SIGN INVARIANCE IN GOODNESS-OF-FIT TESTS FOR TIME SERIES

By T. W. Anderson and M. A. Stephens

Stanford University and Simon Frazer University

First version received July 1997

Abstract. A goodness-of-fit test for a stationary stochastic process may be based on a functional of the difference between the sample standardized spectral distribution and a hypothesized standardized spectral distribution. Theorems are given to show that under certain conditions the distribution of such a functional based on observations from a process $\{y_t\}$ indexed by a parameter θ is the same for $\theta = \theta_0$ and for $\theta = -\theta_0$. The results are illustrated by three examples of time series processes.

Keywords. Stationary processes; standardized spectral distributions; Kolmogorov–Smirnov statistic; Cramér–von Mises statistic.

1. INTRODUCTION

Goodness-of-fit statistics for testing time series models, e.g. the AR(1) and MA(1) models, can be based on a comparison of the standardized spectral distribution of the process and the observed empirical standardized spectral distribution. The standardized spectral distribution is the spectral distribution divided by the variance of the process; see Anderson (1993) for more details. Tests of the above type include those based on the Kolmogorov–Smirnov and Cramér–von Mises statistics. In some important cases, the distribution of a given statistic does not depend on the sign of a parameter or a vector of parameters; we use the term 'sign invariance' to describe this property.

Anderson and Stephens (1993) considered the Cramér–von Mises statistic for testing the AR(1) model defined by

$$y_t = \phi y_{t-1} + u_t \tag{1}$$

with known value of the parameter ϕ, and gave tables of the asymptotic distributions of the test statistic. The tables were given for various positive values of the first-order correlation ρ, which for the model above equals ϕ. The authors gave similar tables for testing the MA(1) model defined by

$$y_t = u_t + \alpha u_{t-1} \tag{2}$$

with known value of the parameter α. The tables were again given for positive values of the first-order correlation ρ, which for this model equals $\alpha/(1+\alpha^2)$. In both models, the u_t were assumed to be normally distributed with zero mean.

Although this was not explicitly stated, the tables can be used also for negative values of ρ by entering the table with $|\rho|$; i.e., the statistic is sign-invariant.

In this paper we prove a general theorem which verifies this property of sign invariance and, more importantly, extends it to the case where the parameter is estimated. The theorem applies as well to finite samples, and to other statistics for testing fit.

Before giving the theorem, we need the following preliminaries. Consider a stochastic process $\{y_t\}$ with mean zero and covariances $\sigma_y(h) = Ey_t y_{t+h}$. The variance is then $\sigma_y(0)$ and the autocorrelation function is $\rho_y(h) = \sigma_y(h)/\sigma_y(0)$. The standardized spectral density is

$$f_y(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho_y(h) \cos(\lambda h). \tag{3}$$

Note that

$$\rho_y(h) = \int_{-\pi}^{\pi} \cos(\lambda h) f_y(\lambda)\, d\lambda. \tag{4}$$

The standardized spectral distribution is

$$F_y(\lambda) = 2\int_0^{\lambda} f_y(\nu)\, d\nu = \frac{1}{\pi}\left\{\lambda + 2\sum_{h=1}^{\infty} \rho_y(h) \frac{\sin(\lambda h)}{h}\right\}. \tag{5}$$

Suppose $y = (y_1, \ldots, y_T)$ is a vector whose components are the values in a sample of size T from $\{y_t\}$. The hth autocorrelation of the sample is $r_y(h) = \sum_{t=1}^{T-h} y_t y_{t+h} / \sum_{t=1}^{T} y_t^2$.

