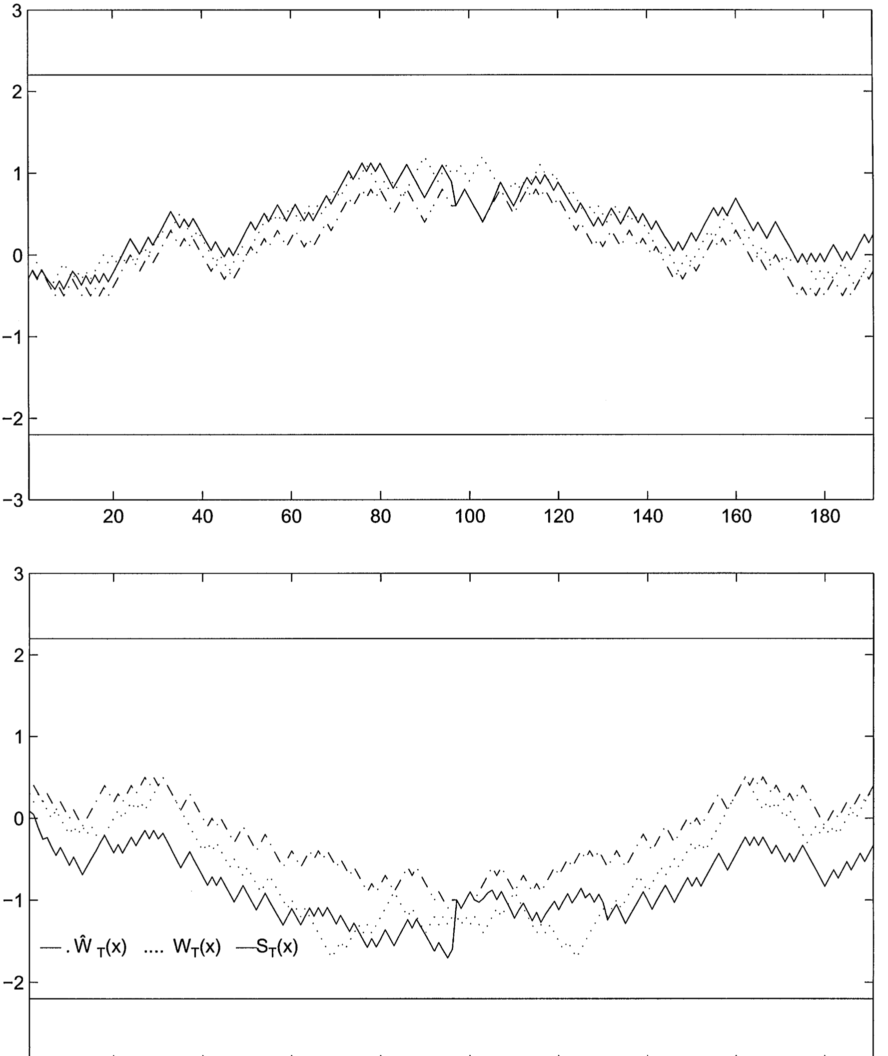

Fig. 1. Two sample paths of Gaussian observations.

S_T show strong evidence of asymmetry in the first sample. Recall that $W_T(x)$ is not observed in general. If one had used $\max|\hat{W}_T(x)|$ as the test statistic, one would have falsely accepted symmetry in the first sample because $\hat{W}_T(x)$ evidently lies within the standard error bands for all values of x. However, the CS lies outside the confidence band and correctly rejects symmetry. In

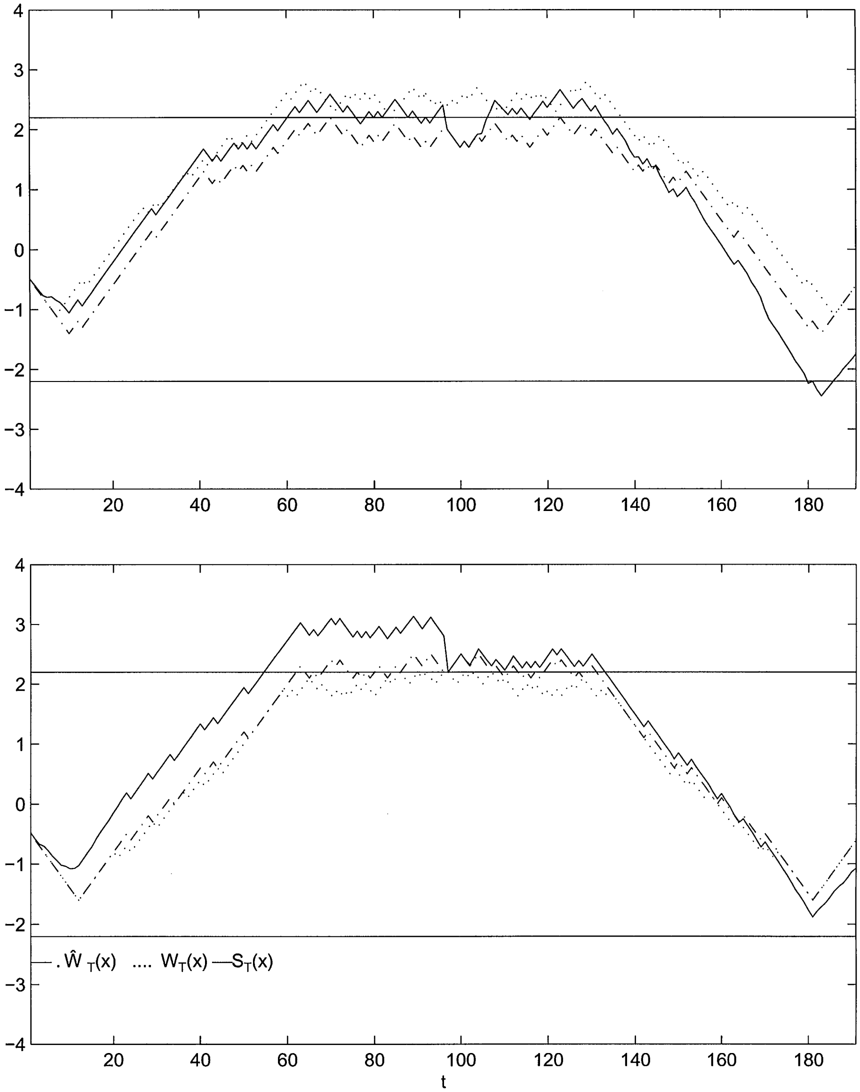

Fig. 2. Two sample paths of χ^2 observations.

the second sample, evidence of asymmetry is weaker, but the transformed process $S_T(x)$ still indicates asymmetry (in fact, stronger evidence of asymmetry than $W_T(x)$). This shows that the proposed test has power. Section 3 provides a formal analysis on power.

3. Power analysis

3.1. Local alternatives

As noted earlier, tests of symmetry based on estimated densities will not have root-T local power because root-T local departures from a symmetric density will be smoothed away by kernel smoothing, and the density estimator will converge to the underlying symmetric density (with a slower rate than $\sqrt{T}$). Although we use kernel smoothing to estimate f and g in the martingale transformations, our tests do not depend entirely on estimated densities and hence still have local power. To formally show that this is the case, we consider alternatives for which the disturbance e_t forms a triangular array. The distribution function of this array is described, for $t = 1, 2, \ldots, T$, by

$$e_{Tt} \sim \left(1 - \frac{\delta}{\sqrt{T}}\right) F(x) + \frac{\delta}{\sqrt{T}} H(x), \tag{8}$$

where F is the distribution function of a symmetric random variable, and H is that of an asymmetric random variable. We assume H satisfies Assumption A1 imposed on F. Define $v(x) = H(x) + H(-x) - 2H(0)$. It is easy to see that $v(x) \equiv 0$ if and only if $H(x)$ is the distribution function of a symmetric random variable. But $H(x)$ is asymmetric by assumption, hence $v(x) \not\equiv 0$. The following theorem summarizes the properties of the martingale transformed processes under the local alternative.

