A Test for Conditional Symmetry in Time Series Models*

Jushan Bai [†] Serena Ng [‡]

August 26, 1998

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 could be infinite. The size and power of the test are satisfactory even for small samples. In addition, the proposed test is shown to have non-trivial power against root-T local alternatives. Applying the test to various time series, we reject conditional symmetry in inflation, exchange rate and stock returns. Among the non-financial time series considered, we find that investment, the consumption of durables, and manufacturing employment also reject conditional symmetry. Interestingly, these are series whose dynamics are believed to be affected by fixed costs of adjustments.

JEL Classification: C12; C22.

Key Words and Phrases: conditional symmetry, empirical distribution function, kernel estimation, Brownian motion, ARCH/GARCH, nonlinear time series.

---

*We thank seminar participants at the Camp Econometrics for useful comments. This paper is also prepared for the NBER/NSF Time Series Seminar to be held in Chicago in September, 1998. The first author acknowledges financial support from the National Science Foundation under Grant SBR-9709508.

[†]Department of Economics, Boston College, Email: jushan.bai@bc.edu.

[‡]Department of Economics, Boston College. Email: serena.ng@bc.edu

Correspondence: Jushan Bai, Department of Economics, Boston College, Chestnut Hill, MA, 02467.

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 conditional distribution of Y_t, conditional on X_t, is symmetric with respect to the conditional mean $E(Y_t|X_t)$. More formally, consider the following nonlinear time series regression model:

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

where $h(X_t, \beta)$ is the conditional mean, $\sigma^2(X_t, \lambda)$ is the conditional variance, and e_t are disturbances with zero mean and unit variance and are 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 cdf of e_t, respectively.

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

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

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

$$Y_t = X_t'\beta + \sigma_t e_t$$

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

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

where $\lambda = (\alpha, \delta, \gamma)$. The test statistic to be developed in this article can still be applied even though the dimension of the conditioning variables in these cases is infinite.

The rest of this paper is organized as follows. A review of applications and some macroeconomic motivations for interest in conditional symmetry are given in Section 2. In Section 3, we propose a test statistic for conditional symmetry, analyze its asymptotic properties, and highlight its generality vis-á-vis alternative tests in the literature. Simulation experiments and empirical applications are provided in Section 5, and Section 6 concludes.

2 The Use of Conditional Symmetry in Estimation and Inference

From a statistical point of view, the primary interest in whether or not $f(e)$ is symmetric arises because when the symmetry assumption fails, the mean is no longer the only natural measure of the location of a distribution. This has important implications in a number of contexts. In this section, we discuss the importance of conditional symmetry for i) use of the QMLE in estimating models with time varying volatility, ii) adaptive estimation, and iii) efficient bootstrapping. Relevance of the symmetry assumption to macroeconomic applications, as well as the distinction between conditional and unconditional symmetry, are also discussed.

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). While these models are usually estimated by quasi-maximum likelihood (QMLE), the asymptotic properties of the QMLE estimator is known only for the special case of a Gaussian likelihood. In particular, Lee and Hansen (1991) and Lumsdaine (1996) showed that when the model correctly specifies both the conditional mean and the conditional variance, the QMLE estimator is consistent for the parameters of the GARCH(1,1) model with a Gaussian likelihood even when the assumption of normality is false.

An increasing number of applications has, however, assumed a non-Gaussian likelihood. 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. These studies were motivated by the fact that the innovations in financial time series usually have fat tails and are sometimes asymmetric. See, for example, Diebold (1988) in the context of exchange rates. In a recent paper, Newey and Steigerwald (1997) studied the conditions under which a non-Gaussian QMLE is consistent. The key for consistency is what the authors referred to as an identification condition which requires that the quasi-log likelihood has a unique maximum at the true conditional mean and the relative scale parameter. The identification issue arises because non-Gaussian densities are not necessarily best summarized by the natural location (the mean) and scale (the standard deviation), and in such cases, side conditions are necessary for consistent estimation and identification of the ARCH/GARCH parameters.

Newey and Steigerwald showed formally that when the likelihood is non-Gaussian, the identification condition can still be satisfied if both the true and the assumed innovation density is symmetric around zero and is unimodal.[1] Intuitively, conditional symmetry facilitates identification because under symmetry, the mean, median, and mode (assuming unimodality) of the innovations coincide. In consequence, the conditional mean is restored as the natural location parameter. Newey and Steigerwald also showed that consistency of the QMLE estimates can still be obtained if conditional symmetry fails, but that it would require an additional parameter which identifies the location of the innovation distribution to be specified and estimated. Of course, the QMLE estimates will be less efficient when symmetry holds but the additional location parameter is estimated. A direct use of our proposed test for conditional symmetry is to determine whether estimation of this additional location parameter is necessary.

The assumption of conditional symmetry is also important for adaptive estimations. Suppose we are interested in estimating a set of parameters, θ, associated with a model with innovations e whose shape is unknown. The estimation of θ is said to be adaptive when the information bound on θ is the same whether or not the density of e is known. That is to say, an adaptive estimator shares the same asymptotic optimality properties as a maximum likelihood estimator. Thus, an adaptive estimator can be seen as a MLE estimator when the shape of the likelihood is unknown. Bickel (1982) considered the conditions for adaptation in the context of a semi-parametric model $P_{\theta,G}$ characterized by a set of finite dimensional parameters, θ, and a shape nuisance parameters, G. Bickel established that the necessary condition for adaptation is the orthogonality between the scores for θ and the scores for the scalar parameter τ, where τ parameterizes $G(\cdot)$. In a linear regression setting, Bickel showed that, if conditional symmetry holds, the slope parameters can achieve the same information bound whether or not the error density is known. Bickel's analysis was extended to a number of time series models. For (conditionally) homoskedastic ARMA models, Kriess (1987) showed that the parameters can be estimated adaptively if e is conditionally symmetric. Linton (1993) discussed a

[1] The exception is the special case when the true conditional mean is centered around zero.

reparameterization of an ARCH process to achieve adaptation. Adaptive estimation of error correction models was discussed in Hodgson (1998), and Newey (1988) showed that the parameters of a linear regression model can be estimated adaptively by generalized methods of moments. In each of these applications, adaptation is achieved under the maintained assumption of conditional symmetry. The usefulness of this assumption arises because when e is symmetric, the scores are antisymmetric around zero[2], thus satisfying the Bickel orthogonality condition.[3]

Knowledge about the properties of e_t also has efficiency implications for bootstrapping. The general bootstrap procedure for nonparametric and semiparametric 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 $\hat{e}_t$.

To be precise, suppose the interest is in the critical values of the t-statistic associated with the model $y_t = X_t'\beta + e_t$. The typical bootstrap procedure is to obtain the j-th draw (y_t^j, x_t^j) from $[(y_1, x_1), (y_2, x_2), \ldots, (y_T, x_T)]$ with equal probability $1/T$, calculate the t statistic each time and resample J times. With this method, the empirical critical values for the t-statistic are obtained without assumptions made about the distributional properties of e_t. But, suppose it is known that e_t is symmetrically distributed. Then one can exploit this structure by drawing $2T$ points, $[(\hat{e}_1, X_1), \ldots, (\hat{e}_T, X_T), (-\hat{e}_1, X_1), \ldots, (-\hat{e}_T, X_T)]$ for each resampled set of data. This is analogous to assuming that y_t comes from a two-point conditional distribution. If this parametric assumption is correctly imposed, the bootstrapped standard errors will be more efficient because more information is used to bootstrap the critical values. Evidently, efficiency gains are possible if we know the distribution of e_t is symmetric.

2.1 Conditional Asymmetry versus Business Cycle Asymmetries

Apart from the statistical implications, whether or not conditional symmetry holds is an issue that is of macroeconomic interest in its own right. 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 the sign of the forecast errors

[2]Let $\psi : R \to r$. ψ is antisymmetric if $\psi(y) = -\psi(-y)$.

[3]The assumption of symmetry is sufficient but not necessary for adaptation, see, e.g., Gonzalez-Rivera (1997). However, she shows that adaptation holds only for a narrow class of nonsymmetric densities.

are not equally likely. As well, the distributional properties of e_t could be useful to our understanding of the impulse and propagating mechanism of macroeconomic dynamics.