From the sample, the standardized sample spectral density (periodogram) is then

$$I_y(\lambda) = \frac{1}{2\pi} \sum_{h=1}^{T-1} r_y(h) \cos(\lambda h), \tag{6}$$

and the standardized sample spectral distribution is

$$\hat{F}_y(\lambda) = 2\int_0^{\lambda} I_y(\nu)\, d\nu. \tag{7}$$

In the AR(1) model defined by (1), replacing ϕ by $-\phi$ yields a process $\{x_t\}$ with autocovariance function

$$\rho_x(h) = (-1)^h \rho_y(h), \qquad h = 0, 1, \ldots \tag{8}$$

The standardized spectral density of this process is

$$
\begin{aligned}
f_x(\lambda) &= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}(-1)\rho_y(h)\cos(\lambda h) \\
&= \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\rho_y(h)\cos\{(\pi-\lambda)h\} \\
&= f_y(\pi-\lambda) \qquad (9)
\end{aligned}
$$

and

$$
\begin{aligned}
F_x(\lambda) &= 2\int_0^{\lambda} f_x(\nu)\,d\nu = 2\int_0^{\lambda} f_y(\pi-\lambda)\,d\lambda \\
&= 2\int_{\pi-\lambda}^{\pi} f_y(\mu)\,d\mu \\
&= 1 - F_y(\pi-\lambda). \qquad (10)
\end{aligned}
$$

We shall say that two processes related by (8), (9) and (10) are *covariant*.

2. SIGN INVARIANCE FOR A PAIR OF PROCESSES

In what follows, we discuss functionals $K[\cdot]$ with arguments of the form $\hat{F}_y(\lambda) - F_y(\lambda)$. Such a functional will be denoted by $Q(y)$. Note that the first term of the argument of K involves a sample from a process $\{y_t\}$ and the second involves the spectral distribution of the process itself. However, these two uses of y should not cause confusion. If $K[\cdot]$ is a functional such that $K[g] = K[h]$ when $g(\lambda)$ is related to $h(\lambda)$ by $h(\lambda) = -g(\pi-\lambda)$, $0 \leqslant \lambda \leqslant \pi$, we shall say that $K[\cdot]$ is a *symmetric* functional.

We now state the following theorem.

THEOREM 1. *Let $\{y_t\}$ and $\{x_t\}$ be two covariant stationary Gaussian processes with means zero and with autocorrelations $\rho_y(h)$ and $\rho_x(h)$ respectively. Let $y = (y_1, \ldots, y_T)$ be a sample of size T from $\{y_t\}$, and $x = (x_1, \ldots, x_T)$ be a sample of size T from $\{x_t\}$. Suppose $K[g]$ is a symmetric functional. Then the functional*

$$Q(y) = K[\hat{F}_y(\cdot) - F_y(\cdot)] \qquad (11)$$

has the same distribution as $Q(x) = K[\hat{F}_x(\cdot) - F_x(\cdot)]$.

PROOF. We first prove the theorem for the processes $\{y_t\}$ and $\{z_t\}$, where $\{z_t\}$ is a process defined by $z_t = (-1)^t y_t$. This process has mean zero and the autocorrelations are $\rho_z(h) = (-1)^h\rho_y(h)$; the standardized spectral density of the z_t process is $f_z(\lambda) = f_y(\pi - \lambda)$ and the standardized spectral distribution is $F_z(\lambda) = 1 - F_y(\pi - \lambda)$.

A sample from $\{z_t\}$ may be constructed from the y sample by calculating $z_t = (-1)^t y_t, \quad t = 1, \ldots, T$. The hth autocorrelation of the sample is $r_z(h) = (-1)^h r_y(h)$; for the z sample we have

$$I_z(\lambda) = \frac{1}{2\pi} \sum_{h=1}^{T-1} r_z(h) \cos(\lambda h) = I_y(\pi - \nu) \tag{12}$$

and

$$\hat{F}_z(\lambda) = 2\int_0^{\lambda} I_z(\nu)\, d\nu = 1 - \hat{F}_y(\pi - \lambda). \tag{13}$$

It follows that

$$\hat{F}_z(\lambda) - F_z(\lambda) = -\{\hat{F}_y(\pi - \lambda) - F_y(\pi - \lambda)\}. \tag{14}$$

The assumption that $K[\cdot]$ is symmetric implies

$$K[\hat{F}_z(\cdot) - F_z(\cdot)] = K[\hat{F}_y(\cdot) - F_y(\cdot)] \tag{15}$$

and therefore the functionals $Q(z)$ and $Q(y)$ have the same distribution.