Theorem 3. Assume Assumptions A1–A6 *hold. Under the local alternative of* (8), *we have*

$$S_T(x) \Rightarrow B(1 - 2F(x)) + \delta[v(x) + \phi_v(x)], \qquad x < 0,$$

$$S_T(x) \Rightarrow B(2F(x) - 1) + \delta[v(x) + \phi_v(-x)], \quad x \geqslant 0,$$

where

$$\phi_v(x) = \int_x^0 \left[\dot{f}(y) \left(\int_{-\infty}^{y} g(z)^2 f(z)\,\mathrm{d}z \right)^{-1} \int_{-\infty}^{y} g(z)\,\mathrm{d}v(z) \right] \mathrm{d}y.$$

To show that the proposed tests have local power, we need to establish that $v(x) + \phi_v(x) \not\equiv 0$. Consider the integral equation (or functional relationship):

$$v(x) + \phi_v(x) \equiv 0. \tag{9}$$

Obviously, $v(x) \equiv 0$ is a solution to the above equation. The following lemma provides a general solution to Eq. (9) and is proved in Bai (2000).

Lemma 2. The only nonzero solution to the integral equation (9) *is* $v(x) = a[f(x) - f(0)]$, *where* $a \neq 0$ *is an arbitrary constant.*

A different constant a corresponds to a different solution, but this is the only class of nonzero solutions to the integral equation. Notice that by the definition of $v(x) = H(x) + H(-x) - 2H(0)$, in order for (9) to hold, we must also have

$$H(x) + H(-x) - 2H(0) = a[f(x) - f(0)]. \tag{10}$$

Therefore, to show that the tests have nontrivial power, we need only show that no asymmetric distribution will satisfy (10) and thus (9) under the maintained assumption that $E(e_{Tt}) = 0$. Under the local alternative (8), $F(x)$ is the distribution function of a symmetric random variable and thus has a zero mean. That is, $\int x \, \mathrm{d}F(x) = 0$. Under the assumption that $E(e_{Tt}) = 0$, the distribution $H(x)$ must also have a zero mean. Differentiate the identity (10), multiply by x, and then integrate both sides, we have $\int_{-\infty}^{\infty} xh(x)\,\mathrm{d}x - \int_{-\infty}^{\infty} xh(-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} a\dot{f}(x)\,\mathrm{d}x$, where $h(x)$ is the density of $H(x)$. Each of the two terms on the left-hand side is zero and the right-hand side is $-a$, which shows $a=0$. But then $H(x)+H(-x)-2H(0)=0$ by (10), which in turn implies $H(x)$ is a symmetric distribution, contradicting the assumption on H. Thus, under the zero-mean restriction on the errors, imposed by Assumption A1, no asymmetric distribution can satisfy (9). This leads to the following:

Corollary 2. Assume Assumptions A1–A6 *hold. Then the proposed tests have nontrivial local power against all departures from symmetry.*

3.2. Fixed alternatives

In this subsection, we show that the proposed tests are consistent against fixed alternatives and that the tests diverge at the rate of $\sqrt{T}$. Let $v(x) = F(x) + F(-x) - 2F(0)$. Under the alternative hypothesis that e_t has an asymmetric distribution, $v(x) \not\equiv 0$.

Theorem 4. Assume that Assumptions A1–A6 *hold and that* e_t *has an asymmetric distribution. Then*

$$|S_T(x) - \sqrt{T}[v(x) + \phi_v^-(x)]| = O_\mathrm{p}(1), \quad x < 0$$

and

$$|S_T(x) - \sqrt{T}[v(x) - \phi_v^+(x)]| = O_p(1), \quad x > 0,$$

where $O_p(1)$ is uniform over x and

$$\phi_v^-(x) = \int_x^0 \left[\dot{f}(y) \left(\int_{-\infty}^y g(z)^2 f(z)\,\mathrm{d}z \right)^{-1} \int_{-\infty}^y g(z)\,\mathrm{d}v(z) \right] \mathrm{d}y,$$

$$\phi_v^+(x) = \int_0^x \left[\dot{f}(y) \left(\int_y^{\infty} g(z)^2 f(z)\,\mathrm{d}z \right)^{-1} \int_y^{\infty} g(z)\,\mathrm{d}v(z) \right] \mathrm{d}y.$$

In addition, the test is consistent.

Again, by Lemma 2, the only class of nonzero solution to the integral equation $v(x) + \phi_v^-(x) \equiv 0$ is given by $v(x) = a[f(x) - f(0)]$ for some $a \neq 0$. This implies

$$F(x) + F(-x) - 2F(0) = a[f(x) - f(0)]. \tag{11}$$

However, no asymmetric distribution can satisfy the above restriction. This is because the left-hand side of (11) is an even function, implying that $f(x)$ is an even function. But if $f(x)$ is an even function, then e_t has a symmetric distribution and a contradiction is arrived. Thus, under a fixed alternative, $v(x)+\phi_v^-(x) \not\equiv 0$, and $CS^- = \sqrt{T}\max_{x<0} |v(x)+\phi_v^-(x)| + O_p(1) = O_p(\sqrt{T})$. This establishes the consistency of CS^- under a fixed alternative. Similarly, $v(x) - \phi_v^+(x) \not\equiv 0$. Thus, all the three tests are consistent. To our knowledge, the rate of divergence is faster than the existing consistent tests in the literature.

4. Simulations

We first present simulations to assess the size and power of the tests. In addition to some well-known distributions such as the normal and t, we also consider distributions from the generalized lambda family. This family encompasses a range of symmetric and asymmetric distributions that can be easily generated since they are defined in terms of the inverse of the cumulative distribution $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0 < u < 1$. The λ parameters are taken from Table 1 of Randles et al. (1980).

4.1. Testing for symmetry in the demeaned series

In this subsection, we only use a constant as the conditioning variable. The demeaned data are then standardized to have unit variance. The sizes of the tests are assessed by drawing random variables from seven symmetric distributions, which include normal, t_5, mixture normal, and four symmetric λ distributions. To assess power, we simulate data from eight asymmetric distributions, which include lognormal, $\chi^2_{(2)}$, exponential, and five asymmetric λ distributions. The distributions, along with the coefficient of skewness (τ) and

Table 1
Size of the tests: (H_0: symmetry around the mean) (asymptotic nominal size = 0.05)[a]

	$T=50$			$T=100$			$T=200$			τ	κ
	CS	CS^-	CS^+	CS	CS^-	CS^+	CS	CS^-	CS^+		
S1	0.037	0.011	0.031	0.051	0.023	0.039	0.049	0.026	0.040	0.0	3.0
S2	0.067	0.015	0.057	0.081	0.023	0.069	0.071	0.025	0.056	0.0	9.0
S3	0.042	0.013	0.035	0.042	0.024	0.031	0.045	0.024	0.041	0.0	2.5
S4	0.044	0.028	0.025	0.047	0.024	0.035	0.044	0.022	0.038	0.0	3.0
S5	0.078	0.021	0.065	0.087	0.028	0.069	0.075	0.022	0.066	0.0	6.0
S6	0.106	0.028	0.088	0.110	0.034	0.085	0.091	0.033	0.050	0.0	11.6
S7	0.134	0.036	0.110	0.140	0.029	0.124	0.117	0.045	0.092	0.0	126.0

[a]Notes: S1: $N(0,1)$; S2: t_5; S3: $e_1 I_{z\le 0.5} + e_2 I_{z>0.5}$, where $z \sim U(0,1)$ $e_1 \sim N(-1,1)$, and $e_2 \sim N(1,1)$; S4: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = 0.19754$, $\lambda_3 = 0.134915$, $\lambda_4 = 0.134915$; S5: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.08$, $\lambda_4 = -0.08$; S6: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -0.397912$, $\lambda_3 = -0.16$, $\lambda_4 = -0.16$; S7: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.24$, $\lambda_4 = -0.240$.

