

Power Computations for Intervention Analysis

A. I. McLEOD

Department of Statistical and Actuarial Sciences
University of Western Ontario
London, Ontario N6A 5B7, Canada
(aimcleod@uwo.ca)

E. R. VINGILIS

Department of Family Medicine
University of Western Ontario
London, Ontario N6G 4X8, Canada
(evingili@uwo.ca)

In many intervention analysis applications, time series data may be expensive or otherwise difficult to collect. In this case the power function is helpful, because it can be used to determine the probability that a proposed intervention analysis application will detect a meaningful change. Assuming that an underlying autoregressive integrated moving average (ARIMA) or fractional ARIMA model is known or can be estimated from the preintervention time series, the methodology for computing the required power function is developed for pulse, step, and ramp interventions with ARIMA and fractional ARIMA errors. Convenient formulas for computing the power function for important special cases are given. Illustrative applications in traffic safety and environmental impact assessment are discussed.

KEY WORDS: Autocorrelation and lack of statistical independence; Autoregressive integrated moving average time series models; Environmental impact assessment; Forecast and actuality significance test; Long-memory time series; Sample size; Two-sample problem.

1. INTRODUCTION

Intervention analysis developed by Box and Tiao (1976a) has been widely used in a variety of applications in engineering, biological, environmental, and social sciences to quantify the effect of a known intervention at time $t = T$ on data collected as a time series, $z_t, t = 1, \dots, n$. In its simplest form, intervention analysis itself may be regarded as a generalization of the two-sample problem to the case where the error or noise term is autocorrelated. It is well known that the usual two-sample procedures are not robust against alternatives involving autocorrelation (Box, Hunter, and Hunter 1978, sec. 3.1). The purpose of this article is to describe methods for computing the sample size necessary to detect an intervention with a prescribed power and level. It is shown by simulation experiments that these methods can be accurate even in moderately small samples. Statistical power computations have also been studied by Tiao et al. (1990) and Weatherhead et al. (1998) for particular types of intervention analysis models used for trend detection with environmental time series. The present article extends and refines these results.

It is assumed that for $t < T + b$, where b is the delay parameter, the time series is generated by a fractional autoregressive integrated moving average model, ARIMA(p, d, q) with fractional differencing parameter $|f| < .5$. Stationary short-memory time series models, $d = f = 0$, are used in environmental impact assessment (Box and Tiao 1976a; Tiao et al. 1990; Noakes and Campbell 1992; Weatherhead et al. 1998; Hipel and McLeod 1994, sec. 19.4.5) and in quality control (Jiang, Tsui, and Woodall 2000), as well as in many other areas of science and technology. Nonstationary models with $d = 1$ and/or long-memory models with $0 < f < .5$ have numerous applications in the physical and engineering sciences, such as quality control and industrial time series (Luceño 1995; Box and Luceño 1997), Internet traffic (Cao, Cleveland, Lin, and Sun 2001), daily solar irradiance (Kärner 2002), levels of Lake Huron (Roberts 1991, pp. 319–320), daily wind speed (Haslett and Raftery 1989), and various types of hydrologic time series (Beran 1994; Hipel and McLeod 1994).

In general, we may write the fractional ARIMA model for the preintervention series as

$$\nabla^{d+f} z_t = \xi + \theta(B)/\phi(B)a_t, \quad t = 1, \dots, T + b - 1, \quad (1)$$

where ξ is the constant term, d is the differencing parameter, $\nabla = 1 - B$, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, and B is the backshift operator on t . The innovations, denoted by $a_t, t = 1, \dots, n$, are assumed to be independent and normally distributed with mean 0 and variance σ_a^2 . It is also assumed that $\phi(B) = 0$ and $\theta(B) = 0$ have no common roots and that all roots are outside the unit circle.

2. SIMPLE INTERVENTION ANALYSIS MODEL

2.1 Introduction

The simple intervention analysis (SIA) model may be written as

$$\nabla^d z_t = \xi + \omega \nabla^d B^b I_t^{(T)} + \nabla^{-f} \frac{\theta(B)}{\phi(B)} a_t, \quad t = 1, \dots, n, \quad (2)$$

where $I_t^{(T)}$ is the intervention series, ω is the parameter indicating the magnitude of the intervention, and $\nabla^{-f} \theta(B)/\phi(B) a_t$ is the stationary error component. In this article three types of intervention series are used, the step, pulse, and ramp series, defined by

$$I_t^{(T)} = S_t^{(T)} = \begin{cases} 0 & \text{if } t < T \\ 1 & \text{if } t \geq T, \end{cases} \quad (3)$$

$$I_t^{(T)} = P_t^{(T)} = \begin{cases} 0 & \text{if } t \neq T \\ 1 & \text{if } t = T, \end{cases} \quad (4)$$

or

$$I_t^{(T)} = R_t^{(T)} = \begin{cases} 0 & \text{if } t < T \\ t - T + 1 & \text{if } t \geq T. \end{cases} \quad (5)$$

In practice, two of the most common models for the error are the autoregressive AR(1), and integrated moving average IMA(1), which correspond to $p = 1, d = 0, q = 0$ and $p = 0, d = 1, q = 1$. In the case of a step intervention, the SIA model implies that for $t \geq T + b$, an increase of ω occurred. So the SIA model with a step intervention can be regarded as the time-series generalization of the standard two-sample test for a change in location. In practice, this is one of the most frequently applicable models. Pulse interventions are useful for dealing with outliers (Chang, Tiao, and Chen 1988). A ramp intervention has been used to model the recovery trend in stratospheric ozone (Reinsel et al. 2002).

The SIA model may be generalized by allowing for multiple interventions and other types of interventions, as well as for seasonal ARIMA errors and possible covariates (Tiao et al. 1990; Weatherhead et al. 1998; Reinsel 2002; Reinsel et al. 2002). All of these situations are easily handled with the methods discussed in Sections 2.2 and 2.3. Power computations, although possible, are less useful when applied to dynamic response interventions, for the reasons explained in Appendix B.

2.2 Information Matrix

Letting $\lambda_1 = (\xi, \omega)$ and $\lambda_2 = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, f)$, it is shown in Appendix A that the expected Fisher information matrix is block diagonal with blocks, \mathcal{I}_{λ_1} and \mathcal{I}_{λ_2} corresponding to λ_1 and λ_2 . For the first block,

$$\mathcal{I}_{\xi, \omega} = \sigma_a^{-2} \mathbf{J}' \Gamma_n^{-1} \mathbf{J}, \tag{6}$$

where $\sigma_a^{-2} \Gamma_n^{-1}$ is the inverse of the covariance matrix of the stationary component and \mathbf{J} is an $n \times 2$ matrix with 1 in the first column and $\nabla^d I_t^{(T)}, t = 1, \dots, n$, in the second column. The Trench algorithm (Golub and Van Loan 1983) provides a computationally efficient method for computing Γ_n^{-1} . An expression essentially equivalent to (6) was obtained by Tiao et al. (1990) and Weatherhead et al. (1998) using generalized least squares. Assuming approximate normality of the estimates, the asymptotic variance of the maximum likelihood estimate of ω is found by taking the (2, 2) element of the inverse of (6),

$$\sigma_{\hat{\omega}} = \sqrt{(\mathcal{I}_{1,1} / (\mathcal{I}_{1,1} \mathcal{I}_{2,2} - \mathcal{I}_{1,2}^2))}, \tag{7}$$

where $\mathcal{I}_{i,j}$ denotes the (i, j) entry in the matrix $\mathcal{I}_{\xi, \omega}$. If the constant term, ξ , is not present, then $\sigma_{\hat{\omega}} = 1 / \sqrt{\mathcal{I}_{2,2}}$. When there is

an extensive amount of data before the intervention, it is sometimes helpful to simply correct the series by its sample mean and assume that $\xi = 0$ (Tiao et al. 1990).