Several other notions of asymmetry have also been used in macroeconomics. Beaudry and Koop (1993), for example, were interested in asymmetry in persistence. That is, whether the dynamic response of output to positive and negative shocks is the same. In our notation, asymmetry in persistence arises when β in the conditional mean model depends on the sign of e_t. Clearly, asymmetry in persistence can occur in the presence of conditional symmetry. Others such as Neftci (1994) and Hamilton (1989) asked whether the behavior of GDP during expansions is similar to that during recessions. Since these studies are mainly concerned with symmetry of the series itself rather than its innovations, unconditional symmetry is arguably the object of interest. Except in special cases such as one which will be discussed below, conditional symmetry does not, in general, imply unconditional symmetry. A test for unconditional symmetry will be developed in our companion paper.

There are, however, instances when economic behavior and/or structure naturally gives rise to conditional asymmetry. A specific example is given by the "No news is good news" model of Campbell and Hentschel (1992). The authors' objective was to provide a formal explanation for the correlation between volatility and returns. Their basic motivation was that a large piece of good news about future dividends will increase future expected volatility, lowers the stock price, and dampens the positive impact of the dividend news. In contrast, a large piece of negative news will also increase volatility and lower the stock price, but now it exaggerates the negative impact of the dividend news. In consequence, "volatility feedback" will generate stock price movements that are correlated with future volatility, giving rise to the phenomenon of "predictive asymmetry". Campbell and Hentschel formally showed, using a quadratic ARCH specification for volatility, that the residuals of returns, conditional on volatility, should be skewed. In other words, the residuals in a model of log returns on volatility should be asymmetrically distributed. This is precisely our notion of conditional asymmetry.

3 The Test Statistic

Skewness, or the third moment, is perhaps the statistic that naturally comes to mind when the object of interest is symmetry of a distribution. Hsieh (1988), for example, performed diagnostics on the standardized estimated residuals using the coefficient of skewness. However, tests based on the skewness coefficient will not be consistent since there are many distributions that are asymmetric and yet their skewness coefficients are zero. 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. Fan and Gencay (1995) proposed a test based on the idea that under symmetry, $2\int f(x)f(-x)dx = \int f^2(x)dx + \int f^2(-x)dx$. Ahmad and Li (1996)'s test is based on $\int_{-\infty}^{\infty}[f(x) - f(-x)]^2 dF = 0$. Zheng (1998), on the other hand, constructed a test on the basis that under symmetry, $1 - F(-x) - F(x) = 0$. All these tests (as does ours) only require consistent estimates of the parameters β. As well, the unknown density function $f(x)$ in these tests is estimated by the kernel smoothing method. Regression quantile estimators also contain information about conditional symmetry. If $\beta(\tau)$ is the quantile regression estimator for $\tau \in (0,1)$, $\beta(\tau) + \beta(1-\tau) = 2\beta(1/2)$ under symmetry. Newey and Powell (1988) considered a test with this as a starting point, but used a somewhat different criterion function than quantile regressions to obtain "asymmetric least squares estimates", from which the test for conditional symmetry is based.

There are several important aspects that distinguish the test proposed in this paper from the aforementioned tests. First, previous authors consider only i.i.d. data and the results for time series observations are not available in the literature. 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 non-stationary. Second, the conditioning variables permitted in our framework can be infinite dimensional, allowing the use of ARMA as well as GARCH models as the specification for the conditional mean and the conditional variance. Third, most of the existing tests are based on comparisons of nonparametrically estimated density functions. Due to nonparametric smoothing, these tests do not have root-T local power. Our test is based on empirical distribution functions and has nontrivial power against root-T local alternatives.

Suppose $\{e_t, t = 1, \ldots, T\}$ is *i.i.d.* with density $f(e)$, distribution $F(e)$, 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 function of e_t $(t = 1, ..., T)$ and that of $-e_t$ $(t = 1, ..., T)$. Define the empirical process, $U_T^+(x)$, based on e_t, as

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

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

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

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

$$W_T(x) = U_T^+(x) - U_T^-(x) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big[I(e_t \leq x) - I(-e_t \leq x) \Big] \tag{3}$$

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 values of e_t and the number of positive values of e_t. Under symmetry, $W_T(x)$ should be small at all values of x.

In view of the mathematical identity

$$W_T(x) = W_T(-x), \quad (a.s.)$$

one can consider either positive or negative values of x in the construction of W_T. We have the following:

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

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, $B(2[1 - F(x)])$ $(x \geq 0)$ is a time-reversed Brownian motion on $[0,1]$.

If e_t were observed, the $\max |W_T(x)|$ statistic could be used as a test for symmetry. Furthermore, critical values are readily available since the distribution of the maximum of a Brownian motion is well known. But $\{e_t\}$ is the sequence of innovations of a nonlinear time series model, which we do not observe. Therefore, we consider a feasible statistic based upon the estimated residuals, $\hat{e}_t$, and use martingale transformation methods to obtain a test that is asymptotically distribution free. The transformation method was first studied by Khmaladze (1981) and was recently extended in several directions by Bai (1997).

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

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

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

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

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,

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

where $f(x)$ is the density of e_t and ξ_{1T}, given in (11) in the Appendix, is a stochastically bounded random variable (that does not depend on x). Since the limiting distribution of $\hat{W}_T(x)$ depends on f as well as the estimated parameters, the limiting distribution (and hence critical values) will not be asymptotically distribution free.

To circumvent this problem, we use the martingale transformation method (see, Bai (1997)) 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 such that

$$\int_{-\infty}^{\infty} (f_T - f)^2 dx = o_p(1), \quad \text{and} \quad \int_{-\infty}^{\infty} (g_T - g)^2 dF = o_p(1). \tag{4}$$

We use kernel method to construct f_T and g_T from the residuals $\hat{e}_t$. For $x \le 0$, define

$$S_T(x) = \hat{W}_T(x) - \hat{W}_T(0) + \int_x^0 h_T^-(y)dy \tag{5}$$

where

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

For $x > 0$, define

$$S_T(x) = \hat{W}_T(x) - \hat{W}_T(0) - \int_0^x h_T^+(y)dy \tag{6}$$

where

$$h_T^+(y) = g_T(y) f_T(y) \Big[\int_y^\infty g_T(z)^2 f_T(z) dz \Big]^{-1} \int_y^\infty g_T(z) d\hat{W}_T(z).$$

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

$$CS_T^+ = \max_{x \geq 0} |S_T(x)|,$$

$$CS_T^- = \max_{x \leq 0} |S_T(x)|.$$

Theorem 1 *Under assumptions A1-A6 in Appendix A and conditional symmetry, we have*

$$\begin{aligned} S_T(x) &\Rightarrow B(1 - 2F(x)), \qquad x \leq 0 \\ S_T(x) &\Rightarrow B(2F(x) - 1), \qquad x > 0 \\ CS_T^- &\xrightarrow{d} \max_{0 \leq s \leq 1} |B(s)|, \\ CS_T^+ &\xrightarrow{d} \max_{0 \leq s \leq 1} |B(s)|. \end{aligned}$$

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

The proof of the theorem assumes a very general specification of the conditional mean of the form $h(\Omega_t, \beta)$ and a GARCH(1,1) error process. Extension to general GARCH(p,q) is straightforward. The asymptotic critical values of the test can be obtained from its analytical density function of the random variable $\max_{0 \leq s \leq 1} |B(s)|$. Alternatively, they can be easily obtained by simulation. The critical values corresponding to 1%, 5%, and 10% levels of significance are 2.78, 2.21, and 1.91, respectively. Although the terms involved in the tests seem complicated, they can be easily computed as discussed in the Appendix B.

The theorem suggests that one can use either CS_T^- or CS_T^+ to test for conditional symmetry. This result arises because f is an even function and g is an odd function under the null hypothesis. It can easily be shown that if f and g were used in the transformations and f is even and g is odd, then we would have $S_T(x) = S_T(-x)$ for all x and thus the exact relationship $CS_T^+ = CS_T^-$. Because f_T and g_T are consistent for f and g, it can be shown that the transformation based on f_T and g_T is asymptotically equivalent to that based on f and g (see Lemma 6 in the Appendix). This implies that $S_T(x) = S_T(-x) + o_p(1)$, where $o_p(1)$ is uniform over x. Therefore, $CS_T^- = CS_T^+ + o_p(1)$.