Table 2
Power of the tests (based on 5% asymptotic critical values)[a]

	$T=50$			$T=100$			$T=200$			τ	κ
	CS	CS^-	CS^+	CS	CS^-	CS^+	CS	CS^-	CS^+		
A1	0.977	0.942	0.832	1.000	1.000	0.996	1.000	1.000	1.000	6.18	113.9
A2	0.882	0.803	0.542	0.995	0.991	0.775	1.000	1.000	0.981	2.0	9.0
A3	0.878	0.795	0.541	0.997	0.993	0.759	1.000	1.000	0.982	2.0	9.0
A4	0.566	0.487	0.314	0.850	0.815	0.669	0.9982	0.958	0.998	0.5	2.2
A5	0.418	0.315	0.205	0.697	0.647	0.262	0.972	0.961	0.631	1.5	7.5
A6	0.307	0.177	0.180	0.416	0.350	0.146	0.647	0.626	0.140	2.0	21.2
A7	0.932	0.870	0.664	0.999	0.998	0.870	1.000	1.000	0.997	3.16	23.8
A8	0.961	0.915	0.729	1.000	1.000	0.929	1.000	1.000	0.999	3.8	40.7

[a]Notes: A1: lognormal: $\exp(e)$, $e \sim N(0,1)$; A2: χ^2_2; A3: exponential: $-\ln(e)$, $e \sim U(0,1)$; A4: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = 1.0$, $\lambda_3 = 1.4$, $\lambda_4 = 0.25$; A5: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0075$, $\lambda_4 = -0.03$; A6: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.1$, $\lambda_4 = -0.18$; A7: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.001$, $\lambda_4 = -0.13$; A8: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0001$, $\lambda_4 = -0.17$.

kurtosis (κ), are given in Tables 1 and 2. The choice of these distributions is motivated by a number of considerations. First, they are used in previous studies of testing symmetry and thus provide benchmarks for comparing size and power. Second, these distributions have a wide range of skewness and kurtosis and thus should cover the sample skewness and kurtosis of many economic series encountered in practice. For example, for the empirical data listed in Table 7, the sample kurtosis is in the range of 3–20, while the kurtosis in the simulated data varies from 3 to 130. Third, some of distributions are chosen to evaluate the robustness of test statistics when certain assumptions are not met. For example, the t_5 distribution does not have finite sixth moment, which is required by the $\hat{\pi}$ test.

The asymptotic critical values of the tests are 2.78, 2.20, and 1.91 at the 1%, 5%, and 10%, levels, respectively. To conserve space, we only report the results for the 5% test. The sizes of the tests in Table 1 indicate that both CS^- and CS^+ tend to be undersized, but the size of CS is generally more accurate. Of the seven symmetric distributions considered, the CS is slightly oversized under S5, with larger size distortions under S6 and S7. Notice that S6 and S7 have large kurtosis. The results for power are reported in Table 2. The power being reported is not adjusted for size because the null hypothesis is satisfied by a class of distributions (composite hypothesis), each of which has different size properties. In spite of this caveat, the CS statistic has substantially more power than CS^- or CS^+. This is to be expected since the CS rejects conditional symmetry if either CS^- or CS^+ rejects, or both. This gain in power is nontrivial. For example, in the χ_2^2 case, the rejection rate for the CS is over 0.9, while the CS^+ rejects only 54% of the time in small samples.

All tests considered have low power for cases A4–A6. Zheng's test also has low power when tested against these alternatives. Note that these distributions also have large kurtosis. Indeed, the results reported in Randles et al. (1980) for an unconditional test of symmetry also show the same phenomenon. Thick tailed distributions seem to pose both size and power problems for testing skewness, problems that are not unique to our proposed tests. For other asymmetric distributions, the CS generally has good power even when the sample size is small. For example, the distributions A7 and A8 were also considered in Zheng (1998). The CS test has a marked improvement in power over Zheng's test for these two distributions. Compared with the results of Fan and Gencay (1995), who also examined distributions A1–A3, our CS test has comparable power, rejecting the null hypothesis over 90% of the time even when the sample size is small. This comparison of power may not be fair due to the lack of size adjustment. Nevertheless, the CS stacks up well with tests in the literature that are applicable to i.i.d. data only.

4.2. Testing for conditional symmetry in time series regressions

To consider the size and power of the tests in a more general setting, we consider the following regression models:

1. $y_t = a + \sum_{i=1}^{k} x_{it} + e_t,\ x_{it} \sim$ i.i.d., $i = 1, \ldots, k$;
2. AR(1): $y_t = \rho y_{t-1} + e_t,\ \rho = 0.5, 0.8$;
3. MA(1): $y_t = e_t + \rho e_{t-1},\ \rho = 0.5, 0.8$;

Table 3
Size and power of the CS and $\hat{\pi}$ tests: regression model with i.i.d. regressors[a]

DGP 1: $y_t = 1 + \sum_{i=1}^{k} X_{ti} + e_t$
Regression: $y_t = \alpha + \sum_{i=1}^{k} X_{ti}\beta_i + e_t$

			Model					
Test	k	T	1	2	3	4	5	6
CS	1	50	0.053	0.074	0.042	0.937	0.880	0.980
$\hat{\pi}$			0.035	0.044	0.026	0.510	0.427	0.322
CS	1	100	0.046	0.067	0.038	0.999	0.998	1.000
$\hat{\pi}$			0.043	0.030	0.026	0.753	0.571	0.437
CS	1	200	0.042	0.061	0.044	1.000	1.000	1.000
$\hat{\pi}$			0.039	0.040	0.039	0.898	0.661	0.651
CS	4	50	0.052	0.051	0.030	0.813	0.716	0.783
$\hat{\pi}$			0.047	0.069	0.027	0.592	0.564	0.584
CS	4	100	0.046	0.062	0.042	0.999	0.984	0.994
$\hat{\pi}$			0.045	0.049	0.029	0.811	0.663	0.817
CS	4	200	0.042	0.081	0.037	1.000	1.000	1.000
$\hat{\pi}$			0.047	0.053	0.047	0.911	0.772	0.925

[a]Notes: Model 1: $e_t \sim N(0,1)$; Model 2: $e_t \sim t_5$; Model 3: $e_t \sim N(-1,1)I_{z<0.5} + N(1,1)I_{z>=0.5}$, $z \sim U(0,1)$; Model 4: $e_t \sim \chi_2^2$; Model 5: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.001$, $\lambda_4 = -0.13$; Model 6: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0<u<1$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0001$, $\lambda_4 = -0.17$.

4. GARCH(1,1): $y_t = 1 + u_t$, $u_t = \sigma_t e_t$, $\sigma_t^2 = \phi_0 + \phi_1\sigma_{t-1}^2 + \phi_2 u_{t-1}^2$; $\phi_0 = 2$, $\phi_1 = 0.5$, $\phi_2 = 0.3$; $\phi_0 = 2$, $\phi_1 = 0.9$, $\phi_2 = 0.05$.

The errors e_t are drawn from three symmetric distributions (normal, t-distribution, and mixture normal) to assess size, and from three asymmetric distributions (chi-square, and two lambda distributions) to assess power (see note to Table 3). After e_t is drawn, the population mean and standard deviation of e_t are used to standardize the series. The models are estimated, and the estimated residuals are tested for conditional symmetry.

Since the CS test dominates CS^+ and CS^- on both theoretical and empirical grounds, we only consider the CS test hereafter. Table 3 reports results based on the least squares residuals of a linear model with a constant and k i.i.d. variables as regressors. Table 4 reports results using residuals from least squares estimation of an AR(1) model. Since the sampling properties of the conditional skewness coefficient is known when the conditional mean is estimated by least squares, we also report the size and power of $\hat{\pi}$.

Compared to the results in Table 2 which did not include the random regressors, the power of CS is generally lower in small samples. As in Fan and Gencay (1995), power also decreases as the number of regressors increases in small samples. But this is a small sample phenomenon. For

Table 4
Size and power of the CS and $\hat{\pi}$ tests: AR(1)

DGP 2: $y_t = \rho y_{t-1} + e_t$
Regression: $y_t = \alpha + \beta y_{t-1} + e_t$

			Model					
Test	ρ	T	1	2	3	4	5	6
CS	0.5	50	0.044	0.061	0.032	0.938	0.862	0.908
$\hat{\pi}$			0.033	0.043	0.043	0.489	0.409	0.383
CS	0.5	100	0.046	0.087	0.046	1.000	0.998	1.000
$\hat{\pi}$			0.044	0.033	0.026	0.731	0.560	0.554
CS	0.5	200	0.047	0.071	0.037	1.000	1.000	1.000
$\hat{\pi}$			0.037	0.039	0.046	0.891	0.750	0.660
CS	0.8	50	0.047	0.073	0.035	0.941	0.850	0.899
$\hat{\pi}$			0.024	0.035	0.023	0.489	0.409	0.382
CS	0.8	100	0.047	0.075	0.047	1.000	0.996	0.996
$\hat{\pi}$			0.036	0.042	0.056	0.745	0.582	0.545
CS	0.8	200	0.053	0.066	0.042	1.000	1.000	1.000
$\hat{\pi}$			0.052	0.029	0.047	0.915	0.742	0.679

sample sizes that we are likely to encounter in economic analysis, increasing the number of regressors should not have implications for power. Evidently, the CS has good size and power even the autoregressive coefficient is large.