The results of Pierce (1972) provide a computationally efficient approximation to (6) when $f = 0$. From Pierce [1972, eq. (3.2)] we can write the Fisher information for (ξ, ω) based on n observations as

$$\mathcal{I}_{\xi, \omega} = \sigma_a^{-2} \begin{pmatrix} n\kappa^2 & \kappa \sum_t v_t \\ \kappa \sum_t v_t & \sum_t v_t^2 \end{pmatrix}, \tag{8}$$

where $\kappa = -\phi(1)/\theta(1)$ and $v_t = -\phi(B)/\theta(B)w_t$, where $w_t = \nabla^d I_t^{(T)}$. Without loss of generality, we take $b = 0$, because if $b > 0$, the formulas hold with T replaced by $T + b$. Provided that T is not too small and T is not too close to n , (8) yields almost identical values to the more exact formula given in (6). New explicit expressions, using Pierce's approximation for AR(1) and IMA(1) cases, are given in Tables 1 and 2 for step, pulse and ramp interventions.

From (6), it follows that for consistency of the estimates $\hat{\xi}$ and $\hat{\omega}$, $\mathcal{I}_{\xi, \omega}/n$ (or, equivalently, $\mathbf{J}'\mathbf{J}/n$) must converge to a nonsingular matrix. For the intervention analysis models defined by (2)–(5), this occurs provided that

$$\frac{1}{n} \sum_{t=1}^n \nabla^d I_t^{(T)} \rightarrow c, \quad c > 0, c \neq 1. \tag{9}$$

If the constant term, ξ , is assumed to be known or 0, then only $c > 0$ is needed. This result is certainly not the whole story from an application standpoint. In Section 2.5 we show, using simulation experiments, that the empirical variances may be accurately estimated from (7) even when (9) is not satisfied.

2.3 Power and Sample Size

The null hypothesis $\mathcal{H}_0: \omega = 0$ can be tested using two asymptotically equivalent methods. The first method, referred to as the “Z-test,” uses $Z = \hat{\omega} / \hat{\sigma}_{\hat{\omega}}$, where $\hat{\omega}$ is the maximum likelihood estimate for ω and $\hat{\sigma}_{\hat{\omega}}$ is its estimated standard error. Note that $\sigma_{\hat{\omega}}$, the standard error of $\hat{\omega}$, depends only on the underlying ARIMA model in the preintervention period, and so it can be estimated before the postintervention data are obtained. A second, asymptotically equivalent method is to use a likelihood ratio test.

Table 1. Information Matrix for Simple Intervention Analysis With AR(1) Errors

Type	Information matrix entries
Step	$\mathcal{I}_{1,2}/\sigma_a^2 = (n - T)(1 - \phi_1)^2 + 1 - \phi_1$
	$\mathcal{I}_{2,2}/\sigma_a^2 = (n - T)(1 - \phi_1)^2 + 1$
Pulse	$\mathcal{I}_{1,2}/\sigma_a^2 = 1 - \phi_1^2$
	$\mathcal{I}_{2,2}/\sigma_a^2 = 1 - \phi_1^2$
Ramp	$\mathcal{I}_{1,2}/\sigma_a^2 = (1 + n - T)(1 - \phi_1)(2 + n - T - (n - T)\phi_1)/2$
	$\mathcal{I}_{2,2}/\sigma_a^2 = (1 + n - T)(6 + 7n + 2n^2 - 7T - 4nT + 2T^2 - 8n\phi_1$
	$\quad - 4n^2\phi_1 + 8T\phi_1 + 8nT\phi_1 - 4T^2\phi_1 + n\phi_1^2 + 2n^2\phi_1^2 - T\phi_1^2$ $\quad - 4nT\phi_1^2 + 2T^2\phi_1^2)/6$

NOTE: This table gives the (1, 2) and (2, 2) entries, $\mathcal{I}_{1,2}/\sigma_a^2$ and $\mathcal{I}_{2,2}/\sigma_a^2$. For each intervention type, $\mathcal{I}_{1,1}/\sigma_a^2 = n(1 - \phi_1)^2$, and the (2, 1) entry is obtained by symmetry.

Table 2. Information Matrix for Simple Intervention Analysis With IMA(1) Errors

Type	Information matrix entries
Step	$\mathcal{I}_{1,2}/\sigma_a^2 = (1 - \theta_1)^{-2}(1 - \theta^{n+1-T})$ $\mathcal{I}_{2,2}/\sigma_a^2 = (1 - \theta_1^2)^{-1}(1 - \theta^{2(n+1-T)})$
Pulse	$\mathcal{I}_{1,2}/\sigma_a^2 = (1 - \theta_1)^{-1}\theta_1^{n-T}$ $\mathcal{I}_{2,2}/\sigma_a^2 = (1 + \theta_1)^{-1}2(1 + \theta^{2(n-T)+1})$
Ramp	$\mathcal{I}_{1,2}/\sigma_a^2 = (1 - \theta_1)^{-3}(n + 1 - T + \theta_1^{n+2-T} - (n + 2 - T)\theta)$ $\mathcal{I}_{2,2}/\sigma_a^2 = (1 + \theta)^{-1}(1 - \theta_1)^{-3}(2\theta^{2n+T}(1 + \theta) - \theta^{4+2n} + \theta^{2T}(n + 1 - T - 2\theta - (2 + n - T)\theta^2))$

NOTE: For $\theta_1 = 0$, set $\theta_1^0 = 1$. The table gives the (1, 2) and (2, 2) entries, $\mathcal{I}_{1,2}/\sigma_a^2$ and $\mathcal{I}_{2,2}/\sigma_a^2$. For each intervention type, $\mathcal{I}_{1,1}/\sigma_a^2 = (n - 1)/(1 - \theta_1)^2$, and the (2, 1) entry is obtained by symmetry.