This means that not only CS_T^- and CS_T^+ have the same asymptotic distribution, but also they are asymptotically equivalent. These results also suggest a third alternative test, defined as the maximum of CS_T^- and CS_T^+:

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

since it has the same distribution as CS_T^+ and CS_T^-. We state this result as a corollary.

Corollary 1 *Under the null hypothesis of conditional symmetry and the conditions of Theorem 1, we have*

$$CS_T \xrightarrow{d} \sup_{0 \le s \le 1} |B(s)|.$$

The CS_T statistic has two advantages and is our preferred statistic. First, it has better power since under the alternative, the equivalence of CS_T^- and CS_T^+ breaks down. Second, even if the null is true, for finite samples, f_T may not be exactly even and g_T may not be exactly odd, and thus we do not expect that CS_T^- and CS_T^+ to be exactly the same. By using the statistic CS_T, the user is free from having to decide which test to use. Simulations show that the CS_T test has substantial power advantages over CS_T^- and CS_T^+.

It is instructive to examine graphically how the untransformed empirical process $\hat{W}_T(x)$ and the transformed process $S_T(x)$, both based on estimated residuals, differ from $W_T(x)$ based on the true residuals. To this end, we use two samples of normal and two samples of $\chi^2_{(2)}$ observations to evaluate the three processes. Each sample consists of 100 observations of a standardized random variable. That is, $Y_i \sim N(0,1)$ or $Y_i \sim (\chi^2_{(2)} - 2)/\sqrt{2})$. The residual is 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. Figures 1 and 2 plot the three processes evaluated at 200 points, x_k $(k = 1, 2, \ldots, 200)$ of which half are positive and half are negative. In addition, these points are located symmetrically around zero. The dashed line and the light-solid 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. Recall that $W_T(x)$ and $\hat{W}_T(x)$ are symmetric about zero, therefore their graph should be symmetric about the middle point x_k for $k = 100$. If the null hypothesis is true, the process S_T should almost be symmetric about the middle point, and under the alternative hypothesis, S_T will not be symmetric. These features are all confirmed by the two figures.

The 95% confidence interval is given by [-2.2, 2.2] and is also shown in the graphs. The departure of $\hat{W}_T(x)$ from $W_T(x)$ in all cases is apparent and indicates the effects of parameter estimation. 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. We see that in the normal case, $S_T(x)$ and $W_T(x)$ are quite close to each other, showing the effectiveness of the martingale transformations. In particular, the test statistic $CS_T = \max|S_T(x)|$ is close to $\max|W_T(x)|$. In the normal case, symmetry cannot be rejected.

For the case of χ^2 observations, the $W_T(x)$ process indicates strong evidence of asymmetry in the first sample but weaker, albeit significant, evidence of asymmetry in the second. Note that $W_T(x)$ is not observable for general models in practice. If one uses $\max|\hat{W}_T(x)|$ as a test statistics, one would falsely reject symmetry for the first sample because $\hat{W}_T(x)$ evidently lies within the standard error bands for all values of x. However, $CS_T = \max|S_T(x)|$ clearly lies outside the confidence band and the statistic correctly rejects symmetry. In the second sample, the transformed process $S_T(x)$ shows stronger evidence of asymmetry than implied by $W_T(x)$. These results show that the proposed test has power. An analysis of local asymptotic power will formally be given in Theorem 2 below.

3.1 Local Power Analysis

As discussed earlier, existing tests of symmetry based entirely on estimated densities do not have root-T local power. This is because root-T local departure from a symmetric density will be smoothed away by kernel smoothing and the resulting 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 has local power. To show that the proposed test has non-trivial power against root-T local alternatives, we consider alternatives for which the disturbances e_t form a triangular array. The distribution function of this array is described by, for $t = 1, 2, ..., T$,

$$e_{Tt} \sim (1 - \delta/\sqrt{T})F(x) + \delta/\sqrt{T}H(x) \tag{7}$$

where F is the distribution function of a symmetric random variable and H is that of a non-symmetric random variable and hence $1 - H(x) - H(-x) \not\equiv 0$. Define $v(x) = H(x) + H(-x) - 2H(0)$. Then it is easy to show that $v(x) \equiv 0$ if and only if $H(x)$ is a distribution function of a symmetric random variable. By the assumption on $H(x)$, it follows that $v(x) \not\equiv 0$. In addition, we assume H satisfies assumption A1 imposed on F.

Theorem 2 *Assume A1-A6 hold. Under the local alternative of (7), we have*

$$S_T(x) \Rightarrow B(1-2F(x)) + \delta v(x) + \delta\phi^-(x), \qquad x < 0$$

$$S_T(x) \Rightarrow B(2F(x)-1) + \delta v(x) - \delta\phi^+(x), \qquad x \geq 0$$

where

$$\phi^-(x) = \int_x^0 \dot{f}(y)(\int_{-\infty}^y g(z)^2 f(z)dy)^{-1}\int_{-\infty}^y g(z)dv(z)dy$$

$$\phi^+(x) = \int_0^x \dot{f}(y)(\int_y^\infty g(z)^2 f(z)dy)^{-1}\int_y^\infty g(z)dv(z)dy$$

Since the limiting distribution is different under the local alternatives, this implies that the test statistic has non-trivial local power.

4 Simulations

In this section, we first present simulations to assess the size and power of the tests when the conditional model includes just a constant. 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, Fligner, Policello and Wolfe (1980). We then conduct simulations for time series regression models commonly encountered in economic analysis.

4.1 Testing for Symmetry in the Demeaned Series

The symmetric distributions we consider are:

1. S1: $N(0,1)$;
2. S2: t_5;
3. S3: $e_1 I_{z\leq .5} + e_2 I_{z>.5}$, where $z \sim U(0,1)$, $e_1 \sim N(-1,1)$, and $e_2 \sim N(1,1)$;
4. S4: $\lambda_1 = 0$, λ_2=.19754, λ_3=.134915, λ_4=.134915;
5. S5: $\lambda_1 = 0$, λ_2=-1, λ_3=-.08, λ_4=-.08;
6. S6: $\lambda_1 = 0$, λ_2=-.397912, λ_3=-.16, λ_4=-.16;

7. S7: $\lambda_1 = 0$, λ_2=-1, λ_3=-.24, λ_4=-.24;

The size of the test is assessed by considering the number of rejections using the asymptotic critical values of 2.78, 2.20, and 1.91 at the 1, 5, and 10 percent levels respectively. To conserve space, we only report the results for the 5% test in Table 1. Let α_3 denote the coefficient of skewness and α_4 of kurtosis. These parameters are also reported in Table 1 for convenience. In all cases, we use only a constant as the conditioning variable. The estimated residuals are centered around the mean and standardized before applying the tests. Thus, we use the mean as the location parameter to test for conditional symmetry of the demeaned series around zero. For notational simplicity, we shall drop the subscript T associated with the test statistics in the following discussion.

The results in Table 1 indicate that the CS test generally has good size. The exceptions are in the S6 and S7 distributions when the sample size is small. These two distributions have large kurtosis, and the results suggest that a larger sample size might be required for the CS test to be accurate when a symmetric distribution has heavy tails. It is also clear from the results that both CS^- and CS^+ tend to be undersized.

The power of the tests is assessed by considering the following asymmetric series:

1. A1: lognormal: $\exp(e)$, $e \sim N(0,1)$;
2. A2: χ^2_2;
3. A3: exponential: $-\ln(e)$, $e \sim N(0,1)$,
4. A4: $\lambda_1 = 0$, λ_2=1.0, λ_3=1.4,$\lambda_4 = .25$;
5. A5: $\lambda_1 = 0$, λ_2=-1, λ_3=-.0075, λ_4=-.03;
6. A6: $\lambda_1 = 0$, λ_2=-1, λ_3=-.1, λ_4=-.18;
7. A7: $\lambda_1 = 0$, λ_2=-1, λ_3=-.001, λ_4=-.13;
8. A8: $\lambda_1 = 0$, λ_2=-1, λ_3=-.0001, λ_4=-.17;

The simulation results are given in Table 2. The result that stands out is that the CS statistic has substantially more power than the CS^- or the CS^+ tests. This is to be expected since the CS test rejects conditional symmetry if CS^- or CS^+ rejects, or both. This gain in power is non-trivial. For example, in the χ^2_2 case, the CS rejects over 90% while the CS^+ rejects only 54% of the time in small samples.

