

Friday April 18, 2008.



Advances in Statistical Computing For Linear Time Series



A. Ian McLeod
aimcleod@uwo.ca

Introduction

New methods and their implementation in R packages are discussed.

- **general linear time series model: estimation, simulation and forecasting**
- **cluster computing in R**
- **faster subset AR and ARMA for long time series**

improved portmanteau goodness-of-fit test

- medical time series
- Burg estimator

Itsa Package

A. Ian McLeod, Hao Yu, Zinovi L. Krougly (2007). Algorithms for Linear Time Series Analysis: With R Package, *Journal of Statistical Software* 23, 5. <http://www.jstatsoft.org/v23/i05>

■ Durbin-Levinson Algorithm

Fast and accurate. $O(k^2)$ flops

Given a time series $z_t, t = 1, 2, \dots$ with TACVF, $\gamma_k, k = 0, 1, \dots$ we can determine $\phi_{k,j}, j = 1, \dots, k$ the coefficients in the optimal linear predictor $\hat{z}_t = \phi_{k,1} z_{t-1} + \dots + \phi_{k,k} z_{t-k}$ using the DL recursion.

Set $\phi_{1,1} = \gamma_1 / \gamma_0$ and $v_1 = (1 - \phi_{1,1}^2) \gamma_0$, where v_k denotes the variance of the k step linear predictor. Then for $k = 2, 3, \dots$ we can iteratively obtain,

$$\phi_{k,k} = (\gamma_k - \phi_{k-1,1} \gamma_{k-1} - \dots - \phi_{k-1,k-1} \gamma_1) / v_{k-1} \quad \text{DL-1}$$

$$\begin{pmatrix} \phi_{k,1} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{k,k-1} \end{pmatrix} = \begin{pmatrix} \phi_{k-1,1} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{k-1,k-1} \end{pmatrix} - \phi_{k,k} \begin{pmatrix} \phi_{k-1,k-1} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{k-1,1} \end{pmatrix} \quad \text{DL-2}$$

and

$$v_k = v_{k-1} (1 - \phi_{k,k}^2). \quad \text{DL-3}$$

AR To Pacf $\phi \rightarrow \zeta$

$$\phi_{k-1,i} = (\phi_{k,i} + \phi_{k,k} \phi_{k,k-i}) / (1 - \phi_{k,k}^2), \quad (1)$$

$$k = p - 1, \dots, 1; i = 1, \dots, k - 1$$

From DL recursion we can write,

$$\phi_{k-1,i} = \phi_{k,i} + \phi_{k,k} \phi_{k-1,k-i} \quad (2)$$

by symmetry,

$$\phi_{k-1,k-i} = \phi_{k,k-i} + \phi_{k,k} \phi_{k-1,i} \quad (3)$$

Subing (2) in (3) and simplifying yields (1).

Monahan JF (1984). "A Note on Enforcing Stationarity in Autoregressive Moving Average Models." *Biometrika*, 71, 403-404.

■ **Pacf To AR** $\zeta \rightarrow \phi$

$$\phi_{k,i} = \phi_{k-1,i} - \zeta_k \phi_{k-1,k-i};$$

$$k = 2, \dots, p; i = 1, \dots, k - 1.$$

This follows directly from eqn. (1).

■ Exact Likelihood Computation

Consider n successive observations from a stationary Gaussian process with mean zero and autocovariance function γ_k . We assume that the autocovariance function is specified by a parameter vector β together with the innovation variance σ_a^2 . Using the Durbin-Levinson algorithm with the γ_k Exact one-step prediction error using the last t observations,

$$e_t = z_t - \phi_{t,1} z_{t-1} - \dots - \phi_{t,t-1} z_1 \quad (4)$$

Then e_t are independent normal with mean zero and variance $\sigma_a^2 v_{t-1}$, where v_{t-1} denotes the variance of the $t-1$ step linear predictor produced by the Durbin-Levinson algorithm. Hence the loglikelihood function for β and σ_a^2 may be written,

$$L(\beta) = -\frac{n}{2} \mathbf{Log}(S(\beta))$$

where

$$S(\beta) = \sum_{t=1}^n e_t^2 / v_t.$$

This method seems to be first discussed by in W.K. Li (1981). *Topics in Time Series Analysis*. Ph.D. thesis, University of Western Ontario.