The asymptotic theoretical power function for the Z-test of the null hypothesis $\mathcal{H}_0 : \omega = 0$ against the two-sided alternative at level α is $\Pr\{|\hat{\omega}| > \mathcal{Z}_{1-\alpha/2}\sigma_{\hat{\omega}}|\omega\}$, where $\mathcal{Z}_{1-\alpha/2}$ is the upper $(1 - \alpha/2)$ -quantile in the standard normal distribution. For brevity, the asymptotic theoretical power function is referred to simply as the ‘‘power function.’’ In practice, this power function is approximated by replacing $\sigma_{\hat{\omega}}$ by an estimate, $\hat{\sigma}_{\hat{\omega}}$, based either on the preintervention data or on other prior knowledge. Often it is more convenient to use the rescaled parameter, $\delta = \omega/\sigma$, where σ^2 is the variance of the stationary error component, because in this case knowledge of σ^2 is not needed. The power function may be expressed in terms of δ as

$$\Pi(\delta) = \Phi(-\mathcal{Z}_{1-\alpha/2} - \delta\sigma/\sigma_{\hat{\omega}}) + 1 - \Phi(\mathcal{Z}_{1-\alpha/2} - \delta\sigma/\sigma_{\hat{\omega}}), \tag{10}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal. If the variance of the preintervention series, σ^2 , is known or estimated, then the power function for ω is $\Pi(\omega/\sigma)$. Equation (10) should be adjusted if only a one-sided alternative is under consideration.

As was done by Tiao et al. (1990), it is sometimes of interest to estimate the amount of additional data needed to detect an intervention of a specified magnitude with a prescribed power. The power function $\Pi(\delta)$ may be expressed more fully as a function of the test level α and the other underlying parameters n and T , so we can write the power function more fully as $\Pi(\delta, \alpha, n, T)$. For a fixed $\alpha = \alpha^{(0)}$, $\delta = \delta^{(0)}$, and a prescribed power $\Pi^{(0)}$, we may estimate the number of additional data values, m , that are required by numerically solving the equation $\Pi(\delta^{(0)}, \alpha^{(0)}, T + m - 1, T) = \Pi^{(0)}$. If, as in the geophysical datasets considered by Tiao et al. (1990), there is extensive preintervention data, we may assume that the mean is known and take $T = 1$ and solve $\Pi(\delta^{(0)}, \alpha^{(0)}, m, 1) = \Pi^{(0)}$. This technique is illustrated in Section 2.4, where it is also explained that in some situations, due to the limitations imposed by the model, there is no solution for m .

In general, the power and sample size computations for interventions with ARIMA and fractional ARIMA errors are easily done using an advanced quantitative programming environment, such as Mathematica, MatLab, S, or Stata. In the case of SIA with AR(1) or IMA(1) errors, power computations can even be done on a hand calculator.

2.4 Numerical Illustrations

In this section the power and sample size computations are illustrated for the SIA with a step intervention with AR(1), IMA(1), and fractionally differenced white noise. First, an approximation to the detection limit, δ' , is derived for the step intervention in an SIA model with unknown mean and stationary short-memory errors, with $f = d = 0$ and a fixed number, $T - 1$, of preintervention observations. The variance of the estimate, $\hat{\delta}$, may be written as $\text{var}(\hat{\delta}) \doteq \gamma_\delta/T$, where $\gamma_\delta = \sum_{k=-\infty}^{\infty} \gamma_k/\gamma_0$, γ_k is the autocovariance function for the stationary preintervention series, and $\gamma_0 = \sigma^2$. To achieve 90% power, $\Pr\{(\hat{\delta} - \delta')/\text{SE}(\hat{\delta}) > 1.96 - \delta'/\text{SE}(\hat{\delta})\} \doteq .9$. Hence, $2 - \delta'/\text{SE}(\hat{\delta}) \doteq -1.3$. So $\delta' \doteq 3.3\text{SE}(\hat{\delta})$.

Using Table 1, the power curve for the AR(1) with unknown mean, $n = 50$, $T = 25$, and $\phi_1 = .5$, $\sigma_\omega = .526681$. With $\sigma = 1/\sqrt{1 - \phi_1^2} = 1.1547$, the power curve is $\Pi(\delta) = 1 + \Phi(-1.960 - 2.192 \times \delta) - \Phi(1.960 - 2.192 \times \delta)$. This and the power curve obtained by letting $n \rightarrow \infty$ are shown in Figure 1, as is the approximate detection level, $\delta' \doteq \gamma_\delta/\sqrt{T} = 1.14$. For comparison, the exact value of δ' found by numerically solving $\Pi(\delta', .5, 10^9, 25) = .9$ is $\delta' = 1.12$. Assuming an unknown mean and that $T = 25$, we can find m , the number of additional observations needed to achieve a prescribed power level. For example, for 90% power with $\delta^{(0)} = 1.5$, solving

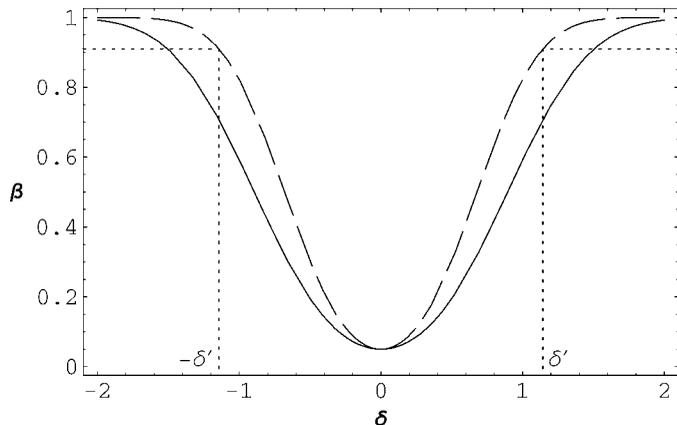


Figure 1. Comparison of Power Curves for $n = 50$, $T = 25$, and $n = \infty$, $T = 25$. The solid curve represents for $n = 50$, $T = 25$; the dashed curve, $n = \infty$, $T = 25$. The approximate detection limit, $\delta' \doteq 1.143$, is also shown.