The tests generally have good power even when the sample size is small. For example, the distributions A7 and A8 were also considered in Zheng (1998), but while Zheng's test records power of about 20% for a sample size of 100 in both cases, our test has power over 90%. 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. All the tests considered have low power for cases A4, A5 and A6 unless the sample size is large (say, more than 200 observations). These two distributions are only mildly asymmetric. Note that it is not so much mild asymmetry per se that causes the tests to have low power, but rather that power is low when the kurtosis to skewness ratio is large. Interestingly, the test developed by Zheng (1998) also has power problems in this parameter space. As well, the results of Randles et al. (1980) for testing unconditional symmetry in *i.i.d.* data also exhibit the same phenomenon. In all, our test 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 test in a more general setting, we consider the following data generating processes:

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): $y_t = \rho y_{t-1} + e_t$, $\rho = .5, .8$;

3. MA(1): $y_t = e_t + \rho e_{t-1}$, $\rho = .5, .8$;

4. GARCH(1,1): $y_t = 1 + u_t$, $u_t = \sqrt{h_t} e_t$, $h_t = \phi_0 + \phi_1 h_{t-1} + \phi_2 u_{t-1}^2$; $\phi_0 = 2$, $\phi_1 = .5$, ϕ_2=.3 and .45,

where e_t is drawn from one of the following six distributions:

1. $e_t \sim N(0, 1)$;

2. $e_t \sim t_5$;

3. $e_t \sim N(-1, 1)I_{z<.5} + N(-1, 1)I_{z>=.5}$, $z \sim U(0, 1)$;

4. $e_t \sim \chi_2^2$;

5. A7: $\lambda_1 = 0$, λ_2=-1, λ_3=-.001, λ_4=-.13;

6. A8: $\lambda_1 = 0$, λ_2=-1, λ_3=-.0001, λ_4=-.17;

These are, respectively, normal, t-distribution, mixture normal, chi-square, and two lambda distributions. The first three are symmetric and are used to assess size, while distributions four through six are asymmetric and are used to assess power. After e_t is drawn, the population mean and standard deviation of e_t are used to standardize the series.

The results for the CS test are reported in Tables 3 through 6, noting that the (unreported) CS^+ and CS^- have similar properties. The results in Table 3 are based on a non-dynamic regression model that has a constant and k *i.i.d.* variables as regressors. Compared to the results in Table 2 which did not include the random regressors, power is lower when $T = 50$. Furthermore, power decreases as the number of regressors increases. The test developed in Fan and Gencay (1995) exhibit the same property. However, this is a small sample phenomenon because power is comparable to those in Table 2 at larger sample sizes. Thus, for sample sizes that we are likely to encounter in economic analysis, increasing the number of regressors should not have implications for power.

Models two, three and four are dynamic models and these results are reported in Tables 4 through 6. The test generally has good size. The probability of rejecting symmetry is close to 100 percent when $T \geq 100$. Even when T is small, the power is usually well over 70 percent. Furthermore, the results are robust even when the AR or the MA root is near unit circle. The results are also robust when the error process is close to being an IGARCH.

4.3 Empirical Applications

The tests are applied to seventeen macroeconomic time series. Data for GDP, the GDP deflator, the consumption of durables, final sales, the consumption of non-durables, residential investment, and non-residential investment are taken from the national accounts and are therefore quarterly data. The various exchange rates, the unemployment rate, employment, M2, CPI are monthly series. The 30 day interest rate, and M2 are weekly data.[4] With the exception of the interest rate and the unemployment rate, we take logarithms of the data. All data are then first differenced, and in each case, the conditional mean is estimated using an AR(2) model. The residuals are then used to test for conditional symmetry. We also considered the CRSP (equal weight) daily stock returns (not differenced) to which we fit an AR(2) model.

[4]All data are taken from the Economic Time Series Page, and URL is: vos.business.uab.edu/data.data/htm.

The results in Table 7 find evidence of conditional asymmetry at the 1% level in the Japan-US exchange rate and stock returns, and at the 5% level for the consumption of durables and manufacturing employment. We reject conditional symmetry in the CPI inflation series and non-residential investment at the 10% level. The evidence for the Japan-US exchange rate, stock returns, and durables is particularly convincing because both the CS^+ and CS^- reject the null hypothesis. Hsieh (1988) has presented evidence for skewness in daily exchange rates, and French, Schwert and Stambaugh (1987) find skewness in stock returns. While these authors suggested evidence for unconditional symmetry, we find evidence for asymmetry even after the data is conditioned on their lags. Among the non-financial time series considered, investment, the consumption of durables, and manufacturing employment reject conditional symmetry. This finding is interesting because the dynamics of these series are often believed to be affected by fixed costs of adjustments.

It is useful to put into perspective these results for conditional symmetry vis-á-via the evidence for cyclical asymmetry in the macroeconomic literature. Using the coefficient of skewness as the test statistic, Delong and Summers (1985) find little evidence for asymmetry in postwar U.S. output growth and industrial production, but find some evidence of asymmetry in the unemployment rate series. Because of data dependence, Delong and Summers calculate the critical value for their skewness coefficient by Monte Carlo simulations for AR(2) models. Since we accept conditional symmetry for the growth rate of GDP and industrial production, for the specified conditional mean, this also implies unconditional symmetry. The reason is that a linear ARMA model for X_t can also be written as: $X_t = \sum_{i=0}^{\infty} a_i e_{t-i}$. It follows that if the e_i are symmetric, X_t will also be symmetric (if e_t and $-e_t$ have the same distribution then X_t and $-X_t$ also have the same distribution). Hence our evidence of conditional symmetry is in fact consistent with Delong and Summers' evidence for unconditional symmetry in the two series. We note that, however, if each e_t is asymmetric, it does not necessarily imply X_t is asymmetric.

5 Conclusion

In this paper, we propose a test for conditional symmetry in dynamic models. Unlike other tests that exist in the literature, our test is valid whether or not the data are i.i.d. and is suited for time series applications. We highlight a number of econometric applications where the assumption of conditional symmetry is invoked. The proposed test is asymptotically distribution free and simulations show that it has good finite size and power properties. Applying the test to macroeconomic data, we reject conditional symmetry in some financial time series which have previously rejected unconditional symmetry. We find evidence of conditional asymmetry in macroeconomic variables whose dynamics are thought to be affected by fixed cost of adjustments.

Table 1: Size of the Tests: (H_0: Symmetry Around the Mean)
(asymptotic nominal size=0.05)

	T=50			T=100			T=200			α_3	α_4
	CS	CS^-	CS^+	CS	CS^-	CS^+	CS	CS^-	CS^+		
Normal	.037	.011	.031	.051	.023	.039	.049	.026	.040	0	3.0
t_5	.067	.015	.057	.081	.023	.069	.071	.025	.056	0	9.0
S3	.042	.013	.035	.042	.024	.031	.045	.024	.041	0	
S4	.044	.028	.025	.047	.024	.035	.044	.022	.038	0	3.0
S5	.078	.021	.065	.087	.028	.069	.075	.022	.066	0	6.0
S6	.106	.028	.088	.110	.034	.085	.091	.033	.050	0	11.6
S7	.134	.036	.110	.140	.029	.124	.117	.045	.092	0	126.0

Table 2: Power of the Tests (based on 5% asymptotic critical values)

	T=50			T=100			T=200			α_3	α_4
	CS	CS^-	CS^+	CS	CS^-	CS^+	CS	CS^-	CS^+		
Lognormal	.977	.942	.832	1.000	1.000	.996	1.000	1.000	1.000		
χ^2_2	.882	.803	.542	.995	.991	.775	1.000	1.000	.981		
Exponential	.878	.795	.541	.997	.993	.759	1.000	1.000	.982		
A4	.566	.487	.314	.850	.815	.669	.9982	.958	.998	.5	2.2
A5	.418	.315	.205	.697	.647	.262	.972	.961	.631	1.5	7.5
A6	.307	.177	.180	.416	.350	.146	.647	.626	.140	2.0	21.2
A7	.932	.870	.664	.999	.998	.870	1.000	1.000	.997	3.16	23.8
A8	.961	.915	.729	1.000	1.000	.929	1.000	1.000	.999	3.8	40.7