For the rest of the proof we note that, since the process $\{y_t\}$ is Gaussian, its distribution is characterized by the variance and the autocorrelation sequence $\rho_y(h)$, or by the variance and the standardized spectral distribution (5). Since the autocorrelation of $\{x_t\}$ is the same as that of $\{z_t\}$, the two processes have the same distributional properties, and therefore the functionals $Q(x)$ and $Q(y)$ have the same distribution. ∎

3. A PROCESS INDEXED BY A PARAMETER

We now consider a process indexed by a parameter, which perhaps must be estimated, and functionals based on this process. Suppose the process $\{y_t(\theta)\}$, abbreviated $\{y_t\}$, is indexed by a parameter θ. The hth autocorrelation function will now be denoted by $\rho(h|\theta)$, and the standardized spectral density and standardized spectral distribution become $f(\lambda|\theta)$ and $F(\lambda|\theta)$. A sample from the process when the parameter is θ will be denoted by $y = (y_1, y_2, \ldots, y_T)$ as before. The functional $Q(y)$ will now be written

$$Q(y, \theta) = K[\hat{F}_y(\cdot) - F_y(\cdot|\theta)]. \tag{16}$$

Of particular interest are functionals based on the process when the parameter takes the value θ, and those from the process when the parameter has value $-\theta$. We also need an estimator of θ with the following special property. Suppose $\hat{\theta}_y = \hat{\theta}(y_1, \ldots, y_T)$ is an estimator of θ based on a sample $(y_1, \ldots, y_T)$ from process $\{y_t\}$, and suppose $\hat{\theta}_y^*$ is the value of $\hat{\theta}(y_1, -y_2, \ldots, (-1)^T y_T)$, i.e., the value obtained when the estimator is evaluated from the sample changed by changing alternate signs. Then if we have the property that $\hat{\theta}_y^* = -\hat{\theta}_y$, we shall call the estimator $\hat{\theta}(y_1, \ldots, y_T)$ a *covariant* estimator.

THEOREM 2. *Let* $\{y_t(\theta)\}$ *be a family of stationary Gaussian processes such that for every pair of parameter values* θ, $-\theta \in \Theta$ *the processes are covariant. Let* $K[\cdot]$ *be a symmetric functional. Let* $\hat{\theta}_y$ *be the value obtained from a covariant estimator* $\hat{\theta}$, *and let*

$$Q(y, \hat{\theta}_y) = K[\hat{F}_y(\cdot) - F(\cdot|\hat{\theta}_y)]. \tag{17}$$

Then

(i) $Q(y, \theta)$ *has the same distribution as* $Q(x, -\theta)$ *when y and x are samples from covariant processes with parameter values* θ *and* $-\theta$;

(ii) $Q(y, \hat{\theta}_y)$ *has the same distribution as* $Q(x, \hat{\theta}_x)$ *when y and x are samples from covariant processes with parameter vales* θ *and* $-\theta$ *and* $\hat{\theta}_y$ *and* $\hat{\theta}_x$ *are covariant estimators.*

PROOF. Part (i) is essentially a restatement of Theorem 1. Part (ii) follows because the distribution of $Q(x, \hat{\theta}_x)$ is the same as that of $Q(z, \hat{\theta}_z)$ and

$$\begin{aligned} Q(z, \hat{\theta}_z) &= K[\hat{F}_z(\cdot) - F(\cdot|\hat{\theta}_z)] \\ &= K[-\hat{F}_y(\pi - \cdot) + F(\pi - \cdot|\hat{\theta}_y)]. \end{aligned} \tag{18}$$

■

Note that a corollary of the theorem is that the distribution of a covariant estimator $\hat{\theta}$, say $\hat{\theta}_1$, from a given population with parameter θ is the same as the distribution of $-\hat{\theta}_2$, where $\hat{\theta}_2$ is the estimator from the distribution with parameter $-\theta$.