Comparing the CS with $\hat{\pi}$, the size of $\hat{\pi}$ is comparable to that of CS in static regressions (see Table 3). However, in dynamic regressions (Table 4), $\hat{\pi}$ tends to be slightly undersized. Most importantly, the power of CS is always higher than $\hat{\pi}$. The conditional skewness coefficient has power at $T=200$ observations comparable to the CS with $T=50$. There is a nontrivial trade-off between power and computation ease. Interestingly, the $\hat{\pi}$ test has correct size for the t_5 distribution, even though sixth moment of t_5 does not exist.

Results for MA(1) and GARCH(1,1) are reported in Tables 4– 6. Note that in the GARCH(1,1) case, it is $\hat{e}_t = \hat{u}_t / \hat{\sigma}_t$ that is being tested. The CS generally has good size and power, correctly rejecting symmetry with probability close to one when $T \geqslant 100$. Even when T is small, the power is usually well over 70%. The results are robust even when the error process is close to being an IGARCH.

4.3. Empirical applications

The tests are applied to 21 macroeconomic time series. Data for GDP, the GDP deflator, the consumption of durables, final sales, the consumption of nondurables, residential investment, and nonresidential investment are taken

Table 5
Size and power of the *CS* test: MA(1) regressor

DGP 3: $y_t = e_t + \rho e_{t-1}$
Regression: $y_t = \alpha + e_t + \rho e_{t-1}$

		Model					
ρ	T	1	2	3	4	5	6
0.5	50	0.045	0.060	0.047	0.833	0.780	0.816
0.5	100	0.050	0.084	0.043	0.991	0.987	0.995
0.5	200	0.041	0.065	0.042	1.000	1.000	1.000
0.8	50	0.051	0.062	0.037	0.793	0.740	0.758
0.8	100	0.039	0.090	0.049	0.995	0.980	0.991
0.8	200	0.045	0.063	0.036	1.000	1.000	1.000

Table 6
Size and power of the test: GARCH(1,1) regressor

DGP 4: $y_t = 1 + u_t$, $u_t = e_t\sigma_t$, $\sigma_t^2 = \phi_0 + \phi_1\sigma_{t-1}^2 + \phi_2 u_{t-1}^2$
Regression: GARCH(1,1) with Gaussian likelihood

		Model[a]					
ϕ	T	1	2	3	4	5	6
(20.0, 0.5, 0.3)	50	0.035	0.072	0.036	0.937	0.884	0.905
(20.0, 0.5, 0.3)	100	0.047	0.080	0.044	0.995	0.989	0.994
(20.0, 0.5, 0.3)	200	0.046	0.069	0.048	1.000	1.000	1.000
(20.0, 0.9, 0.05)	50	0.053	0.093	0.040	0.758	0.810	0.873
(20.0, 0.9, 0.05)	100	0.052	0.060	0.031	0.989	0.989	0.992
(20.0, 0.9, 0.05)	200	0.047	0.097	0.041	1.000	1.000	1.000

[a]Models 1 to 6 are the same as in Table 3.

from the national accounts and are quarterly data. The unemployment rate, employment, M2, CPI are monthly series. The 30 day interest rate, and M2 are weekly data.[6] With the exception of the interest rate and the unemployment rate (which we do not take logs), we take first difference of the logarithm of the data. We then estimate an AR(2) model for each series by least squares. The residuals are then used to test conditional symmetry.[7] We also considered three exchange rates (in logged first differences), and the

[6] All data (except for the stock returns which are obtained from CRSP) are taken from the Economic Time Series Page, and URL is vos.business.uab.edu/data.htm.

[7] GARCH errors are also considered but we only report results with nonGARCH errors in order to compare with the skewness coefficient ($\hat{\pi}$) test, which would require a new theory under GARCH errors. Moreover, the test statistics are found similar with or without GARCH errors.

Table 7
Application to macroeconomic data[a]

Sample	Series	CS	CS^-	CS^+	$\hat{\pi}$	$\hat{\tau}$	$\hat{\kappa}$
71:1-97:12	Canada–US ex. rate	1.896	1.896	1.186	1.554	0.226	3.139
71:1-97:12	German–US ex. rate	0.814	0.643	0.814	−0.644	−0.134	3.499
71:1-97-12	Japan–US ex. rate	3.968	3.033	3.968	−2.369	−0.481	3.905
48:1-97:12	Unemployment rate	0.750	0.750	0.719	0.074	0.015	9.476
46:1-97:12	Ind. prod.	1.233	1.178	1.233	1.659	0.963	12.322
59:1-97:4	Inflation (GDP)	1.021	1.021	0.739	0.598	0.175	4.475
59:1-97:4	GDP	0.756	0.586	0.756	−0.633	−0.275	4.830
47:1-97:12	Inflation (CPI)	2.283	2.283	1.607	1.181	0.799	8.059
92:01:03-96:05:10	30 day int. rate	0.949	0.949	0.882	0.978	0.930	10.376
92:01:06-96:05:13	M2	0.863	0.832	0.863	−0.167	−0.025	3.013
59:3-96:4	Con. durables	2.641	2.115	2.641	−1.958	−0.854	5.116
59:3-96:4	Con. non-durables	1.117	0.806	1.117	0.785	0.260	4.454
46:1-96:11	Employment	1.504	1.022	1.504	−1.615	−0.299	3.778
69:3-97:4	Investment	0.568	0.392	0.568	−1.440	−0.724	5.716
46:1-97:12	Manu. employment	2.256	1.013	2.256	−0.326	−0.239	18.125
46:1-97:12	Non-Manu. employment	1.330	0.891	1.330	1.007	0.302	8.281
59:3-97:4	Final sales	0.988	0.954	0.988	0.416	0.257	5.893
59:3-97:4	Non-resid. invest	2.190	1.479	2.190	−0.891	−0.273	3.825
59:3-97:4	Resid. invest	0.906	0.906	0.687	−0.555	−0.211	5.203
90:01:02-96:12:31	Stock returns(V)	1.795	1.795	1.394	−2.426	−0.421	5.202
90:01:02-96:12:31	Stock returns(E)	5.577	3.998	5.577	−2.466	−0.594	7.864

[a]Note: The critical values for the CS tests are 2.78, 2.20 and 1.91 at the 1, 5, and 10% levels respectively. The two tailed critical values for $\hat{\pi}$ are $\pm$2.32, 1.96, and 1.64. (V) denotes value weighted, and (E) denotes equal weighted, both exclude dividends.

value as well as the equal weighted CRSP daily stock returns. For these financial series, we simply remove the mean. The sample skewness and kurtosis coefficients for the 21 series of estimated residuals are also reported.

The results in Table 7 indicate evidence of conditional asymmetry at the 1% level in the Japan–US exchange rate and equal weighted stock returns, at the 5% level for the CPI inflation, consumption of durables and manufacturing employment, and at the 10% level for nonresidential investment. The statistics for the US/Canada exchange rate and value weighted stock returns are also close to being significant at the 10% level. The evidence for the Japan–US exchange rate, stock returns, and durables is particularly convincing because the CS^+ and CS^- reject the null hypothesis. The findings for financial series are consistent with those of French et al. (1987). Hsieh (1988) finds that the residuals from an exchange rate model with conditional heteroskedasticity are skewed. Our test, which takes into account that estimated residuals are used in the testing, corroborates his evidence. The finding that investment, durables, and manufacturing employment reject conditional symmetry is also interesting because the dynamics of these series are often believed to be affected by fixed costs of adjustments.

We also apply the $\hat{\pi}$ statistic to each of the 21 series. Consistent with the CS test, $\hat{\pi}$ also rejects conditional symmetry in the Japan–US exchange rate, consumption durables, and equal weighted stock returns. However, $\hat{\pi}$ cannot reject conditional symmetry in CPI inflation and manufacturing employment. The only case when $\hat{\pi}$ rejects and the CS does not is the value weighted stock returns.