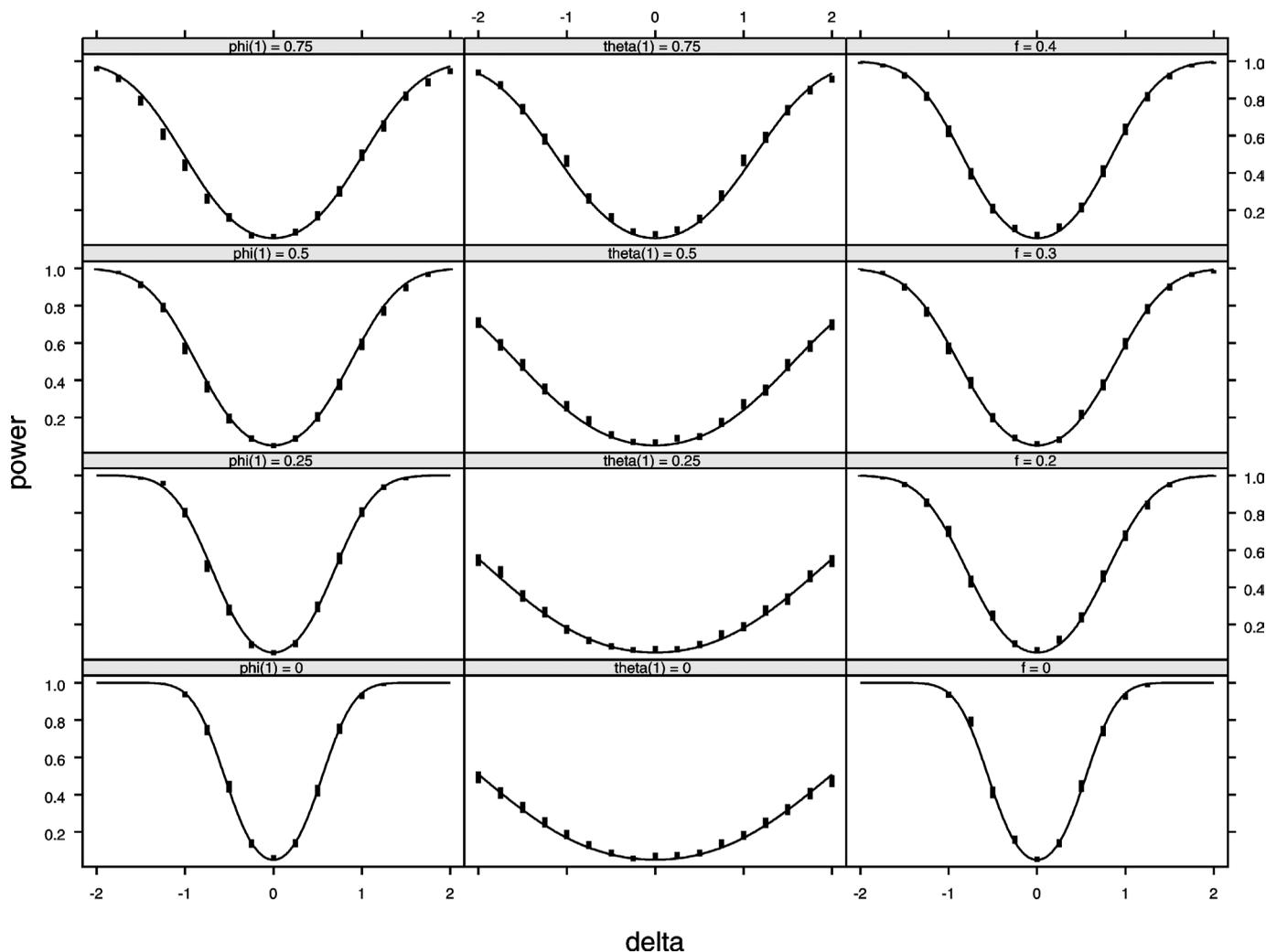


Figure 2. Comparison of Empirical and Theoretical Asymptotic Power in the SIA Model With AR(1), IMA(1), and Fractionally Differenced White Noise. The parameter $\delta = \omega/\sigma$ is the rescaled step size. The solid curve shows the theoretical power defined in (10). The vertical bars show the width of a 95% confidence interval for the empirical power in 1,000 simulations of the model. The AR(1) and IMA(1) parameters ϕ_1 and θ_1 are denoted by $\phi(1)$ and $\theta(1)$.

$\Pi(1.5, .05, 25 + m - 1, 25) = .9$, we find $m = 23$. In the known mean case, taking $T = 1$, we find $m = 10$. In the unknown mean case, if $\delta^{(0)} \leq \gamma_\delta$, then there is no solution, but if the mean is known, then m can always be found.

The middle panels of Figure 2 illustrate the power curves for an IMA(1) with $n = 50$ and $T = 25$. With $\theta_1 = .5$, $\Pi(\delta) = 1 + \Phi(-1.960 - 1.252 \times \delta) - \Phi(1.960 - 1.252 \times \delta)$.

Because long-memory or fractional time series have also been suggested for various types of geophysical data, it is of interest to examine the impact of this type of process on our ability to detect interventions. Table 3 compares the power of a two-sided 5% level test of the fractionally differenced white noise model $p = d = q = 0$ with $f = .2$ and $f = .4$ to the corresponding approximating ARMA(1, 1) when $n = 50$ and $T = 25$. The approximating ARMA(1, 1) model was determined by equating the first two autocorrelations in the fractional model with the first two autocorrelations in the ARMA(1, 1) model and solving to obtain the parameters ϕ_1 and θ_1 . In the first case with $f = .2$, the power is almost identical; in the second case with $f = .4$, the power is slightly higher for the ARMA(1, 1) approximation. This suggests that long-term

memory in the fractional noise model has little effect on the power when the length of the series is moderate, as in this example with $n = 50$ and $T = 25$. For sufficiently long time series, the effect on long memory is much more important, and the ARMA(1, 1) approximation does not hold.

Table 3. Power Function, $\Pi(\delta)$, for Fractionally Differenced White Noise With Parameter f and the Approximating ARMA(1, 1) Model for a Two-Sided 5% Level Test in the SIA Step Intervention Model With $n = 50$ and $T = 25$

δ	$f = .2$	$f = .4$
0	.050, .050	.050, .050
.5	.198, .202	.086, .076
1.	.602, .612	.198, .156
1.5	.914, .920	.384, .291
2.	.993, .994	.602, .468
2.5	1.000, 1.000	.792, .651
3.	1.000, 1.000	.914, .805

NOTE: The first entry in each pair is for the fractional model, and the second is for the ARMA(1, 1) model. The parameters in the approximating ARMA model are $\phi_1 = .667$, $\phi_2 = .451$ and $\theta_1 = .875$, $\theta_2 = .405$ corresponding to $f = .2$ and $f = .4$.

2.5 Simulation Experiment

The power function derived in (10) relies on the asymptotic normality of the maximum likelihood estimator, and so it is helpful to check its accuracy by simulation. We do this by comparing the power function with the empirical power function, $\hat{\Pi}$. For each simulated time series, all parameters in the model were estimated by exact maximum likelihood estimation, and the Z-test was computed. The empirical power, $\hat{\Pi}$, of a two-sided 5% test is then the proportion of times that the absolute value of this Z-statistic exceeded 1.96 in absolute value, and the 95% confidence interval for Π is $\hat{\Pi} \pm 1.96\sqrt{(\hat{\Pi}(1 - \hat{\Pi})/N)}$, where N is the number of simulations. For each model and each parameter setting, $N = 1,000$.