■ Simulation

Three methods are implemented:

- Using Durbin-Levinson, ie. eqn (4), $O(n^2)$ flops

Using Davies-Harte method. **Davies RB, Harte DS (1987). Tests for Hurst Effect. *Biometrika*, 74, 95-101.** An example is given with fractionally-differenced white noise in our package. See help for the function **DHSimulate**. $O(n \log n)$ flops

- **FFT Method.** $O(n \log n)$ flops

Stationary and ergodic time series can approximated by,

$$z_t \doteq a_t + \psi_1 a_{t-1} + \dots + \psi_{t-Q} a_{t-Q} \quad (5)$$

where a_t are the innovations usually assumed to be IID $(0, \sigma_a^2)$ and Q is sufficiently large.

Taking $m = 2^J$ for some positive integer J and generate innovations $a_t, t = 1, \dots, m$. Using eqn. (5), compute $z_t, t = Q + 1, \dots, m$. In R this computation can be done efficiently using **convolve** which uses the FFT.

■ Trench Algorithm and Forecasting

Trench Algorithm Derivation: Siddiqui (1958, Annals) derived an expression for the inverse of the covariance matrix of AR. So to invert an arbitrary positive-definite symmetric Toeplitz matrix of order n , use a linear predictor to obtain order n linear predictor and then use Siddiqui's result. Using DL recursion for the linear predictor, the whole algorithm is only $O(n^2)$ flops instead of $O(n^4)$.

Let $\hat{z}_n(k)$ denote the minimum mean square error linear predictor of z_{n+k} given the data $z = (z_1, \dots, z_n)$ and the autocovariances $\gamma_l, l = 0, \dots, n - 1$. Then as shown by Hamilton (1994, §4.3) the finite-sample predictor may be written

$$\hat{z}_n(k) = \mu + g_k \Gamma_n^{-1}(z - \mu)$$

where $g_k = (\gamma_{n+k-1}, \dots, \gamma_k)$ and $\text{var}(\hat{z}_n(k)) = \gamma_0 - g_k^T \Gamma_n^{-1} g_k$. In many forecasting applications, a model is fitted to the data to the first n observed time series values and then k -step ahead forecasting are needed for each of m additional observations, $(z_{n+1}, \dots, z_{n+m})$. These forecasts may be written

$$\hat{z}_{n+l}(k) = \mu + g_{n,l} \Gamma_{n+l}^{-1} (z^{(n+l)} - \mu),$$

where $g_{n,l} = (\gamma_{n+l+k-2}, \dots, \gamma_k)$ and $z^{(n+l)} = (z_1, \dots, z_{n+l-1})$. Rather than evaluating m matrix inverses of dimension $n+1, \dots, n+m$ we just need to evaluate Γ_n^{-1} and the remaining inverses can be obtained by updating. This effectively reduces the matrix inverse computations from $O(m(n+m)^2)$ flops to just $O(n(n+m))$ flops. Hamilton (1994, §4.4) discusses the use of the Cholesky decomposition for finite sample forecast updating but the Trench algorithm provides a more computationally efficient approach since Cholesky decomposition takes $O(n^4)$ flops.

More generally, we may be interested in computing the k -step predictor for $k = 1, \dots, L$. The vector of forecasts $\hat{z}_n^{(L)} = (z_n(1), \dots, z_n(L))$ may be written,

$$\hat{z}_n^{(L)} = \mu + G_L \Gamma_n^{-1} (z - \mu)$$

where $G_L = (\gamma_{n+i-j})_{n \times L}$, where $i = 1, \dots, L$ and $j = 1, \dots, n$.

For updating from n to $n+m$ we can write,

$$\hat{z}_{n+m}^{(L)} = \mu + G_{L,m} \Gamma_{n+m}^{-1} (z^{(n+m)} - \mu),$$

where $G_{L,m} = (\gamma_{n+m+i-j-1})_{(n+m-1) \times L}$, where $i = 1, \dots, L$ and $j = 1, \dots, n+m-1$.

Hamilton, J.D. (1994). *Time Series Analysis*. New Jersey: Princeton University Press.