It is useful to put into perspective these results for conditional symmetry vis-á-vis the evidence for cyclical asymmetry in the macroeconomic literature. If X_t is an ARMA process, it can equivalently be represented as $X_t = \sum_{i=0}^{\infty} a_i e_{t-i}$. It follows that if e_t is symmetric, X_t will also be symmetric. Our evidence of conditional symmetry in US output growth and industrial production is consistent with the evidence of Delong and Summers (1982) for unconditional symmetry in the two series. We note that, however, if e_t is asymmetric, it does not necessarily imply X_t is asymmetric.

5. Conclusion

In this paper, we propose a consistent test for conditional symmetry in dynamic models. Unlike other tests that exist in the literature, the CS test is valid whether or not the data are i.i.d. and is suited for time series applications. The proposed test is asymptotically distribution free and, in general, has good finite size and power properties.

We also consider the use of the skewness coefficient in testing for conditional symmetry. Although the skewness coefficient is less powerful than the CS test, it is still quite useful for two reasons. The first is computation ease. Second, it is conceivable that the CS test may miss asymmetry in small samples, whereas skewness coefficient can identify it. This is because consistent tests have to take into account departures from symmetry over all directions, whereas the skewness coefficient delivers its best power at one particular direction, the third moment.

In terms of computation, the proposed test has a high one-time programming cost, but in some sense it actually demands less work from applied researchers than does the skewness coefficient test. This is because the only required input for the test is the estimated residuals. Given the residuals, the construction of the test does not depend on the specification of conditional mean or conditional variance. Thus, the same computer code can be used for all specifications.[8] In contrast, to construct the $\hat{\pi}$ test, the asymptotic variance of the estimated skewness coefficient ($\hat{\tau}$) must be derived case by case because it varies with model specifications and estimation methods. Once these issues are taken into account, the proposed test is more versatile and should serve as a useful complement to the skewness coefficient or other tests for symmetry in the literature.

Acknowledgements

We thank seminar participants at the Camp Econometrics and three anonymous referees for many helpful comments. This paper was also presented at the NBER/NSF Time Series Seminar held in Chicago in September, 1998. The first author acknowledges financial support from the National Science Foundation under Grant SBR-9709508.

Appendix A. Proofs

Proof of Lemma 1. First we derive the variance and covariance function for the process $W_T(x)$. For $x, y \leqslant 0$, it is straightforward to show that $EW_T(x)W_T(y) = E\{[I(e_t \leqslant x) - I(-e_t \leqslant x)][I(e_t \leqslant y) - I(-e_t \leqslant y)]\} = 2F(x \wedge y)$, where $x \wedge y = \min\{x, y\}$. The finite dimensional convergence of $W_T(x)$ to normal random variables and tightness follow from standard empirical process theorems. Thus, $W_T(x)$ converges weakly to a Gaussian process. Because a time-scaled Brownian motion $B(2F(x))$ has the same variance and covariance function as $W_T(x)$, it follows that $W_T(x) \Rightarrow B(2F(x))$. Similarly, for $x \geqslant 0$, $y \geqslant 0$, $EW_T(x)W_T(y) = E\{[I(e_t \leqslant x) - I(-e_t \leqslant x)][I(e_t \leqslant y) - I(-e_t \leqslant y)]\} = 2[1 - F(x \vee y)]$, where $x \vee y = \max\{x, y\}$. A time-scaled

[8] The code is available from the authors upon request.

Brownian motion $B(2[1-F(x)])$ has the same variance and covariance function as $W_T(x)$, we have $W_T(x) \Rightarrow B(2[1-F(x)])$ (for $x \geqslant 0$). $\square$

To prove Theorem 2, we need a number of lemmas.

Lemma A.1. Let $B(r)$ be a standard Brownian motion on $[0,1]$ and let g be a function on $[0,1]$ such that $\int_s^1 g^2(v)\,\mathrm{d}v > 0$ for every $s \in [0,1)$. Then

$$J(r) = B(r) - \int_0^r \left[g(s) \left(\int_s^1 g(v)^2 \,\mathrm{d}v \right)^{-1} \int_s^1 g(v)\,\mathrm{d}B(v) \right] \mathrm{d}s$$

is also a standard Brownian motion on $[0,1]$.

Proof. $J(r)$ is Gaussian because it is a linear transformation of $B(r)$. Elementary calculation (although tedious) shows that $EJ(r)J(s) = r \wedge s$. $\square$

Lemma A.2. Let $B(r)$ be a standard Brownian motion on $[0,1]$ and let g be a function on $[0,1]$ such that $\int_0^s g(v)^2\,\mathrm{d}v > 0$ for every $s \in (0,1]$. Then

$$J(r) = B(r) - B(1) + \int_r^1 \left[g(s) \left(\int_0^s g(v)^2 \,\mathrm{d}v \right)^{-1} \int_0^s g(v)\,\mathrm{d}B(v) \right] \mathrm{d}s$$

is a time-reversed Brownian motion on $[0,1]$. That is, $EJ(r)J(s) = 1-(r \vee s)$.

Proof. Again this follows from a direct calculation showing that $EJ(s)J(r) = 1-(s \vee r)$. $\square$

Lemma A.3. Let $B(r)$ be a standard Brownian motion on $[0,1]$ and let $H(x)$ be a distribution function with density function h and $H(0) = 1/2$.

(i) *Let $g(x)$ be a function defined on $(-\infty,0]$ such that $\int_{-\infty}^y g(v)^2 h(v)\,\mathrm{d}v > 0$ for every $y \leqslant 0$. Define $W(x) = B(2H(x))$ (for $x \leqslant 0$). Then the process J^- defined as*

$$J^-(x) = W(x) - W(0) + \int_x^0 \left[g(y)h(y) \left(\int_{-\infty}^y g(v)^2 h(v)\,\mathrm{d}v \right)^{-1} \int_{-\infty}^y g(v)\,\mathrm{d}W(v) \right] \mathrm{d}y$$

is a zero-mean Gaussian process on $(-\infty,0]$ with $EJ^-(x)J^-(y) = 1-2H(x \vee y)$. So $J^-(x)$ is time-scaled and time-reversed Brownian motion on $(-\infty,0]$.

(ii) *Let $g(x)$ be a function defined on $[0,\infty)$ such that $\int_y^\infty g(v)^2 h(v)\,\mathrm{d}v > 0$ for every $y \geqslant 0$. Define $W(x) = B(2[1-H(x)])$ (for $x \geqslant 0$). Then the process*

J^+ defined as

$$J^+(x) = W(x) - W(0) - \int_0^x \left[g(y)h(y) \left(\int_y^{\infty} g(v)^2 h(v)\,\mathrm{d}v \right)^{-1} \int_y^{-\infty} g(v)\,\mathrm{d}W(v) \right] \mathrm{d}y$$

is a zero-mean Gaussian process on $[0,\infty)$ *with variance–covariance function* $EJ^+(x)J^+(y) = 2H(x \wedge y) - 1$ *(for* $x, y \geqslant 0$*). Thus* $J^+(x)$ *is a time-rescaled Brownian motion on* $[0,\infty)$.

Remark. We can write $J^-(x) \stackrel{\mathrm{d}}{=} B(1 - 2H(x))$ because they have the same variance–covariance function. Note that the argument of B is $1 - 2H(x)$ not $2[1 - H(x)]$. Similarly, we can write $J^+(x) \stackrel{\mathrm{d}}{=} B(2H(x) - 1)$.

Proof. Part (i) follows from a change in variable ($r = 2H(x)$) and Lemma A.2. Part (ii) follows from a change in variable and Lemma A.1. □

Lemma A.4. Let $B(r)$ *and* $H(x)$ *be the same as in the above lemma. Suppose that* $W_T(x)$ *is a sequence of stochastic process such that* $W_T(x) \Rightarrow B(2H(x))$ *for* $x \leqslant 0$ *and* $W_T(x) \Rightarrow B(2[1 - 2H(x)])$ *for* $x \geqslant 0$*. Define* J_T^- *as in Lemma* A.3 *part* (i) *but with* $W(\cdot)$ *replaced by* $W_T(\cdot)$ *in the transformation. Define* J_T^+ *as in Lemma* A.3 *part* (ii), *but again replacing* $W(\cdot)$ *by* $W_T(\cdot)$*. Then*

$$J_T^- \Rightarrow J^- \stackrel{\mathrm{d}}{=} B(1 - 2H(\cdot))$$

and

$$J_T^+ \Rightarrow J^+ \stackrel{\mathrm{d}}{=} B(2H(\cdot) - 1)$$

Proof. This follows from the continuous mapping theorem and Lemma A.3. Also see the Remark above. □

Note that the sequence W_T with the said property occurs in Lemma 1.