4. EXAMPLES

EXAMPLE 1: THE AR(1) MODEL. As the first example consider the AR(1) model $\{y_t\}$ in (1). For this model $\theta = \phi$ and $\rho_h(h) = \phi^h$, $h = 0, 1, \ldots$. Let $\{x_t\}$ be defined by (1) with $-\phi$ replacing ϕ, so that $\rho_x(h) = (-\phi)^h = (-1)^h \rho_y(h)$. If $\{y_t\}$ and $\{x_t\}$ are Gaussian the conditions of Theorems 1 and 2 are satisfied. However, the conclusions of Theorem 1 can be extended to a model in which the distribution of u_t is symmetric but not necessarily normal. Then $v_t = (-1)^t u_t$ has the same distribution as u_t, and the distribution of $Q_x(\cdot)$ is the same as the distribution of $Q_z(\cdot)$ where $z_t = (-1)^t y_t$. The estimator of $\rho_y(h)$ is $h(y_1, \ldots, y_T) = r_y(h) = \sum_{t=1}^{T-h} y_t y_{t+h} / \sum_{t=1}^{T} y_t^2$. The estimator $\hat{\theta} = r_y(1)$, and $\hat{\theta}$ is covariant. The spectral density of $\{y_t\}$ is

$$f(\lambda|\phi) = \frac{1-\phi^2}{2\pi(1+\phi^2-2\phi\cos\lambda)}. \tag{19}$$

Goodness-of-fit statistics to which Theorem 2 can now be applied are discussed in the next section.

EXAMPLE 2. Another example is the MA(1) process (2), with u_t as above. Here $\theta = \alpha$, the autocorrelations are $\rho_y(1) = \alpha/(1+\alpha^2)$, $\rho_y(h) = 0$, $h = 2, 3, \ldots,$ and the spectral density is

$$f(\lambda|\alpha) = \frac{1}{2\pi}(1 + 2\alpha\cos\lambda). \tag{20}$$

Here $f(\lambda|\alpha) = f(\pi-\lambda|-\alpha)$, i.e., the models are covariant. The autocorrelation $\rho_y(1)$ can be estimated by $r_y(1)$ and the equation $r_y(1) = \hat{\alpha}/(1+\hat{\alpha}^2)$ can be solved for $\hat{\theta} = \hat{\alpha}$ if $|r_y(1)| \leqslant 1/2$. The estimator of θ is covariant.

EXAMPLE 3: THE GENERAL ARMA PROCESS. The general ARMA process is defined by

$$\textstyle\sum_{j=0}^{p}\beta_j y_{t-j} = \sum_{g=0}^{q}\alpha_g u_{t-g}, \qquad \beta_0 = 1 = \alpha_0 \tag{21}$$

where, as before, the u_t are assumed to be independently normally distributed with zero mean. Let

$$\sum_{j=0}^{p}\beta_j x^{p-j} = \prod_{j=1}^{p}(x - x_j) \tag{22}$$

and

$$\sum_{g=0}^{q}\alpha_g w_{q-g} = \prod_{g=1}^{q}(w - w_g) \tag{23}$$

where $|x_j| < 1$, $j = 1, \ldots, p$, and $|w_g| \leqslant 1$, $g = 1, \ldots, q$. Define the parameter vector θ by

$$\theta = (x_1, \ldots, x_p, w_1, \ldots, w_q)'. \tag{24}$$

Then

$$f_y(\lambda|\theta) = C\frac{\prod_{g=1}^{q}|\{\exp(i\lambda) - w_g\}|^2}{\prod_{j=1}^{p}|\{\exp(i\lambda) - x_j\}|^2} \tag{25}$$

where C is a constant. Then $f_z(\lambda|-\theta) = f_y(\pi-\lambda|\theta)$ for $z_t = (-1)^t y_t$. Thus the general ARMA process is a special case of Theorem 2, when the parameter vector is defined as in (24).