The model in (2) was simulated with $n = 50$ and $T = 25$ and AR(1) errors with $\phi_1 = 0, .25, .5, .75$, $\omega = \delta\sigma$, where $\delta = 0, \pm.25, \dots, \pm 2.0$. The empirical power confidence limits and theoretical power given by (10) are compared in Figure 2. It is seen that (10) provides an accurate approximation. IMA(1) is a commonly occurring nonstationary time series model. Figure 2 also compares the theoretical and empirical powers for the case with $n = 50$ and $T = 25$ using a two-sided Z-test at the 5% level. Once again, it is seen that (10) holds very well despite the small sample size. The values selected for θ_1 are positive, because this is the most common situation in practice. The power improves, as expected, as θ_1 increases from 0 to 1. Notice that this model does not satisfy (9). The last column of Figure 2 compares the empirical and theoretical powers in the case of fractionally differenced white noise, $p = q = d = 0$ for $f = .0, .2, .3, .4$. The approximation to the theoretical power improves with increasing f . The simulations shown in Figure 2 were repeated using the likelihood-ratio test, and essentially equivalent results were obtained.

In conclusion, the simulations in Figure 2 suggest that for practical purposes, if n , T , and $n - T$ are not too small, then the asymptotic theoretical power curve provides a good small-sample approximation. Alternatively, the simulations show that $\hat{\omega}$ is well approximated using its large-sample approximation even for moderately small samples. As already noted, $\sigma_{\hat{\omega}}$ also must be estimated by $\hat{\sigma}_{\hat{\omega}}$ using either the preintervention data or an estimate of its likely autocorrelation function. In practice, as in the example in Section 3.1, a range of likely parameter values is often used to indicate a range of possible power curves.

2.6 Model Uncertainty

Box, Jenkins, and Reinsel (1994) found that both the ARMA(1, 1) and IMA(1) fit series A, chemical process concentrations, about equally well. Both models give similar one-step-ahead forecasts, but very different long-run forecasts. The situation is similar with the power functions for these two models.

Consider a hypothetical step intervention that occurs immediately after the last observation. In this case $T = 198$, and the power curve as a function of ω is tabulated for a few selected values in Table 4 for a two-sided 5% test assuming that m postintervention observations are available for $m = 5$ and $m = 50$. When $m = 5$, the power curves are quite similar, but for $m = 50$, the power increases for the ARMA model but stays essentially the same for the IMA model. For example, Table 4 shows that there is a 75% chance of detecting a change of .6 with just five postintervention observations.

Table 4. Power Comparison for Step Interventions With ARMA(1, 1) and IMA(1) Errors for Series A With $n = 197 + m$ and $T = 198$

ω	ARMA(1, 1)		IMA(1)	
	$m = 5$	$m = 50$	$m = 5$	$m = 50$
.2	.141	.205	.141	.143
.3	.258	.398	.258	.264
.4	.415	.621	.416	.425
.5	.588	.809	.589	.600
.6	.745	.925	.746	.756
.7	.863	.978	.864	.872

NOTE: The models' other parameters are $\{\phi_1 = .9087, \theta_1 = .5758, \sigma_a = 0.3125\}$ and $\{\theta_1 = .7031, \sigma_a = .3172\}$.

2.7 Forecast-Actuality Significance Test

Box and Tiao (1976b) described an omnibus significance test for detecting whether an intervention has occurred. If a_t , $t = T, \dots, n$, denote the one-step-ahead prediction errors of an assumed model, then the test statistic may be written as $Q = \sum_{t=T}^n a_t^2 / \sigma_a^2$. If the intervention has no effect, then Q is approximately chi-squared distributed on $m = n - T + 1$ degrees of freedom. This significance test is easy to apply and does not require specification of an intervention model and its estimation. However, as might be expected, the loss of power can be considerable, as we now demonstrate.

As an example, consider the SIA model with a step intervention. Then it can be shown, using (4) of Box and Tiao (1976b), that $Q = \|\omega \mathbf{1}'_m \boldsymbol{\pi} / \sigma_a + \mathbf{a} / \sigma_a\|^2$, where $\mathbf{1}_m$ denotes the m -dimensional vector with a 1 in each position, $\mathbf{a} = (a_T, \dots, a_n)$, and $\boldsymbol{\pi} = (\pi_{i-j})$ is the lower triangular matrix with (i, j) entry π_{i-j} , where π_k is the coefficient of B^k in the expansion $\nabla^d \phi(B) / \theta(B) = 1 + \pi_1 B + \pi_2 B^2 + \dots$. So Q has a chi-squared distribution with m degrees of freedom and non-centrality parameter $\nu = (\omega^2 / \sigma_a^2) \|\mathbf{1}'_m \boldsymbol{\pi}\|^2$, and hence the large-sample power function can be computed. Figure 3 compares the power of this significance test with that of the SIA model hypothesis test for an example with $n = 120$, $T = 101$, and AR(1) errors. Figure 3 shows that the power of the significance test can be substantially less than that of the intervention analysis hypothesis test.

3. ILLUSTRATIVE APPLICATIONS

3.1 Traffic Safety and Public Policy

On May 1, 1996, liquor bar closing time in Ontario was changed from 1 AM to 2 AM. In a proposed intervention analysis, we wished to examine the possible effect of this change on late-night automobile fatalities. The data for this study comprised the total number of fatalities each month in Ontario during the hours of 11 PM to 4 AM for a period of years before and after May 1, 1996. For comparison, we also collected similar time series data for Michigan and New York. Data for this analysis were expensive to obtain, because raw records needed to be assembled, cleaned, and aggregated from sources in various jurisdictions. Initially, we planned to obtain monthly time series on the total number of fatalities from January 1994 to December 1998. This would yield $n = 60$ observations, and with the intervention occurring at $T = 36$. At additional cost, we could obtain complete monthly time series covering the period