Lemma A.5. Let W_T *satisfy the conditions of Lemma* A.4. *Suppose that* g_T *and* h_T *are estimates of* g *and* h*, respectively, such that*

$$\int_{-\infty}^{\infty} (h_T - h)^2\,\mathrm{d}x = o_{\mathrm{p}}(1) \quad \text{and} \quad \int_{-\infty}^{\infty} (g_T - g)^2\,\mathrm{d}H = o_{\mathrm{p}}(1).$$

Define

$$\tilde{J}_T^-(x) = W_T(x) - W_T(0) + \int_x^0 g_T(y)h_T(y) \left(\int_{-\infty}^{y} g_T(v)^2 h_T(v)\,\mathrm{d}v \right)^{-1} \int_{-\infty}^{y} g_T(v)\,\mathrm{d}W_T(v)\,\mathrm{d}y$$

and

$$\tilde{J}_T^+(x) = W_T(x) - W_T(0) - \int_0^x g_T(y)h_T(y) \left(\int_y^{\infty} g_T(v)^2 h_T(v)\,\mathrm{d}v\right)^{-1} \int_y^{-\infty} g_T(v)\,\mathrm{d}W_T(v)\,\mathrm{d}y.$$

Then

$$\tilde{J}_T^-(x) = J_T^-(x) + o_\mathrm{p}(1), \quad \textit{and} \quad \tilde{J}_T^+(x) = J_T^+(x) + o_\mathrm{p}(1), \tag{A.1}$$

where $o_\mathrm{p}(1)$ *is uniform over* x, *and* J_T^- *and* J_T^+ *are defined in Lemma* A.4. *Therefore*,

$$\tilde{J}_T^-(x) \Rightarrow B(1 - 2H(x)), \quad \text{and} \quad \tilde{J}_T^+(x) \;\Rightarrow\; B(2H(x) - 1). \tag{A.2}$$

Proof. Eq. (A.1) is implied by the result of Bai (2000, Theorem 2). Eq. (A.2) follows from Eq. (A.1) and Lemma A.4. □

This lemma says when g and h are consistently estimated, the limiting distribution will not be affected.

The following lemma is concerned with the residual empirical process. Let $\hat{U}_T^+(x) = T^{-1/2}\sum_{t=1}^T [I(\hat{e}_t \leqslant x) - F(x)]$ be as defined in the text.

Lemma A.6. Under Assumptions A1–A5, *we have*

$$\hat{U}_T^+(x) = U_T^+(x) + f(x)\xi_{1T} + xf(x)\xi_{2T} + o_\mathrm{p}(1), \tag{A.3}$$

where

$$\xi_{1T} = \frac{1}{T}\sum_{t=1}^T \frac{\partial h_t}{\partial \beta}(\beta_0)'\sqrt{T}(\hat{\beta} - \beta_0)/\sigma_t \tag{A.4}$$

and

$$\xi_{2T} = \frac{1}{T}\sum_{t=1}^T \frac{\partial \log \sigma_t(\lambda_0)}{\partial \lambda}\sqrt{T}(\hat{\lambda} - \lambda_0).$$

Proof. Let $\hat{\sigma}_t = \sigma(\tilde{\Omega}_t, \hat{\lambda})$, $\hat{h}_t = h(\tilde{\Omega}_t, \hat{\beta})$, $\sigma_t = \sigma(\Omega_t, \lambda_0)$, and $h_t = h(\Omega_t, \beta_0)$. From $\hat{e}_t = [Y_t - \hat{h}_t]/\hat{\sigma}_t = [\sigma_t e_t - (\hat{h}_t - h_t)]/\hat{\sigma}_t$, we have $\hat{e}_t \leqslant x$ if and only if $e_t \leqslant x(1 + (\hat{\sigma}_t - \sigma_t)/\sigma_t) + (\hat{h}_t - h_t)/\sigma_t$. From Theorem A.2 of Bai (1996), with $a_t = \sqrt{T}(\hat{\sigma}_t - \sigma_t)/\sigma_t$ and $b_t = \sqrt{T}(\hat{h}_t - h_t)/\sigma_t$, we obtain

$$\hat{U}_T^+(x) = U_T^+(x) + f(x)\frac{1}{T}\sum_{t=1}^T b_t + xf(x)\frac{1}{T}\sum_{t=1}^T a_t + o_\mathrm{p}(1).$$

It is easy to show that, under Assumptions A2–A5, $\xi_{1T} - (1/T)\sum_t b_t = o_\mathrm{p}(1)$ and $\xi_{2T} - (1/T)\sum_t a_t = o_\mathrm{p}(1)$. □

Proof of Theorem 2. Note that $\hat{W}_T(x) = \hat{U}_T^+(x) - \hat{U}_T^-(x)$, where $\hat{U}_T^+(x) = T^{-1/2}\sum_{t=1}^T [I(\hat{e}_t \leqslant x) - F(x)]$ and $\hat{U}_T^-(x) = T^{-1/2}\sum_{t=1}^T [I(-\hat{e}_t \leqslant x) - F(x)]$. From $I(-\hat{e}_t \leqslant x) - F(x) = I(\hat{e}_t \geqslant -x) - F(x)$, we have $1 - I(\hat{e}_t < -x) - F(x) = -\{I(\hat{e}_t < -x) - F(-x)\}$ since $1 - F(x) = F(-x)$ under symmetry. Therefore $\hat{U}_T^-(x) = -\hat{U}_T^+(-x)$ (a.s.) under symmetry. Similarly, $U_T^-(x) = -U_T^+(-x)$. It follows from (A.3) that

$$\hat{U}_T^-(x) = U_T^-(x) - f(-x)\xi_{1T} - (-x)f(-x)\xi_{2T} + o_p(1). \tag{A.5}$$

Take the difference of (A.3), (A.5), and use $f(x) = f(-x)$ and $W_T = U_T^+ - U_T^-$, we obtain

$$\hat{W}_T(x) = W_T(x) + 2f(x)\xi_{1T} + o_p(1). \tag{A.6}$$

Next we consider transforming $\hat{W}_T(x) - \hat{W}_T(0)$ for $x \leqslant 0$. Because $f_T(x) - f(x) = o_p(1)$ uniformly in x by Assumption A.6, we can rewrite (A.6) as

$$\hat{W}_T(x) = W_T(x) + 2f_T(x)\xi_{1T} + o_p(1), \tag{A.7}$$

from which we have (subtracting $\hat{W}(0) = W_T(0) + 2f_T(0)\xi_{1T} + o_p(1)$ from above)

$$\hat{W}_T(x) - \hat{W}_T(0) = W_T(x) - W_T(0) + 2[f_T(x) - f_T(0)]\xi_{1T} + o_p(1). \tag{A.8}$$

Define the mapping $\phi_T : \eta \in D[0,1] \to C[0,1]$,

$$\phi_T(\eta)(x) = \int_x^0 g_T(y) f_T(y) \left(\int_{-\infty}^y g_T(v)^2 f_T(v)\,\mathrm{d}v \right)^{-1} \int_{-\infty}^y g_T(v)\,\mathrm{d}\eta(v)\,\mathrm{d}y. \tag{A.9}$$

Then ϕ_T is a linear mapping with $\phi_T(c) = 0$ for any constant c (or random variable not depending on x). In addition, $\phi_T(f_T)(x) = \int_x^0 \dot{f}_T(y)\,\mathrm{d}y = f_T(0) - f_T(x)$. Note that $S_T(x)$ in Eq. (6) can be equivalently rewritten as $S_T(x) = \hat{W}_T(x) - \hat{W}_T(0) + \phi_T(\hat{W}_T)(x)$. By the linearity property of ϕ_T and (A.7), we have

$$\begin{aligned}\phi_T(\hat{W}_T) &= \phi_T(W_T) + \phi_T(f_T)2\xi_{1T} + o_p(1)\\ &= \phi_T(W_T) + [f_T(0) - f_T(x)]2\xi_{1T} + o_p(1).\end{aligned} \tag{A.10}$$