■ Toeplitz Inverse Updating Formula

$$\Gamma_{n+1}^{-1} = \begin{pmatrix} A & -f \\ f' & e \end{pmatrix}$$

$$\Gamma_n = (\gamma_{i-j})_{n \times n}$$

$$g = (\gamma_n, \gamma_{n-1}, \dots, \gamma_1)'$$

$$e = 1 / (\gamma_0 - g' \Gamma_n^{-1} g)$$

$$f = -\Gamma_n^{-1} g e$$

$$A = \Gamma_n^{-1} + \Gamma_n^{-1} g g' \Gamma_n^{-1} e$$

Cluster Computing

Beowulf cluster computer provides economical supercomputing. Difficult to use if you program only in C++. Relatively easy to use with Rmpi. **Hao Yu (2002). Rmpi: Parallel Statistical Computing in R." R News, 2(2), 10-14.**

Our departmental facility has 58 nodes and cost about £30,000. Simulations which would tie up a PC for two days can be one in an hour.

■ Illustrative Application: Forecasting Example

Fractional Gaussian Noise (FGN) has been a popular model for long-memory time series and the discrete-time version is defined by its TACF,

$$\rho_k = \left(|k+1|^{2H} - 2k^{2H} + |k-1|^{2H} \right) / 2$$

A series of length $n = 663$ was simulated (same as Nile Minimum Flow Data). Forecasts at lead times $l = 1, 2, 3$ were computed using origins $m = 563, \dots, 663$ and the empirical RMSE determined using the known parameter $H = 0.831$. To gauge the effect of model uncertainty, the FGN was fit to each of the series, that is $n = 563, \dots, 663$ and for each fit all out-of-sample forecasts were computed for each lead time $l = 1, 2, 3$ and the average RMSE determined. To see the effect of incorrect model specification, an ARMA(2,1) model was fit and the forecasting repeated for each value. All simulations were repeated 1000 times. These computations take about 6 hours on our current configuration.

Lead	FGN	FGN	ARMA(2, 1)
Time	$H = 0.831$	$H = \hat{H}$	$(\phi_1, \phi_2, \theta_1) = (\hat{\phi}_1, \hat{\phi}_2, \hat{\theta}_1)$
1	0.545	0.546	0.546
2	0.594	0.596	0.597
3	0.610	0.612	0.613

R script see: <http://jss-dev.stat.ucla.edu/v23/i05>

■ Illustrative Application: MLE for H

Long memory time series are characterized by $\gamma_k = O(k^{-\alpha})$. For FGN, $\alpha = 2(1 - H)$ and for fractionally differenced white noise, $\alpha = 1 - 2d$. So in this sense, we have $H = d + 1/2$. Now Li (1981) and Li and McLeod (1988) showed that $\text{Var}(\hat{d}) = \pi^2 / (6n)$.

FitAR Package

McLeod AI and Zhang Y (2006). Partial Autocorrelation Parameterization for Subset Autoregression. *Journal of Time Series Analysis*, 27, 599–612.

A. Ian McLeod and Ying Zhang (2008, to appear). Improved Subset Autoregression: With R Package, *Journal of Statistical Software* 23, 5.

■ Subset AR Models:

Principle of Parameter Parsimony suggests considering models such as,

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_9 x_{t-9} + a_t$$

These models may be fit using least-squares. Least squares subset regression algorithms may be used. In general the $\text{AR}_\phi(i_1, i_2, \dots, i_m)$ is defined by

$$x_t = \phi_{i_1} x_{t-i_1} + \phi_{i_2} x_{t-i_2} + \dots + \phi_{i_m} x_{t-i_m} + a_t$$

Alternative to ARMA models.

■ New Subset Models

$(\zeta_{i_1}, \dots, \zeta_{i_m}) \rightarrow (\phi_1, \dots, \phi_p), i_m = p,$ defined using
Pac fToAR and letting $\zeta_i \in (-1, 1)$. The admissible region is simply a cube in m -dimensions.

Barndorff-Nielsen, O. and Schou, G. (1973). On the parametrization of autoregressive models by partial autocorrelations. *Journal of Multivariate Analysis*, 3, 408-419.