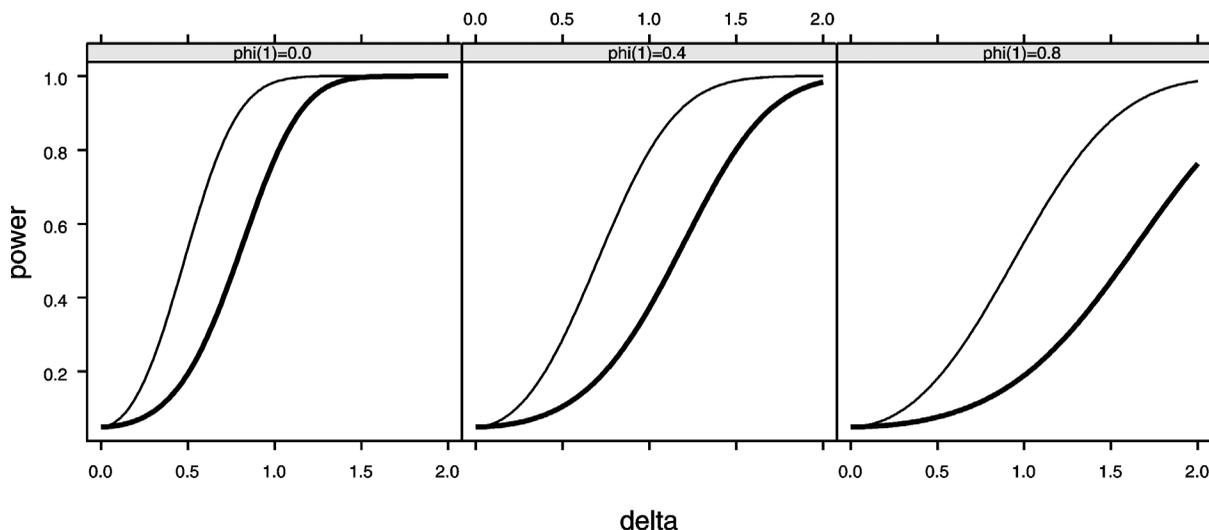


Figure 3. Comparison of Power Functions for a SIA Model With a Step Intervention With AR(1) Errors and the Forecast-Actuality Significance Test for a Two-Sided Test at the 5% Level. The model parameters are $n = 120$, $T = 101$, $\delta = \omega/\sigma$, and $\phi_1 = \phi_1$. The solid thin curve shows the SIA model-based hypothesis test, and the solid thick curve shows the omnibus significance test using Q . Because both power functions are symmetric about $\delta = 0$, only the upper half is shown.

January 1992–December 1998, which corresponds to $n = 84$ and $T = 48$. We were interested to know whether ($n = 60$, $T = 36$) or ($n = 84$, $T = 48$) would be sufficient to detect change of σ or greater with a reasonably high probability, where σ is the standard deviation of the preintervention series.

Based on previous experience with similar time series (Vingilis, Blefgen, Lei, Sykora, and Mann 1988), we expected the time series to exhibit small autocorrelations that may be modeled by an AR(1) model with parameter $\phi_1 \leq .5$. The intervention was expected to cause an increase in late-night fatalities, so a one-sided upper-tail test is appropriate. The power function in this case is $\Pi(\delta) = 1 - \Phi(1.645 - 2.362 \times \delta)$. Table 5 gives the power of a 5% upper-tail test for these two plans for various ϕ_1 . When $\phi_1 = .5$, Table 5 shows that ($n = 84$, $T = 48$) has a 86.7% chance of detecting a step intervention whose magnitude is only one standard deviation of the error component, whereas the corresponding power for ($n = 60$, $T = 36$) is 76.3%. The results of Table 5 demonstrated to the satisfaction of us and the granting agency that ($n = 84$, $T = 48$) had a good chance of detecting a meaningful change and was worth the extra expenditure.

Table 5. Power Comparison for AR(1) Errors for ($n = 60$, $T = 36$) and ($n = 84$, $T = 48$)

δ	$\phi_1 = 0$	$\phi_1 = .25$	$\phi_1 = .5$	$\phi_1 = .75$
.000	.050, .050	.050, .050	.050, .050	.050, .050
.250	.245, .306	.186, .226	.146, .170	.124, .135
.500	.604, .736	.444, .555	.321, .395	.253, .288
.750	.889, .961	.729, .848	.550, .664	.431, .493
1.000	.985, .998	.914, .973	.763, .867	.624, .700
1.250	.999, 1.000	.983, .998	.904, .964	.790, .857
1.500	1.000, 1.000	.998, 1.000	.971, .994	.903, .946
1.750	1.000, 1.000	1.000, 1.000	.994, .999	.963, .984
2.000	1.000, 1.000	1.000, 1.000	.999, 1.000	.989, .996

NOTE: The first entry in each column corresponds to ($n = 60$, $T = 36$) and the second ($n = 84$, $T = 48$).

3.2 Detecting Ozone Turnaround

Tiao et al. (1990) used the SIA model with a ramp intervention with AR(1) errors to model the trend in monthly deseasonalized stratospheric ozone and other environmental variables. For simplicity, Tiao et al. (1990) assumed that the mean of the preintervention series was known. It may be shown that the expression obtained by Tiao et al. (1990, app. A) for $\sigma_{\hat{\omega}}$ is exactly equal to $\sigma_{\hat{\omega}} = 1/\sqrt{I_{2,2}}$ using Table 1 with $n = T$ and $T = 1$. Table 6 compares this result with the corresponding result obtained using the exact expected Fisher information matrix given in (6) for the same parameters as used by Tiao et al. (1990, table 1). When $\phi = .8$, the difference is as high as 17%, but it decreases as the sample size increases. The approximation is very good for parameter values $\leq .6$. For most of the geophysical time series considered by Tiao et al. (1990), the degree of autocorrelation is quite low, so this approximation works well.

Tiao et al. (1990, table 2) also considered the number of years of monthly data needed to detect a ramp intervention for several geophysical time series of interest. In their computations it was assumed that $T = 1$ and that the mean was known. Table 7 computes the number of years of data needed for these time series under the assumptions that the mean is unknown but there are 30 years of prior data. The other assumptions about the data and the form of the intervention are the same as those of Tiao et al. (1990). The parameter δ shown in the table was based

Table 6. Comparison of Exact and Approximate Methods

Number of years	$\phi = .6$	$\phi = .8$
6	-6	-17
7	-5	-15
8	-5	-13
9	-4	-11
10	-4	-10

NOTE: The function $g(T, \phi)$ defined by Tiao et al. (1990) was computed using the exact form of the information matrix (6) and the approximation (8) for selected parameter values given in table 1 of Tiao et al. (1990). The entries in the table show the percentage difference, $100 \times (\text{EXACT} - \text{APPROXIMATE})/\text{EXACT}$.

Table 7. Number of Years, n^* , for 90% Probability of Detecting a Prescribed Trend, δ Using a Two-Sided 5% Test Given 30 Years of Prior Data and Assuming AR(1) Errors With Estimated Parameter $\hat{\phi}_1$

	Tateno	Hohen	Wakkan	Bulawayo	Abidajan
$\hat{\phi}_1$.32	.05	.14	.43	.65
ω	.003	.003	.2	.2	.2
δ	.00758	.00543	.01042	.01282	.01111
n^*	11.6	12.1	8.0	8.6	12.0
n^*_{Tiao}	14	14	10	10	13

NOTE: The last line of the table shows the comparable values given by Tiao et al. (1990, table 2).

on the information supplied by Tiao et al. (1990). Specifically, $\delta = \omega / (12 \times \hat{\sigma})$, where $\hat{\phi}_1$ and $\hat{\sigma}$ are as obtained by Tiao et al. (1990, table 2) and ω is as obtained by Tiao et al. (1990, pp. 20, 510). Note that ω was divided by 12, because the form of the intervention used by Tiao et al. (1990) was $R_t^{(T+1)} / 12$. In conclusion, the estimate of the sample size required shown in Table 7 is in reasonable agreement with the results of Tiao et al. (1990).

4. CONCLUDING REMARKS

We have shown how the power function for an intervention analysis may be computed provided that we have an estimate of the ARIMA parameters in the preintervention time series or in some closely related time series. In the case of the SIA model with AR(1) or IMA(1) errors, the power function can be easily computed using a hand calculator. Such programs are freely available for the Texas Instruments TI-83 at the first author's webpage. *Mathematica* and S software for computing the power functions and all tables and figures described in this article are also available there, as are various other supplements to this article.