Thus, for $x < 0$,

$$\begin{aligned}S_T(x) &= \hat{W}_T(x) - \hat{W}_T(0) + \phi_T(\hat{W}_T)(x)\\ &= W_T(x) - W_T(0) + 2[f_T(x) - f_T(0)]\xi_{1T} + o_p(1) \quad \text{by (A.8)}\\ &\quad + \phi_T(W_T)(x) + [f_T(0) - f_T(x)]2\xi_{1T} + o_p(1) \quad \text{by (A.10)}\\ &= W_T(x) - W_T(0) + \phi_T(W_T)(x) + o_p(1)\end{aligned}$$

$$= \tilde{J}_T^-(x) + o_\mathrm{p} \quad \text{replacing } h_T \; by \; f_T \text{ in Lemma A.5}$$

$$\Rightarrow B(1-2F(x)) \quad \text{by Lemma A.5.}$$

The proof of weak convergence of $S_T(x)$ (for $x > 0$) to $B(2F(x)-1)$ is the same. The convergence of CS_T^- and CS_T^+ follows from the continuous mapping theorem. This completes the proof of Theorem 2. □

Proof of Theorem 3. Let $K_T(x)=(1-\delta/\sqrt{T})F(x)+\delta/\sqrt{T}H(x)$. By (8), $e_{Tt} \sim K_T(x)$. It is easy to show that $-e_{Tt} \sim G_T(x)$, where $G_T(x)=K_T(x)+(\delta/\sqrt{T})[1-H(x)-H(-x)]$. Define

$$Z_T^+(x) = T^{-1/2}\sum_{t=1}^{T}[I(e_{Tt} \leqslant x) - K_T(x)],$$
$$\hat{Z}_T^+(x) = T^{-1/2}\sum_{t=1}^{T}[I(\hat{e}_{Tt} \leqslant x) - K_T(x)]$$

and

$$Z_T^-(x) = T^{-1/2}\sum_{t=1}^{T}[I(-e_{Tt} \leqslant x) - G_T(x)],$$
$$\hat{Z}_T^-(x) = T^{-1/2}\sum_{t=1}^{T}[I(-\hat{e}_{Tt} \leqslant x) - G_T(x)].$$

Again, using the results of Bai (2000 or 1996), we have

$$\hat{Z}_T^+(x) = Z_T^+(x) + f(x)\xi_{1T} + xf(x)\xi_{2T} + o_\mathrm{p}(1), \tag{A.11}$$

where ξ_{1T} and ξ_{2T} are defined earlier. Eq. (A.11) is similar to (A.5). Note that although $Z_T^+(x)$ involves $K_T(x)$ rather than $F(x)$, we have $K_T(x)=F(x)+O(T^{-1/2})$ and $\mathrm{d}K_T(x)/\mathrm{d}x = f(x) + O(T^{-1/2})$. This explains the presence of $f(x)$ in (A.11). We next consider the asymptotic representation for $\hat{Z}_T^-(x)$. Notice that $I(-\hat{e}_{Tt} \leqslant x) - G_T(x) = 1 - I(\hat{e}_{Tt} < -x) - G_T(x) = -[I(\hat{e}_{Tt} < -x) - K_T(-x)]$ because $1-G_T(x) = K_T(-x)$. Thus $\hat{Z}_T^-(x) = -\hat{Z}_T^+(-x)$ (a.s.), and hence from (A.11) (replacing x by $-x$),

$$\hat{Z}_T^-(x) = Z_T^-(x) - f(-x)\xi_{1T} - (-x)f(-x)\xi_{2T} + o_\mathrm{p}(1). \tag{A.12}$$

Adding and subtracting $K_T(x)$ and $G_T(x)$, we have

$$\hat{W}_T(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(\hat{e}_{Tt} \leqslant x) - I(-\hat{e}_{Tt} \leqslant x)]$$
$$= \hat{Z}_T^+(x) - \hat{Z}_T^-(x) + T^{1/2}[K_T(x) - G_T(x)]$$
$$= Z_T^+(x) - Z_T^-(x) + 2f(x)\xi_{1T} \quad \text{from (A.11) and (A.12)}$$
$$+ \delta[H(x) + H(-x) - 1] + o_\mathrm{p}(1).$$

The last expression follows from $T^{1/2}[K_T(x) - G_T(x)] = \delta[H(x) + H(-x) - 1]$. Let $W_T(x) = Z_T^+(x) - Z_T^-(x)$, then $W_T(x) \Rightarrow B(2F(x))$ for $x \leqslant 0$ and $W_T \Rightarrow B(2[1 - F(x)])$ for $x > 0$. This is true because the finite dimensional convergence and tightness for Z_T^+ and Z_T^- are guaranteed by the standard empirical process theory. Moreover, for $x, y \leqslant 0$, $EW_T(x)W_T(y) = K_T(x \wedge y) - K_T(x)K_T(y) + G_T(x \wedge y) - G_T(x)G_T(y) + G_T(x)K_T(y) + K_T(x)G_T(y) \to 2F(x \wedge y)$ because $K_T(x) \to F(x)$ and $G_T(x) \to F(x)$. This yields the weak convergence of $W_T(x)$ for $x < 0$. Similarly, for $x > 0$, $W_T(x) \Rightarrow B(2[1-F(x)]$. In summary,

$$\hat{W}_T(x) = W_T(x) + 2f(x)\xi_{1T} + \delta[H(x) + H(-x) - 1] + o_p(1), \qquad (A.13)$$

with W_T converging weakly to a (time-rescaled) Brownian motion process for both $x < 0$ and $x > 0$. Subtracting $\hat{W}_T(0)$ from above we obtain

$$\hat{W}_T(x) - \hat{W}_T(0) = W_T(x) - W_T(0) + 2[f(x) - f(0)]\xi_{1T} + \delta v(x) + o_p(1),$$

where $v(x) = H(x) + H(-x) - 2H(0)$. Under the local alternative hypothesis, we can still construct consistent estimates for $f(x)$ and $g(x) = \dot{f}/f$. The reason is that we can write $\hat{e}_{Tt} = \varepsilon_t + O_p(T^{-1/2})$, where $\varepsilon_t \sim F(x)$. To see this, from $e_{Tt} \sim (1-\delta T^{-1/2})F(x) + (\delta T^{-1/2})H(x)$, we can write $e_{Tt} = \varepsilon_t + \eta_{Tt}$, where $\eta_{Tt} = 0$ with probability $1 - \delta/\sqrt{T}$ and $\eta_{Tt} = a_t - \varepsilon_t$ with probability $\delta/\sqrt{T}$, here ε_t and a_t are independent such that $\varepsilon_t \sim F(x)$ and $a_t \sim H(x)$. Hence, $\eta_{Tt} = O_p(T^{-1/2})$. In addition, the estimated residuals satisfy $\hat{e}_{Tt} = e_{Tt} + O_p(1/\sqrt{T})$ and thus $\hat{e}_{Tt} = \varepsilon_t + O_p(1/\sqrt{T})$. Let f_T and g_T are estimates of f and g. Define the mapping ϕ_T as in (A.9). Then using the same argument as in the proof of Theorem 2, we have for $x < 0$,

$$\hat{W}_T(x) - \hat{W}_T(0) + \phi_T(\hat{W}_T)(x) = W_T(x) - W_T(0) + \phi_T(W_T)(x) + \delta v(x) + \delta\phi_T(v)(x) + o_p(1).$$

By Lemma A.4, $W_T(x) - W_T(0) + \phi_T(W_T)(x) \Rightarrow B(1 - 2F(x))$. In addition, $\phi_T(v)(x) \to \phi_v(x)$, which is defined in Theorem 3. Thus,

$$S_T(x) \Rightarrow B(1 - 2F(x)) + \delta v(x) + \delta\phi_v(x),$$

obtaining the result for $x < 0$. Similar argument shows that for $x > 0$,

$$S_T(x) \Rightarrow B(2F(x)) + \delta v(x) - \delta\phi_v^*(x),$$

where

$$\phi_v^*(x) = \int_0^x \left[\dot{f}(y) \left(\int_y^\infty g(z)^2 f(z)\,\mathrm{d}z \right)^{-1} \int_y^\infty g(z)\,\mathrm{d}v(z) \right] \mathrm{d}y.$$

It can be shown that $\phi_v^*(x) = -\phi_v(-x)$ (follows from the fact that f and v are even functions and g is an odd function). This obtains Theorem 3 for $x > 0$. The proof of Theorem 3 is complete. □