Table 3: Size and Power of the Test: Regression Model with $i.i.d.$ Regressors

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

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

	$k = 1$					
T/e_t	1	2	3	4	5	6
50	0.0530	0.0740	0.0420	0.9370	0.8800	0.9800
100	0.0460	0.0670	0.0380	0.9990	0.9980	1.0000
200	0.0420	0.0610	0.0440	1.0000	1.0000	1.0000
	$k = 4$					
50	0.0520	0.0510	0.0300	0.8130	0.7160	0.7830
100	0.0460	0.0620	0.0420	0.9990	0.9840	0.9940
200	0.0420	0.0810	0.0370	1.0000	1.0000	1.0000

Table 4: Size and Power of the Test: $AR(1)$

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

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

	$\rho = .5$					
T/e_t	1	2	3	4	5	6
50	0.0440	0.0610	0.0320	0.9380	0.8620	0.9080
100	0.0460	0.0870	0.0460	1.0000	0.9980	1.0000
200	0.0470	0.0710	0.0370	1.0000	1.0000	1.0000
	$\rho = .8$					
50	0.0470	0.0730	0.0350	0.9410	0.8500	0.8990
100	0.0470	0.0750	0.0470	1.0000	0.9960	0.9960
200	0.0530	0.0660	0.0420	1.0000	1.0000	1.0000

Table 5: Size and Power of the Test: $MA(1)$ Regressor

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

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

	$\rho = .5$					
T/e_t	1	2	3	4	5	6
50	0.0450	0.0600	0.0470	0.8330	0.7800	0.8160
100	0.0500	0.0840	0.0430	0.9910	0.9870	0.9950
200	0.0410	0.0650	0.0420	1.0000	1.0000	1.0000
	$\rho = .8$					
50	0.0510	0.0620	0.0370	0.7930	0.7400	0.7580
100	0.0390	0.0900	0.0490	0.9950	0.9800	0.9910
200	0.0450	0.0630	0.0360	1.0000	1.0000	1.0000

Table 6: Size and Power of the Test: $GARCH(1,1)$ Regressor

DGP 4: $y_t = 1 + u_t$, $u_t = e_t\sqrt{h_t}$, $h_t = \phi_0 + \phi_1 h_{t-1} + \phi_2 u_{t-1}^2$

Regression: GARCH(1,1) with Gaussian Likelihood

	$\phi = (2.0, .5, .3)$					
T/e_t	1	2	3	4	5	6
50	0.0350	0.0720	0.0360	0.9370	0.8840	0.9050
100	0.0470	0.0800	0.0440	0.9950	0.9890	0.9940
200	0.0460	0.0690	0.0480	1.0000	1.0000	1.0000
	$\phi = (2.0, .5, .45)$					
50	0.0530	0.0710	0.0370	0.9230	0.8650	0.8880
100	0.0510	0.0640	0.0450	0.9920	0.9940	0.9840
200	0.0400	0.0810	0.0410	1.0000	1.0000	1.0000

Table 7: Application to Macroeconomic Data

Sample	Series	CS	CS^-	CS^+
1971:1-1997:12	Canada-U.S. Ex. Rate	.8846	.8836	.7765
1971:1-1997:12	German-U.S. Ex. Rate	.7578	.7578	.6380
1971:1-1997-12	Japan-U.S. Ex. Rate	2.9671	2.5556	2.9671
1948:1-1997:12	Unemployment Rate	.9462	.7781	.9462
1946:1-1997:12	Ind. Prod.	1.4432	.9341	1.4432
1959:1-1997:4	Inflation (GDP)	1.0002	1.0002	.7386
1959:1-1997:4	GDP	.7561	.7561	.6710
1947:1-1997:12	Inflation (CPI)	2.2834	2.2834	1.6068
1981:10:30-1996:05:10:	30 day Int. Rate	1.3312	1.1996	1.3312
1980:11:03-1998:01:19	M2	1.1083	1.0005	1.1083
1959:3-1996:4	Con. Durables	2.6407	2.1146	2.6407
1959:3-1996:4	Con. Non-Durables	1.1172	.8062	1.1172
1946:1-1996:11	Employment	1.5040	1.0002	1.5040
1946:1-1997:12	Manu. Employment	2.2560	1.0001	2.2560
1959;3-1997:4	Final Sales	.9885	.9536	.9985
1959:3-1997:4	Non-Resid. Invest	2.1900	1.4794	2.1900
1959:3-1997:4	Resid. Invest	.9056	.9056	.6873
1990:01:02-1996:12:31	Stock Returns	3.9237	3.9237	3.4939

The critical values are 2.78, 2.20 and 1.91 at the 1, 5, and 10 percent levels respectively.

Appendix A: Proofs

Proof of Lemma 1: First we derive the variance and covariance function for the process $W_T(x)$. For $x, y \leq 0$, it is straightforward to show that $EW_T(x)W_T(y) = E\{[I(e_t \leq x) - I(-e_t \leq x)][I(e_t \leq y) - I(-e_t \leq y)]\} = 2F(x \wedge y)$, 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 $x \geq 0, y \geq 0$, $EW_T(x)W_T(y) = E\{[I(e_t \leq x) - I(-e_t \leq x)][I(e_t \leq y) - I(-e_t \leq y)]\} = 2[1 - F(x \vee y)]$, where $x \vee y = \max\{x, y\}$. A time-scaled 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 $x \geq 0$). □

To prove Theorem 1, we need a number of lemmas.

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

$$J(r) = B(r) - \int_0^r [g(s)(\int_s^1 g(v)^2 dv)^{-1} \int_s^1 g(v)dB(v)]ds$$

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 $EJ(r)J(s) = r \wedge s$. □.

Lemma 3 *Let $B(r)$ be a standard Brownian motion on [0,1] and let g be a function on [0,1] such that $\int_0^s g(v)^2 dv > 0$ for every $s \in (0, 1]$. Then*

$$J(r) = B(r) - B(1) + \int_r^1 [g(s)(\int_0^s g(v)^2 dv)^{-1} \int_0^s g(v)dB(v)]ds$$

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 $EJ(s)J(r) = s \wedge r$. □

Lemma 4 *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 $\int_{-\infty}^y g(v)^2 h(v)dv > 0$ for every $y \leq 0$. Define $W(x) = B(2H(x))$ (for $x \leq 0$). Then the process J^- defined as

$$J^-(x) = W(x) - W(0) + \int_x^0 [g(y)h(y)(\int_{-\infty}^y g(v)^2 h(v)dv)^{-1} \int_{-\infty}^y g(v)dW(v)]dy$$

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* $\int_y^\infty g(v)^2 h(v)dv > 0$ *for every* $y \geq 0$. *Define* $W(x) = B(2[1 - H(x)])$ *(for* $x \geq 0$*). Then the process* J^+ *defined as*

$$J^+(x) = W(x) - W(0) - \int_0^x [g(y)h(y)(\int_y^\infty g(v)^2 h(v)dv)^{-1} \int_y^{-\infty} g(v)dW(v)]dy$$

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

Remark: We can write $J^-(x) \stackrel{d}{=} B(1 - 2H(x))$, 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 $J^+(x) \stackrel{d}{=} B(2H(x) - 1)$.

Proof: Part (i) follows from a change in variable ($r = H(x)$) and Lemma 3. Part (ii) follows from a change in variable and Lemma 2. □

Lemma 5 *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* $x \leq 0$ *and* $W_T(x) \Rightarrow B(2[1 - 2H(x)])$ *for* $x \geq 0$. *Define* J_T^- *as in Lemma 4 part (i) but with* $W(\cdot)$ *replaced by* $W_T(\cdot)$ *in the transformation. Define* J_T^+ *as in Lemma 4 part (ii), but again replacing* $W(\cdot)$ *by* $W_T(\cdot)$. *Then*

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

and

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

Proof: This follows from the continuous mapping theorem and Lemma 4. Also see the Remark above. □

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

Lemma 6 *Let* W_T *satisfy the conditions of Lemma 5. Suppose that* g_T *and* h_T *are estimates of* g *and* h*, respectively, such that*

$$\int_{-\infty}^\infty (h_T - h)^2 dx = o_p(1) \quad \text{and} \quad \int_{-\infty}^\infty (g_T - g)^2 dH = o_p(1)$$

Define

$$\widetilde{J}_T^-(x) = W_T(x) - W_T(0) + \int_x^0 g_T(y)h_T(y)(\int_{-\infty}^y g_T(v)^2 h_T(v)dv)^{-1} \int_{-\infty}^y g_T(v)dW_T(v)dy$$

and

$$\widetilde{J}_T^+(x) = W_T(x) - W_T(0) - \int_0^x g_T(y)h_T(y)(\int_y^{\infty} g_T(v)^2 h_T(v)dv)^{-1}\int_y^{-\infty} g_T(v)dW_T(v)dy.$$

Then

$$\widetilde{J}_T^-(x) = J_T^-(x) + o_p(1), \quad \text{and} \quad \widetilde{J}_T^+(x) = J_T^+(x) + o_p(1) \tag{8}$$

where $o_p(1)$ is uniform over x, and J_T^- and J_T^+ are defined in Lemma 5. Therefore,

$$\widetilde{J}_T^-(x) \Rightarrow B(1-2H(x)), \quad \text{and} \quad \widetilde{J}_T^+(x) \Rightarrow B(2H(x)-1) \tag{9}$$

Proof: Equation (8) is implied by the result of Bai (1997, Theorem 2). Equation (9) follows from equation (8) and Lemma 5. □

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

We now state the assumptions under which Theorem 1 will be proved.