The emphasis of this article has been on using the power function as an aid in selecting the sample size. In the case of the SIA model, if $\Pi(\omega') = 1 - \beta'$ for a 5% two-sided test of $\mathcal{H}_0: \omega = 0$, then the usual 95% confidence interval for ω will contain 0 with probability β' when $\omega = \omega'$. Thus the power function may be used as an aid in choosing the sample size so that a useful confidence interval is obtained. Instead of the power function, we could have focused on the width of a suitable interval estimate of ω . Because this also depends on an estimate of σ_ω , the methods presented are applicable. It may be noted that overemphasis on hypothesis tests has long been condemned, as was noted many years ago by Cox (1977). Nevertheless, as indicated by Cox (1977), such tests remain important in practice.

The power function depends strongly on the degree of autocorrelation in the preintervention time series. In the stratospheric ozone example, Section 3.2, a long preintervention series was available that enabled the model to be accurately estimated. In other cases, such as the traffic safety example in Section 3.1, the preintervention series is either unavailable or quite short. In such cases prior information may be available that indicates a range of likely models. As discussed in Section 3.1, this still may be very useful for planning purposes. A final note of caution, power computations should be used only before the data analysis is done (Hoening and Heisey 2001; Lenth 2001) and should never be used to compute the observed power after a test of hypothesis has already been carried out.

ACKNOWLEDGMENTS

This research was supported by grants from the NIAAA and NSERC. The authors thank the editor, an associate editor, two referees, and R. J. Kulperger for helpful suggestions.

APPENDIX A: DERIVATION OF THE INFORMATION MATRIX

The log-likelihood function, apart from a constant, may be written as

$$L(\lambda_1, \lambda_2, \sigma_a^2) = -\log(\sigma) - \log(\det(\Gamma_n)) - \frac{1}{2\sigma_a^2} \mathbf{y}' \Gamma_n^{-1} \mathbf{y}, \tag{A.1}$$

where \mathbf{y} is the column vector of length $n - d$ with t th entry $\nabla^d z_t - \xi - \omega \nabla^d S_t^{(T)}$, $t = d + 1, \dots, n$. Then $\partial \mathbf{y} / \partial \xi = (-1, \dots, -1)$. Similarly, $\partial \mathbf{y} / \partial \omega = (-S_1^{(T)}, \dots, -S_n^{(T)})$. Hence

$$\begin{aligned} \mathcal{I}_{\lambda_1} &= -E(\partial_{\lambda_1}^2 L(\lambda_1, \lambda_2, \sigma_a^2)) \\ &= \frac{1}{\sigma_a^2} \mathbf{J}' \Gamma_n^{-1} \mathbf{J}, \end{aligned} \tag{A.2}$$

where \mathbf{J} is as in (6). Because $E(\partial^2 L(\lambda_1, \lambda_2, \sigma_a^2) / (\partial \lambda_1 \partial \lambda_2)) = \mathbf{0}$ and $E(\partial^2 L(\lambda_1, \lambda_2, \sigma_a^2) / (\partial \lambda_1 \partial \sigma_a^2)) = \mathbf{0}$, the information matrix is block diagonal.

APPENDIX B: INTERVENTIONS WITH A DYNAMIC RESPONSE

For completeness, we also discuss the intervention analysis model with a dynamic response to the intervention, which may be written as

$$\begin{aligned} \nabla^d z_t &= \xi + \omega(B) / \delta(B) \nabla^d B^b I_t^{(T)} + \nabla^{-f} \frac{\theta(B)}{\phi(B)} a_t, \\ &t = 1, \dots, n, \end{aligned} \tag{B.1}$$

where $\omega(B) = \omega_0 + \omega_1 B + \dots + \omega_r B^r$ and $\delta(B) = \delta_0 - \delta_1 B - \dots - \delta_s B^s$. For stability of the transfer function, it is assumed that all roots of $\delta(B) = 0$ lie outside the unit circle. As in Appendix A, the exact information matrix for the parameters is $\lambda_1 = (\xi, \omega_0, \dots, \omega_r, \delta_1, \dots, \delta_s) \mathcal{I}_{\lambda_1} = \sigma_a^{-2} \mathbf{J}' \Gamma_n^{-1} \mathbf{J}$, where \mathbf{J} is an $(n - d) \times (2 + r + s)$ matrix with rows $(1, u_t, \dots, u_{t-r}, v_t, \dots, v_{t-s})$ for $t = 1, \dots, n - d$, where $u_{t-j} = \nabla^d (1/\delta(B)) I_{t-j}^{(T)}$ and $v_{t-j} = \nabla^d (\omega(B)/\delta(B)) I_{t-j}^{(T)}$. Alternatively, the large-sample approximation given by Pierce (1972) may be used. The steady-state gain (Box et al. 1994, sec. 10.1.1), which measures the

long-run change of the intervention, is defined by $g = (\omega_0 + \dots + \omega_r)/(1 - \delta_1 - \dots - \delta_s)$. The maximum likelihood estimates for the model may be used to form the estimate of g , \hat{g} . Using a Taylor series linearization, the standard deviation of \hat{g} is given by $\sigma_{\hat{g}} = \sqrt{\mathbf{d}'_g \mathbf{V}_g \mathbf{d}_g}$, where \mathbf{V}_g is obtained by dropping the first row and column from $\mathcal{I}_{\lambda_1}^{-1}$ and $\mathbf{d}_g = (\partial g/\partial \omega_0, \dots, \partial g/\partial \omega_r, \partial g/\partial \delta_1, \dots, \partial g/\partial \delta_s)$. For dynamic intervention analysis models, we may consider testing $\mathcal{H}_0: g = 0$ using the Z test. Notice that when $s > 0$, we need estimates of all parameters in the full intervention model to estimate $\sigma_{\hat{g}}$. This limits the applicability of this approach, because even if the preintervention series were known, it is not likely that such precise information would be available for the intervention parameters. Often the SIA model can be used to get an approximation to the power in this case.