■ Theoretical Developments

Derived the asymptotic distributions of the MLE and the residual autocorrelations. Developed computationally efficient exact MLE algorithm.

■ Exact Loglikelihood

$$L(\phi, \sigma_a^2) = -\frac{1}{2} \text{Log}(\det(\Gamma_n)) - \frac{1}{2} \mathbf{x}' \Gamma_n^{-1} \mathbf{x}$$

where Γ_n is the covariance matrix of $x = (x_1, \dots, x_n)$ and $\phi = (\phi_1, \dots, \phi_p)$.

Champernowne (1948) and Box and Jenkins (1970) showed that

$$\mathbf{x}' \Gamma_n^{-1} \mathbf{x} = \beta' D \beta / \sigma_a^2,$$

where $\beta = (-1, \phi_1, \dots, \phi_p)$ and D is the $(p+1)$ -by- $(p+1)$ matrix with (i, j) -entry

$$D_{i,j} = D_{j,i} = x_i x_j + \dots + x_{n+1-i} x_{n+1-j}.$$

The standardized covariance determinant of order p , $g_p = \det(\Gamma_p / \sigma_a^2)$ where $\Gamma_p = (\gamma_{i-j})$ and $\gamma_k = \text{Cov}(x_t, x_{t-k})$ may be written (Barndorff-Nielsen and Schou, 1973) as

$$g_p = \prod_{j=1}^p (1 - \zeta_j^2)^{-j}.$$

Hence the exact loglikelihood function (5.3.1) may now be written,

$$L(\phi, \sigma_a^2) = -\frac{n}{2} \text{Log}(\sigma_a^2) - \frac{1}{2} \text{Log}(g_p) - \frac{1}{2\sigma_a^2} S(\phi)$$

where $S(\phi) = \beta' D \beta$.

Champernowne, D.G. (1948), Sampling Theory Applied to Autoregressive Sequences. *Journal of the Royal Statistical Society B* 10, 204-242.

■ Concentrated Loglikelihood

Maximizing (5.3.1), $L(\phi, \sigma_a^2)$, over σ_a^2 , we obtain

$$\hat{\sigma}_a^2 = S(\phi) / n$$

and the profile loglikelihood for ϕ can be written,

$$L(\phi) = -\frac{n}{2} \text{Log}(S(\phi)/n) - \frac{1}{2} \text{Log}(g_p). \quad (6)$$

After the initial computation of the matrix D which only needs to be done once, each further evaluation of the likelihood $L(\phi)$ only requires $O(p^2)$ flops.

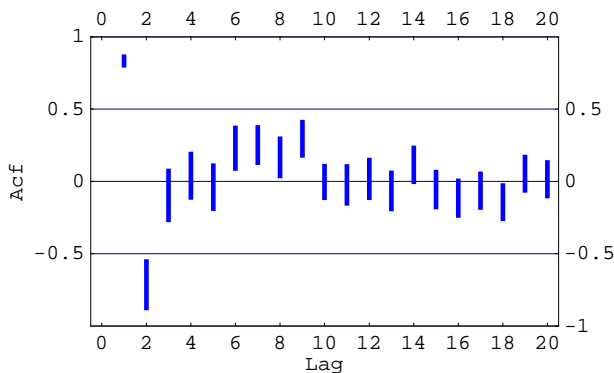
■ Reparameterized Concentrated Loglikelihood

To maximize $L(\phi)$, it is convenient to re-parameterize using the $\zeta = (\zeta_1, \dots, \zeta_p)$ parameters. Initial estimates via Burg algorithm.

$$L(\phi(\zeta)) = -\frac{n}{2} \text{Log}(S(\phi(\zeta))/n) - \frac{1}{2} \text{Log}(g_p). \quad (7)$$

■ Partial Autocorrelation Plot, Series A

Using the Burg or exact MLE we obtain $\hat{\zeta}_k = \hat{\phi}_{k,k}$ and then using Theorem 1 obtain $\text{EstSd}(\hat{\zeta}_k)$. Plot the 95% intervals $\hat{\zeta}_k \pm 1.96 \text{EstSd}(\hat{\zeta}_k)$.