Assumption A1: e_t are iid with cdf $F(x)$ and density $f(x)$. The cdf $f(x)$ is continuously differentiable and $|xf(x)| < M < \infty$ for some $M > 0$.

Assumption A2: $\frac{1}{T}\sum_{t=1}^T \|\frac{\partial h_t}{\partial \beta}\| = O_p(1)$.

Assumption A3: $\max_{1\le t\le T} T^{-1/2}\|\frac{\partial h_t}{\partial \beta}\| = o_p(1)$.

Assumption A4: For every $\epsilon > 0$, there exists $\delta > 0$ such that

$$P\left(\sup_{\|u-v\|\le\delta} T^{-1/2}\sum_{t=1}^T \left\|h_t(\beta_0 + T^{-1/2}u) - h_t(\beta_0 + T^{-1/2}v)\right\| > \epsilon\right) < \epsilon$$

Assumption A5: The estimators satisfy $\sqrt{T}(\hat\beta - \beta_0) = O_p(1)$, and $\sqrt{T}(\hat\lambda - \lambda_0) = O_p(1)$.

Assumption A6: The effect of information truncation is small:

$$T^{-1/2}\sum_{t=1}^T |h(\widetilde{\Omega}_t, \beta) - h(\Omega_t, \beta)| = o_p(1).$$

Proof of Theorem 1. Note that $\hat W_T(x) = \hat U_T^+(x) - \hat U_T^-(x)$ where $\hat U_T^+(x) = T^{-1/2}\sum_{t=1}^T[I(\hat e_t \le x) - F(x)]$ and $\hat U_T^-(x) = T^{-1/2}\sum_{t=1}^T[I(-\hat e_t \le x) - F(x)]$. Under A1-A6, from Bai (1997, Theorems 4 and 5, also see Bai 1996, Theorem A.2), we have

$$\hat U_T^+(x) = U_T^+(x) + f(x)\xi_{1T} + xf(x)\xi_{2T} + o_p(1) \tag{10}$$

where

$$\xi_{1T} = \frac{1}{T}\sum_{t=1}^T \frac{\partial h_t}{\partial \beta}(\beta_0)'\sqrt{T}(\hat\beta - \beta_0)/\sigma_t \tag{11}$$

and

$$\xi_{2T} = \frac{1}{2T}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\Big[\sqrt{T}(\hat{\alpha}-\alpha)\sum_{j=1}^{t}\hat{\delta}^j + \sqrt{T}(\hat{\sigma}_0^2-\sigma_0^2)\hat{\delta}^{t-1}$$

$$+\sqrt{T}(\hat{\delta}-\delta)\sum_{j=1}^{t-1}\hat{\delta}^j\sigma_{t-j}^2 + \sqrt{T}(\hat{\gamma}-\gamma)\sum_{j=1}^{t-1}\hat{\delta}^j(Y_{t-j}-h(\Omega_{t-j},\hat{\beta}))^2\Big].$$

From $I(-\hat{e}_t \le x) - F(x) = I(\hat{e}_t \ge -x) - F(x)$, 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 (10) that

$$\hat{U}_T^-(x) = U_T^-(x) - f(-x)\xi_{1T} - (-x)f(-x)\xi_{2T} + o_p(1). \tag{12}$$

Take the difference of (10), (12), and use $f(x) = f(-x)$ and $W_T = U_T^+ - U_T^-$, we obtain

$$\hat{W}_T(x) = W_T(x) + 2f(x)\xi_{1T} + o_p(1) \tag{13}$$

The above says that estimating parameters in the conditional variance does not affect the process $\hat{W}_T$. But estimating the parameters in the conditional mean does affect $\hat{W}_T(x)$. Next we consider transforming $\hat{W}_T(x) - \hat{W}_T(0)$ for $x \le 0$. Assumption (4) implies that $f_T(x) - f(x) = o_p(1)$ uniformly in x. Thus we can rewrite (13) as

$$\hat{W}_T(x) = W_T(x) + 2f_T(x)\xi_{1T} + o_p(1), \tag{14}$$

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

$$\hat{W}_T(x) - \hat{W}_T(0) = W_T(x) - W_T(0) + 2[f_T(x) - f_T(0)]\xi_{1T} + o_p(1). \tag{15}$$

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

$$\phi_T(\eta)(x) = \int_x^0 g_T(y)f_T(y)(\int_{-\infty}^y g_T(v)^2 f_T(v)dv)^{-1}\int_{-\infty}^y g_T(v)d\eta(v)dy \tag{16}$$

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, $\phi_T(f_T)(x) = \int_x^0 \dot{f}_T(y)dy = f_T(0) - f_T(x)$. Note that $S_T(x)$ in equation (5) 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 (14), we have

$$\phi_T(\hat{W}_T) = \phi_T(W_T) + \phi_T(f_T)2\xi_{1T} + o_p(1) = \phi_T(W_T) + [f_T(0) - f_T(x)]2\xi_{1T} + o_p(1). \tag{17}$$

Thus, for $x < 0$,

$$\begin{aligned} S_T(x) &= \hat{W}_T(x) - \hat{W}_T(0) + \phi_T(\hat{W}_T)(x) \\ &= W_T(x) - W_T(0) + 2[f_T(x) - f_T(0)]\xi_{1T} + o_p(1) \quad \textit{by } (15) \\ &\quad +\phi_T(W_T)(x) + [f_T(0) - f_T(x)]2\xi_{1T} + o_p(1) \quad \textit{by } (17) \\ &= W_T(x) - W_T(0) - \phi_T(W_T)(x) + o_p(1) \\ &= \widetilde{J}_T^-(x) + o_p(1) \quad \text{replacing } h_T \textit{ by } f_T \text{ in Lemma 6} \\ &\Rightarrow B(1 - 2F(x)) \quad \text{by Lemma 6} \end{aligned}$$

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 1. □

Proof of Theorem 2. Let $K_T(x) = (1 - \delta/\sqrt{T})F(x) + \delta/\sqrt{T}H(x)$. By (7), $e_{Tt} \sim K_T(x)$. It is easy to show that $-e_t \sim G_T(x)$, where $G_T(x) = K_T(x) + \frac{\delta}{\sqrt{T}}[1 - H(x) - H(-x)]$. Define:

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

and

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

Again, using the results of Bai (1997), or Bai (1996), we have

$$\hat{Z}_T^+(x) = Z_T^+(x) + f(x)\xi_{1T} + xf(x)\xi_{2T} + o_p(1), \tag{18}$$

where ξ_{1T} and ξ_{2T} are defined earlier. Equation (18) is similar to (12). 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 $dK_T(x)/dx = f(x) + O(T^{-1/2})$. This explains the presence of $f(x)$ in (18). We next consider the asymptotic representation for $\hat{Z}_T^-(x)$. Notice that $I(-\hat{e}_{Tt} \le x) - G_T(x) = 1 - I(\hat{e}_{Tt} < -x) - G_T(x) = -[I(\hat{e}_{Tt} < -x) - K_T(-x)]$ because $1 - G_T(x) = K_T(-x)$. Thus $\hat{Z}_T^-(x) = -\hat{Z}_T^+(-x)$ (a.s.), and hence from (18) (replacing x by $-x$),

$$\hat{Z}_T^-(x) = Z_T^-(x) - f(-x)\xi_{1T} - (-x)f(-x)\xi_{2T} + o_p(1). \tag{19}$$