5. APPLICATION TO GOODNESS-OF-FIT STATISTICS

A functional that is symmetric is the Kolmogorov–Smirnov goodness-of-fit statistic

$$K[\hat{F}_y(\lambda) - F_y(\lambda)] = \sup_{0 \leqslant \lambda \leqslant \pi} |\hat{F}_y(\lambda) - F_y(\lambda)|. \tag{26}$$

Another goodness-of-fit statistic that is symmetric is the Cramér–von Mises statistic

$$\frac{T}{2\pi G^2(\pi)} \int_0^{\pi} \{\hat{F}_y(\lambda) - F_y(\lambda)\}^2 f^2(\lambda)\, d\lambda$$
$$= \frac{T}{4\pi^4 G^2(\pi)} \sum_{r=1}^{\infty} \left\{ \sum_{g=1}^{T-1} \frac{(r_g - \rho_g)(\rho_{r+g} - \rho_{r-g})}{g} \right\}^2 \tag{27}$$

where $G(\pi) = 2\int_0^{\pi} f^2(\lambda)\, d\lambda$. Theorem 2 can be applied in the three examples above to give the results that both the Kolmogorov–Smirnov statistic and the Cramér–von Mises statistic have distributions, for any sample size, which are the same for the parameter θ as for $-\theta$. The Cramér–von Mises statistic has been discussed by Anderson (1993), and has been developed to provide tests for AR(1) and MA(1) processes by Anderson and Stephens (1993). Since the Cramér–von Mises statistic is sign-invariant, Anderson and Stephens gave tables only for positive values of ρ. When the parameters in these models must be estimated, the statistics have been discussed by Anderson *et al*. (1995).

A third statistic, used to test the null hypothesis $\rho_1 = \rho_2 = \cdots = \rho_k = 0$, is the Box–Pierce–Ljung statistic

$$T(T+2) \sum_{g=1}^{k} \frac{r_y^2(g)}{T-g} \tag{28}$$

which is invariant under the transformation $z_t = (-1)^t y_t$. To test the goodness-of-fit of a model it is proposed that the autocorrelations of residuals from a fitted model replace $r_y(g)$ in (28). For example, to test the AR(1) model, estimate ϕ by $\hat{\phi} = r_y(1)$ and calculate the residuals $\hat{u}_t = y_t - \hat{\phi} y_{t-1}$ and then the autocorrelations $r_{\hat{u}}(g)$ of the $\hat{u}_t$. These are then substituted into (28). The resulting statistic is sign-invariant.

It should be noted that in this paper we have assumed $Ey_t = 0$. If $Ey_t = \mu$ then we would define $z_t - \mu = (-1)^t (y_t - \mu)$, and the results would hold. However, if μ is a parameter to be estimated and the estimator is $\sum_{t=1}^{T} y_t / T$, then the estimator is not covariant. Thus Theorem 2 cannot be applied. For instance, in the AR(1) model

$$y_t = \mu + \phi(y_{t-1} - \mu) + u_t \tag{29}$$

when ϕ and μ are estimated, the finite-sample distribution of the Cramér–von

Mises statistic for $\phi = \phi_0$ is not the same as for $\phi = -\phi_0$ (Anderson *et al.*, 1995). However, the asymptotic distribution is sign-invariant (Anderson, 1997).

REFERENCES

ANDERSON, T. W. (1993) Goodness of fit tests for spectral distributions. *Ann. Stat.* 21, 830–47.

—— (1997) Goodness-of-fit tests for autoregressive processes. *J. Time. Ser. Anal*. 18, 321–39.

—— and STEPHENS, M. A. (1993) The modified Cramér–von Mises goodness-of-fit criterion for time series. *Sankhya* 55A, 357–69.

——, LOCKHART, R. A. and STEPHENS, M. A. (1995) Goodness-of-fit tests for the time series models AR(1) and MA(1). Research Report, Department of Mathematics and Statistics, Simon Fraser University.