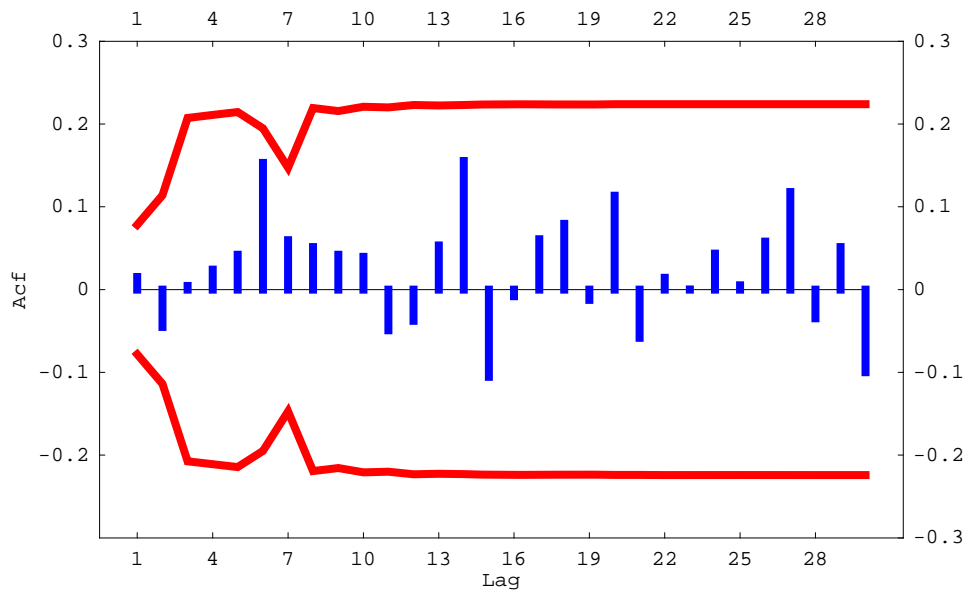


■ Fitted AR_ζ for Series A

	Parameter	SE	Z
ζ_1	0.507853	0.0764974	6.63882
ζ_2	0.201399	0.0722723	2.78667
ζ_6	0.162947	0.0728449	2.23691
ζ_7	0.197549	0.069843	2.82848
μ	17.0624	0.0830781	205.378
σ_a	0.307833		

■ Residual Autocorrelation Plot

As noted by Hosking and Ravishanker (1993) the Bonferonni Inequality may be used to obtain 5% simultaneous significance levels.



■ UBIC Model Selection

Chen J and Chen Z (2008, to appear). Extended Bayesian Information Criteria for Model Selection with Large Model Space. *Biometrika*.
<http://www.stat.ubc.ca/~jhchen/publications.html>

For $AR_\zeta(i_1, \dots, i_m)$, $\mathcal{L} = (-n/2) \log(\hat{\sigma}_a^2)$

Since $\hat{\sigma}_a^2 \approx c_0 \left(1 - \hat{\zeta}_{i_1}^2\right) \cdots \left(1 - \hat{\zeta}_{i_m}^2\right)$

$$\text{UBIC}(i_1, \dots, i_m) = n \log \prod_{k \in \{i_1, \dots, i_m\}} \left(1 - \hat{\zeta}_k^2\right) + m \log(n) + 2 \log \binom{P}{m} \quad (8)$$