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

$$\hat{W}_T(x) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[I(\hat{e}_{Tt} \le x) - I(-\hat{e}_{Tt} \le x)]$$

$$
\begin{aligned}
&= \hat{Z}_T^+(x) - \hat{Z}_T^-(x) + T^{1/2}[K_T(x) - G_T(x)] \\
&= Z_T^+(x) - Z_T^-(x) + 2f(x)\xi_{1T} \qquad \text{from (18) and (19)} \\
&\quad +\delta[H(x) + H(-x) - 1] + o_p(1).
\end{aligned}
$$

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 $x \le 0$ 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 $x, y \le 0$, $EW_T(x)W_T(y) = K_T(x \wedge y) - K_T(x)K_T(y) + G_T(x \wedge y) - G_T(x)G_T(y) + G_T(x)K_T(y) + K_T(x)G_T(y) \to 2F(x \wedge y)$ 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,

$$
\hat{W}_T(x) = W_T(x) + 2f(x)\xi_{1T} + \delta[H(x) + H(-x) - 1] + o_p(1), \tag{20}
$$

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 $\hat{e}_{Tt} = \epsilon_t + O_p(T^{-1/2})$, where $\epsilon_t \sim F(x)$. To see this, from $e_{Tt} \sim (1 - \delta T^{-1/2})F(x) + (\delta T^{-1/2})H(x)$, we can write $e_{Tt} = \epsilon_t + \eta_{Tt}$, where $\eta_{Tt} = 0$ with probability $1 - \frac{\delta}{\sqrt{T}}$ and $\eta_{Tt} = a_t - \epsilon_t$ with probability $\frac{\delta}{\sqrt{T}}$, here ϵ_t and a_t are independent such that $\epsilon_t \sim F(x)$ and $a_t \sim H(x)$. Hence, $\eta_{Tt} = O_p(T^{-1/2})$. In addition, the estimated residuals satisfy $\hat{e}_{Tt} = e_{Tt} + O_p(\frac{1}{\sqrt{T}})$ and thus $\hat{e}_{Tt} = \epsilon_t + O_P(\frac{1}{\sqrt{T}})$. Let f_T and g_T are estimates of f and g. Define the mapping ϕ_T as in (16). Then using the same argument as in the proof of Theorem 1, 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 5, $W_T(x) - W_T(0) + \phi_T(W_T)(x) \Rightarrow B(1 - 2F(x))$. In addition, $\phi_T(v)(x) \to \phi^-(x)$, which is defined in Theorem 2. Thus,

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

obtaining the result for $x < 0$. The case of $x > 0$ is similar. The proof of Theorem 2 is complete. □

Appendix B: Computation of the Statistic

From the definition of the test statistics, the feasible CS_T tests require estimation of the components of $h_T^{\pm}$ [see equations (5) and (6)]. Consider first the terms $\int_{-\infty}^{y} g_T(z) d\hat{W}_T(z)$ and $\int_{y}^{\infty} g_T(z) d\hat{W}_T(z)$. Note that these can equivalently be represented as:

$$\begin{aligned} \int_{-\infty}^{y} g_T d\hat{W}_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [g_T(\hat{e}_t) I(\hat{e}_t \leq y) - g_T(-\hat{e}_t) I(-\hat{e}_t \leq y)]; \\ \text{and} \qquad \int_{y}^{\infty} g_T d\hat{W}_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [g_T(\hat{e}_t)(\hat{e}_t \geq y) - g_T(-\hat{e}_t) I(-\hat{e}_t \geq y)]. \end{aligned}$$

The remaining components of $h_T^{\pm}$ can be obtained as follows. First, consistent estimates of the density and its derivative are obtained non-parametrically. 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. Second, the integration (over z) of $g_T(z)^2 f_T(z)$ is approximated by summations. This makes the computation straightforward. Simulations show that the size and power of the tests are not affected by these approximations. Gauss and Splus programs are available from the authors on request.

References

Ahmad, I.A., and Qi Li (1996) Testing Symmetry of an Unknown Density Function by Kernel Method, manuscript, Department of Economics, University of Guelph.

Beaudry, P. and Koop.,G. (1993). Can Recessions Permanently Change Output? *Journal of Monetary Economics*, 31, 149-163.

Bai, J. (1996). Testing for Parametric Constancy in Linear Regressions: An Empirical Distribution Function Approach. *Econometrica* 64, 597-622.

Bai, J. (1997) Testing Parametric Conditional Distributions of Dynamic Models. Manuscript, Department of Economics, MIT.

Bickel, P. J. (1982), On Adaptive Estimation, *Annals of Statistics* **10**, 647–671.

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31**, 307–27.

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. and Newey, W. K. (1998), Efficient Bootstraps for Semiparametric Models, mimeo, M.I.T.

Campbell, J. Y. and Hentschel, L. (1992), No News is Good News, *Journal of Financial Economics* **31**, 281–318.

Delong, J. B. and Summers, L. H. (1982), Are Business Cycle Symmetrical, in *American Business Cycle: Continuity and Change*, Gordon, R. J. eds., University of Chicago Press, Chicago.

Diebold, F. X. (1988), *Empirical Modeling of Exchange Rate Dynamics*, Springer, New York.

Engle, R. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of Variance of U.K. Inflation, *Econometrica* **50**, 987–1007.

Fan, Y. and 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. and Stambaugh, R. (1987), Expected Stock Returns and Volatility, *Journal of Financial Economics* **19**, 3–29.

Gonzalez-Rivera, G. (1997), A note on adaptation in GARCH models, *Econometric Reviews* 16, 55-68.

Granger, C. W. and Newbold, P. (1974), Spurious Regressions in Econometrics, *Journal of Econometrics* **2**, 111–120.

Hamilton, J. D. (1989), A New Approach to the Econometric Analysis of Non-stationary Time Series and the Business Cycle. *Econometrica* 57, 357-384.

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.

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. and Hansen, B. E. (1991), 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.

Neftci, S. N. (1986), Are Economic Time Series Asymmetric over the Business Cycle? *Journal of Political Economy*, **92**, 307-328.

Newey, W. K. (1988), Adaptive Estimation of Regression Models via Moment Restrictions, *Journal of Economics* **38**, 301–339.

Newey, W. K. and Powell, J. L. (1988), Asymmetric Least Squares Estimation and Testing, *Econometrica* **55**, 819–847.

Newey, W. K. and Steigerwald, D. G. (1997), Asymptotic Bias for Quasi-Maximum Likelihood Estimators in Conditional Heteroskedastic Models, *Econometrica* **65**, 587–599.

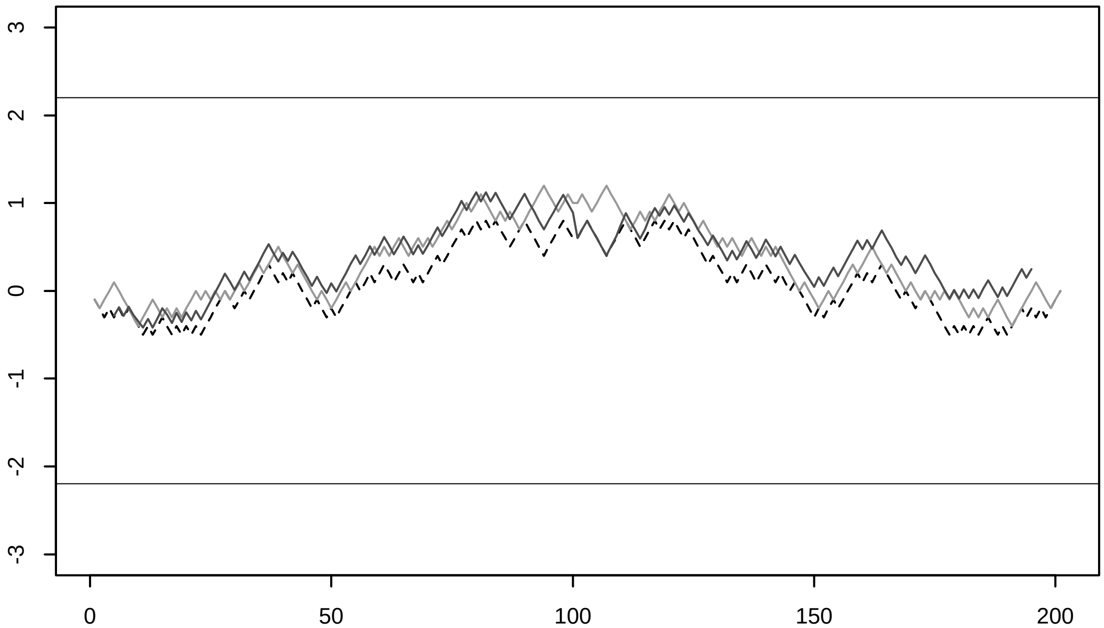
Randles, R. H., Fligner, M. A., Policello, G. E. and Wolfe, D. A. (1980), An Asymptotically Distribution-Free Test for Symmetry Versus Asymmetry, *Journal of the American Statistical Association* **75**, 168–172.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