Proof of Theorem 4. From $I(-\hat{e}_t \leqslant x) = 1 - I(\hat{e}_t \leqslant x)$ (a.s.), we have

$$\begin{aligned}\hat{W}_T(x) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(\hat{e}_t \leqslant x) - I(-\hat{e}_t \leqslant x)]\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(\hat{e}_t \leqslant x) - F(x)]\\ &\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(\hat{e}_t \leqslant -x) - F(-x)] + \sqrt{T}[F(x) + F(-x) - 1]\\ &= \hat{U}_T^+(x) + \hat{U}_T^+(-x) + \sqrt{T}[F(x) + F(-x) - 1],\end{aligned}$$

where $\hat{U}_T^+(x)$ is defined earlier. Apply Lemma A.6 twice with x and $-x$, respectively, we have

$$\begin{aligned}\hat{W}_T(x) &= W_T(x) + [f(x) + f(-x)]\xi_{1T} + x[f(x) - f(-x)]\xi_{2T}\\ &\quad + \sqrt{T}v_1(x) + o_{\mathrm{p}}(1),\end{aligned}$$

where $v_1(x) = F(x) + F(-x) - 1$ and $W_T(x) = U_T^+(x) + U_T^+(-x)$, which again converges weakly to a Brownian motion. In passing, it is pointed out that the above equation reduces to (A.7) under the null hypothesis of symmetric distribution. Similar to (A.8), we have

$$\begin{aligned}\hat{W}_T(x) - \hat{W}_T(0) &= W_T(x) - W_T(0) + [f(x) + f(-x) - 2f(0)]\xi_{1T}\\ &\quad + x[f(x) - f(-x)]\xi_{2T} + \sqrt{T}v(x) + o_{\mathrm{p}}(1), \qquad \text{(A.14)}\end{aligned}$$

where $v(x) = F(x) + F(-x) - 2F(0)$. To obtain $S_T(x)$, we apply the transformation $\phi_T(\cdot)$ [see (A.9)] to $\hat{W}_T(x) - \hat{W}_T(0)$ for $x < 0$. Except for the term $\sqrt{T}v(x)$, all terms on the right-hand side of (A.14) are stochastically bounded. They are still stochastically bounded after the martingale transformation (it is no longer important as to whether the terms involving ξ_{1T} and ξ_{2T} can be eliminated from the transformation). The dominating term after transformation is $\sqrt{T}[v(x) + \phi_T(v)(x)]$. From $\phi_T(v)(x) \to \phi_v^-(x)$, which is defined in Theorem 4, we have $S_T(x) = \sqrt{T}[v(x) + \phi_v^-(x)] + O_{\mathrm{p}}(1)$. This completes the proof of Theorem 4 for $x < 0$. The proof for $x > 0$ is similar and is omitted. Unlike the case of local alternatives, $\phi_v^-(x)$ and $\phi_v^+(x)$ do not necessarily have any relationship because f is no longer an even function under the alternative hypothesis. □

Appendix B. Computation of the statistic

To compute the test statistic, we need to evaluate $\hat{W}_T(x)$ and $h_T^{\pm}(x)$. For every given x, $\hat{W}_T(x)$ is simply the difference between the number of $\hat{e}_t$ and the number of $-\hat{e}_t$ less than or equal to x, then divided by $\sqrt{T}$. To evaluate $h_T^{\pm}(x)$, we need to estimate the density f and its derivative g. They are estimated by nonparametric method (see below for details), and are denoted by f_T and g_T, respectively. Consider first the terms $\int_{-\infty}^{y} g_T(z)\,\mathrm{d}\hat{W}_T(z)$ and $\int_{y}^{\infty} g_T(z)\,\mathrm{d}\hat{W}_T(z)$. Note that these can equivalently be represented as

$$\int_{-\infty}^{y} g_T\,\mathrm{d}\hat{W}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[g_T(\hat{e}_t)I(\hat{e}_t \leqslant y) - g_T(-\hat{e}_t)I(-\hat{e}_t \leqslant y)];$$

and

$$\int_{y}^{\infty} g_T\,\mathrm{d}\hat{W}_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[g_T(\hat{e}_t)(\hat{e}_t \geqslant y) - g_T(-\hat{e}_t)I(-\hat{e}_t \geqslant y)].$$

Next, the integration (over z) of $g_T(z)^2 f_T(z)$ is approximated by summations. After obtaining $h_T^{\pm}(y)$ [see Eqs. (6) and (7)], the integration of $\int_x^0 h_T^{\pm}(y)\,\mathrm{d}y$ is also approximated by summations. This makes the computation straightforward. Simulations show that the size and power of the tests are not affected by these approximations. Finally, the maximum of $S_T(x)$ is obtained by searching over $2T$ ordered data points of $\hat{e}_t$ and $-\hat{e}_t$ $(t = 1, 2, \ldots, T)$.

When estimating the density and its derivative, we use the Gaussian kernel with a plug-in bandwidth as discussed in Silverman (1986). For the Gaussian kernel, the bandwidth which minimizes the approximate mean integrated squared error in estimating the density is given by $1.06\sigma T^{-1/5}$, where T is the sample size, and σ is the standard error of the variable whose density is to be estimated. All the simulation and empirical results are obtained using this systematic choice of bandwidth. However, in results unreported, we increase and decrease the bandwidth by as much as 20% and the size and power functions are robust to such variations in the bandwidth. Gauss and Splus programs are available from the authors on request.

References

Ahmad, I.A., Li, Qi., 1996. Testing symmetry of an unknown density function by kernel method. Manuscript, Department of Economics, University of Guelph.

Bai, J., 1996. Testing for parametric constancy in linear regressions: an empirical distribution function approach. Econometrica 64, 597–622.

Bai, J., 2000. Testing parametric conditional distributions of dynamic models. Manuscript, Department of Economics, Boston College.

Bickel, P.J., 1982. On adaptive estimation. Annals of Statistics 10, 647–671.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542–547.

Brown, B.W., Newey, W.K., 1998. Efficient Bootstraps for Semiparametric Models. Mimeo, MIT, Cambridge, MA.

Campbell, J.Y., Hentschel, L., 1992. No news is good news. Journal of Financial Economics 31, 281–318.
Delong, J.B., Summers, L.H., 1982. Are business cycle symmetrical. In: Gordon, R.J. (Ed.), American Business Cycle: Continuity and Change. University of Chicago Press, Chicago.
Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of variance of UK inflation. Econometrica 50, 987–1007.
Fan, Y., Gencay, R., 1995. A consistent nonparametric test of symmetry in linear regression models. Journal of the American Statistical Association 90, 551–557.
French, K., Schwert, G., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.
Hodgson, D.J., 1998. Adaptive estimation of error correction models. Econometric Theory 14, 44–69.
Hsieh, D.A., 1988. The statistical properties of daily foreign exchange rates: 1974–1983. Journal of International Economics 24, 129–145.
Kendall, A., Ord, K., 1994. Advanced Theory of Statistics. Edward Arnold, London.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-tests. Theory of Probability and its Applications XXVI, 240–257.
Kriess, J.P., 1987. On adaptive estimation in stationary ARMA processes. Annals of Statistics 15, 112–133.
Lee, S., Hansen, B.E., 1994. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory 10, 29–52.
Linton, O., 1993. Adaptive estimation in ARCH models. Econometric Theory 9, 539–569.
Lumsdaine, R.L., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.
Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.
Newey, W.K., 1988. Adaptive estimation of regression models via moment restrictions. Journal of Economics 38, 301–339.
Newey, W.K., Powell, J.L., 1988. Asymmetric least squares estimation and testing. Econometrica 55, 819–847.
Newey, W.K., Steigerwald, D.G., 1997. Asymptotic bias for quasi-maximum likelihood estimators in conditional heteroskedastic models. Econometrica 65, 587–599.
Parzen, E., 1962. On estimation of a probability density and mode. Annals of Mathematical Statistics 35, 1065–1076.
Randles, R.H., Fligner, M.A., Policello, G.E., Wolfe, D.A., 1980. An asymptotically distribution-free test for symmetry versus asymmetry. Journal of the American Statistical Association 75, 168–172.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Zheng, J.X., 1998. Consistent specification testing for conditional symmetry. Econometric Theory 14, 139–149.