As a numerical illustration, consider the dynamic step intervention model, $z_t = \xi + \omega_0(1)/(1 - \delta_1 B)S_t^{(T)} + a_t/(1 - \phi_1 B)$, $t = 1, \dots, n$. Taking $n = 50$, $T = 25$, $\xi = 0$, $\phi_1 = .5$, and $\sigma_a^2 = 1$, Table B.1 compares the power of a 5% two-sided test, $\mathcal{H}_0: g = 0$, where $g = \omega_0(1)$, with that of the Z-test $\mathcal{H}_0: \omega_0^{(2)} = 0$ in the corresponding SIA model defined by $z_t = \xi + \omega_0^{(2)}S_t^{(T)} + a_t/(1 - \phi_1 B)$, where $\omega_0^{(2)} = g$ and the other parameter settings are the same. On an intuitive basis, the effect in the SIA model is slightly larger, so one might expect the power in the SIA model to be slightly larger. Table B.1 shows, comparing the first two entries in each triplet, that this is exactly what happens. The third entry in each triplet in Table B.1 is the empirical power of a two-sided 5% test of $\mathcal{H}_0: \omega_0^{(2)} = 0$ when the SIA model is fitted to a time series generated by the dynamic step intervention model. A total of 1,000 simulations were used for each model. The empirical power is predicted well by the theoretical asymptotic power for the SIA model. These simulations were repeated with various values of the parameter ϕ , and similar results were found when $-1 < \phi \leq .5$. For $\phi_1 > .5$, there was a much larger difference between the asymptotic theoretical power of the dynamic model and the step model. For example, with $\phi_1 = .9$, $\omega_1 = .75$ and $\delta_1 = .75$, the asymptotic power for the two-sided 5% level gains test was only .199, whereas the predicted power using a SIA step intervention was .972. The empirical power of the two-sided 5% level test of $\mathcal{H}_0: \omega_1 = 0$ in the step SIA model was .283. The general conclusion reached was that the step SIA model provides a useful approximation to the more complicated dynamic step intervention model provided that the autocorrelation is not too large. Further simulation results are available in the online supplements.

Table B.1. Power Comparisons of the Dynamic Step Intervention Model With Simple Step Intervention When $n = 50$ and $T = 25$

δ_0	$\omega_0 = .5$	$\omega_0 = .75$	$\omega_0 = 1.0$
.25	.226, .252, .241	.416, .490, .466	.879, .972, .880
.50	.439, .490, .445	.745, .827, .758	.997, 1.000, .974
.75	.673, .732, .692	.937, .972, .932	1.000, 1.000, .955

NOTE: The first entry in each triplet shows the theoretical power of a 5% two-sided test of $\mathcal{H}_0: g = 0$, where $g = \omega_0^{(1)}/(1 - \delta_1)$ in the dynamic step intervention model $z_t = \xi + \omega_0^{(1)}/(1 - \delta_1 B)S_t^{(T)} + a_t/(1 - \phi_1 B)$ with $\xi = 0$, $\phi_1 = .5$, and $\sigma_a^2 = 1$. The second entry is the theoretical power of a 5% test of $\mathcal{H}_0: \omega_0^{(2)} = 0$ in the SIA model, $z_t = \xi + \omega_0^{(2)}S_t^{(T)} + a_t/(1 - \phi_1 B)$, where $\omega_0^{(2)} = \omega_0^{(1)}/(1 - \delta_1)$ and all other parameters are the same as in the dynamic model. The third entry is the empirical power, based on 1,000 simulations, for a two-sided 5% test of $\mathcal{H}_0: \omega_0^{(2)} = 0$ when the SIA model is fitted to a time series generated by the dynamic step intervention model.

[Received August 2003. Revised October 2004.]

REFERENCES

- Beran, J. (1994), *Statistics for Long Memory Processes*, London: Chapman & Hall.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: Wiley.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control* (3rd ed.), San Francisco: Holden-Day.
- Box, G. E. P., and Luceño, A. (1997), *Statistical Control by Monitoring and Feedback Adjustment*, New York: Wiley.
- Box, G. E. P., and Tiao, G. C. (1976a), "Intervention Analysis With Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 70, 70–79.
- (1976b), "Comparison of Forecast and Actuality," *Applied Statistics*, 25, 195–200.
- Cao, J., Cleveland, W. S., Lin, D., and Sun, D. X. (2001), "On the Nonstationarity of Internet Traffic," *Performance Evaluation Review: Proc. ACM Sigmatics*, 29, 102–112.
- Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.
- Cox, D. R. (1977), "The Role of Statistical Significance Tests," *Scandinavian Journal of Statistics*, 4, 49–70.
- Golub, G., and Van Loan, C. F. (1983), *Matrix Computations*, Baltimore: Johns Hopkins University Press.
- Haslett, J., and Raftery, A. E. (1989), "Space–Time Modelling With Long-Memory Dependence: Assessing Ireland's Wind Power Resource," *Applied Statistics*, 38, 1–21.
- Hipel, K. W., and McLeod, A. I. (1994), *Time Series Modelling of Water Resources and Environmental Systems*, Amsterdam: Elsevier.
- Hoenig, J. M., and Heisey, D. M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis," *The American Statistician* 55, 19–24.
- Jiang, W., Tsui, K. L., and Woodall, W. H. (2000), "A New SPC Monitoring Method: The ARMA Chart," *Technometrics*, 42, 399–410.
- Kärner, O. (2002), "On Nonstationarity and Antipersistence in Global Temperature Series," *Journal of Geophysical Research*, 107, D20, 4415.
- Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, 55, 187–193.
- Luceño, A. (1995), "Choosing the EWMA Parameter in Engineering Process Control," *Journal of Quality Technology*, 27, 162–168.
- Noakes, D. J., and Campbell, A. (1992), "Use of Geoduck Clams to Indicate Changes in the Marine Environments of Ladysmith Harbour, British Columbia," *Environmetrics*, 3, 81–97.
- Pierce, D. A. (1972), "Least Squares Estimation in Dynamic-Disturbance Time Series Models," *Biometrika*, 59, 73–78.
- Reinsel, G. C. (2002), "Trend Analysis of Upper Stratospheric Umkehr Ozone Data for Evidence of Turnaround," *Geophysical Research Letters*, 29(10), doi: 10.1029/2002GL014716.
- Reinsel, G. C., Weatherhead, E. C., Tiao, G. C., Miller, A. J., Nagatani, R. M., Wuebbles, D. J., and Flynn, L. E. (2002), "On Detection of Turnaround and Recovery in Trend for Ozone," *Journal of Geophysical Research*, 107(D10), doi: 10.1029/2001JD000500.
- Roberts, H. (1991), *Data Analysis for Managers With Minitab* (2nd ed.), San Francisco: Scientific Press.
- Tiao, G. C., Reinsel, G. C., Xu, D., Pedrick, J. H., Zhu, X., Miller, A. J., DeLuisi, J. J., Mateer, C. L., and Wuebbles, D. J. (1990), "Effects of Autocorrelation and Temporal Sampling Schemes on Estimation of Trend and Spatial Correlation," *Journal of Geophysical Research*, 95, 20507–20517.
- Vingilis, E., Blefgen, H., Lei, H., Sykora, K., and Mann, R. (1988), "An Evaluation of the Deterrent Impact of Ontario's 12-Hour Licence Suspension Law," *Accident Analysis and Prevention*, 20, 9–17.
- Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X. L., Choi, D., Cheang, W. K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E. (1998), "Factors Affecting the Detection of Trends: Statistical Considerations and Applications to Environmental Data," *Journal of Geophysical Research*, 103, 17149–17161.
- Wolfram, S. (1999), *The Mathematica Book* (4th ed.), Champaign, IL/Cambridge, U.K.: Wolfram Media/Cambridge University Press.