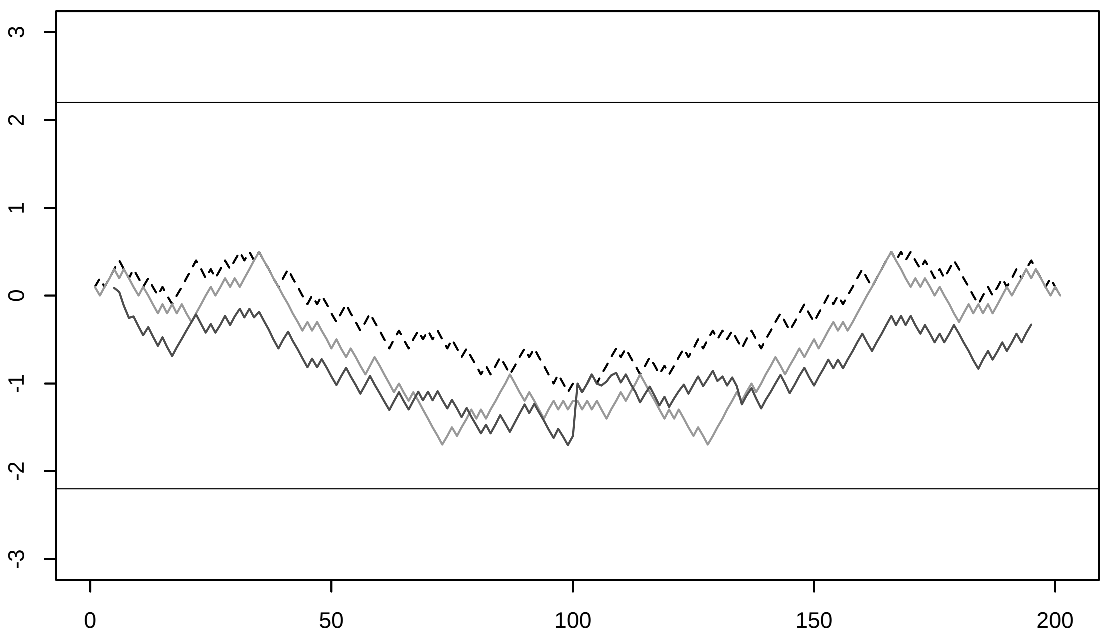
Stoker, T. M. (1986), Consistent Estimation of Scaled Coefficients, *Econometrica* **54:6**, 1461–1482.

Zheng, J. X. (1998), Consistent Specification Testing for Conditional Symmetry, *Econometric Theory* **14**, 139–149.

Figure 1: Test of Symmetry for Simulated Normal Observations. Dashed line: $\hat{W}_T(x)$. Solid line: $S_T(x)$. Light-solid line: $W_T(x)$. Two horizontal lines: the 95% confidence band.

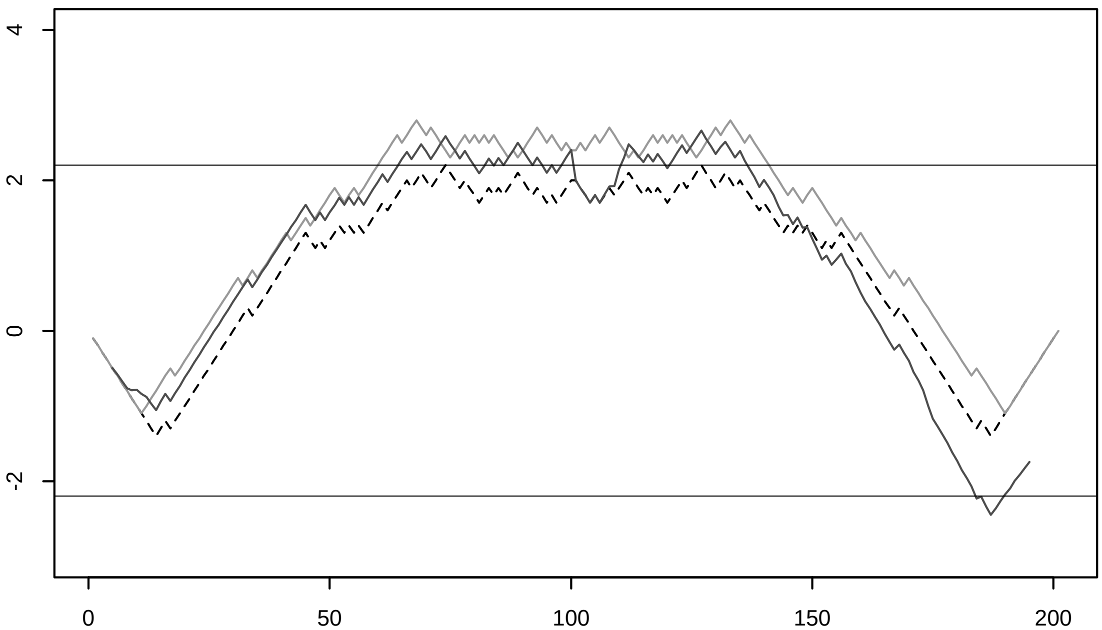


(a)

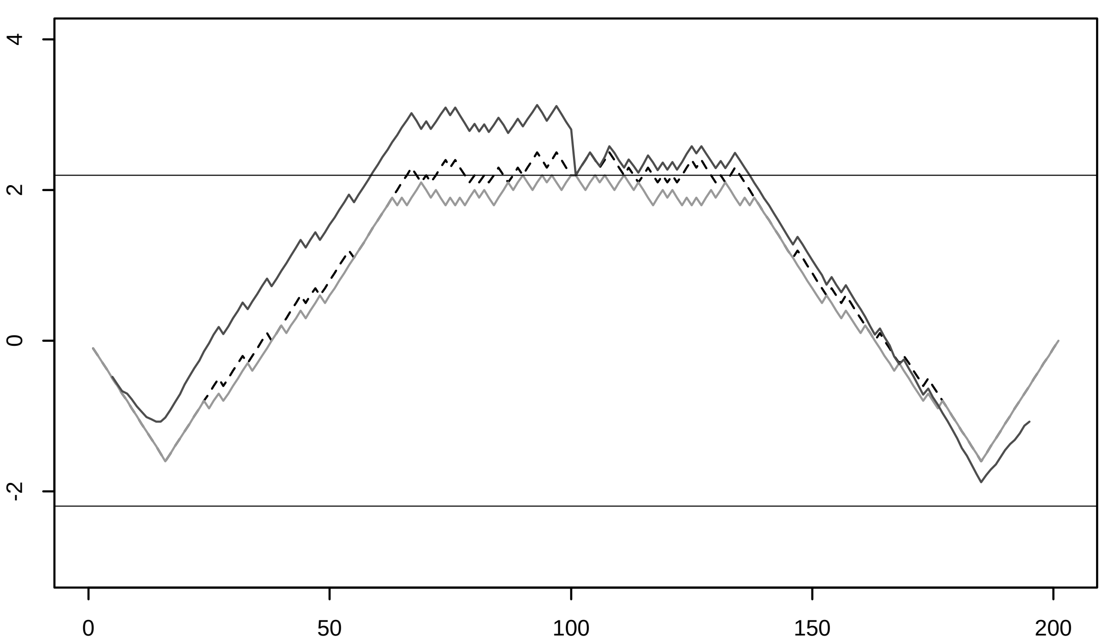


(b)

Figure 2: Test of Symmetry for Simulated Chi-Square Observations. Dashed line: $\hat{W}_T(x)$. Solid line: $S_T(x)$. Light-solid line: $W_T(x)$. Two horizontal lines: the 95% confidence band.



(a)



(b)