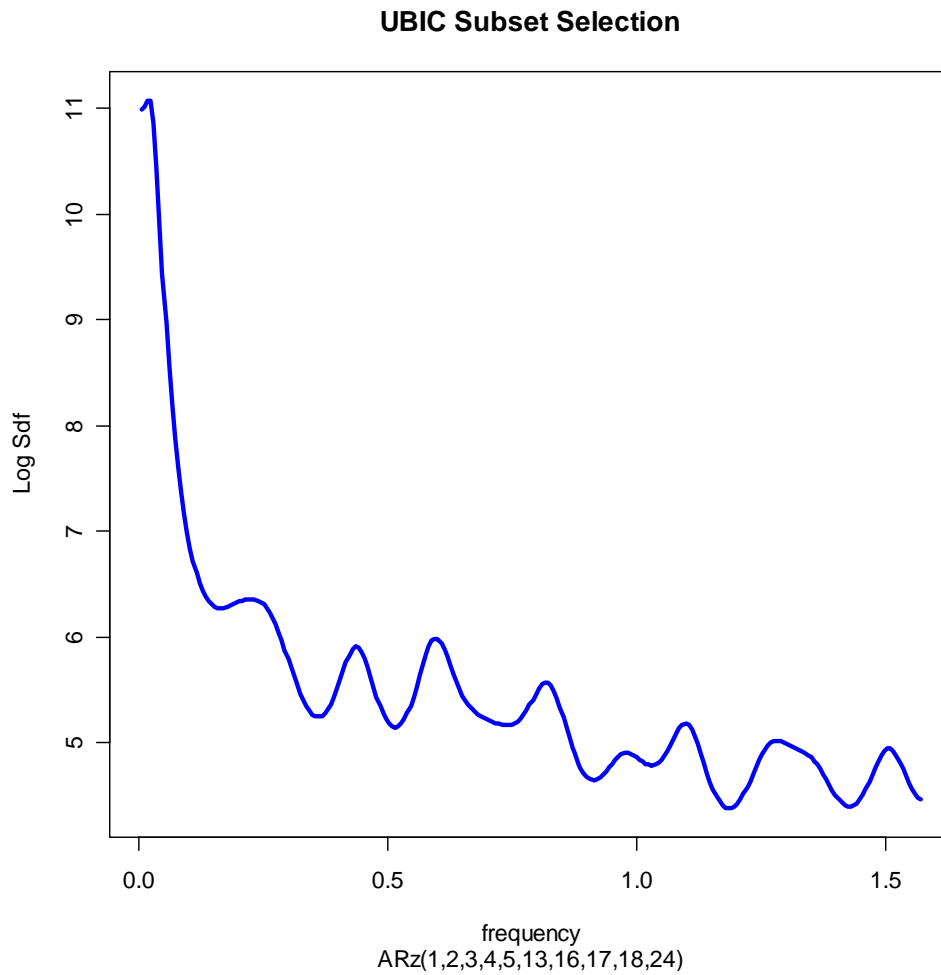
We don't need to search all subsets. Just arrange $\hat{\zeta}_k^2$ in ascending order and proceed with the evaluation of the UBIC.

■ Monthly Sunspot Series, 1749-1983, $n = 2820$

Taking $L = 200$ the best models were determined.

Model	IC	m	\mathcal{L}	AIC	BIC	UBIC
AR $_{\zeta}$	AIC	59	-7616.0	15 352.0	15 708.7	15 977.4
AR $_{\zeta}$	BIC	15	-7697.8	15 427.6	15 522.8	15 549.8
AR $_{\zeta}$	UBIC	10	-7724.6	15 471.3	15 536.7	15 566.1
AR	AIC	28	-7674.3	15 406.5	15 578.9	na
AR	BIC	18	-7702.0	15 442.1	15 555.0	na

The AR $_{\zeta}$ with UBIC still has better BIC than the best AR.



Faster ARMA MLE

McLeod AI and Zhang Y (2008). Faster ARMA Maximum Likelihood Estimation. *Computational Statistics and Data Analysis*, 52, 2166–2176.

R Package: FitARMA

■ AR Likelihood Approximation

Consider

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

where $a_t \sim \text{NID}(0, \sigma_a^2)$. Let $\alpha_1, \dots, \alpha_r$ be the coefficients in the best linear r -th order predictor which may be obtained by D-L for given $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$. By taking r sufficiently large, the Kullback-Leibler Divergence can be made small. In practice it suffices with $r = 30$.

■ Kullback-Leibler Divergence

Let $g(z)$ and $h(z)$ be two probability density functions. Then the Kullback-Leibler discrepancy or divergence of $g(z)$ from $h(z)$ is given by

$$I(g, h) = \mathbb{E} \left\{ \log \frac{g(z)}{h(z)} \right\} \quad (9)$$

The discrepancy between two normal distributions with mean zero and covariance matrices Σ_g and Σ_h

$$I(g, h) = \frac{1}{2} \left(\text{tr}(\Sigma_g \Sigma_h^{-1}) - \log(|\Sigma_g| / |\Sigma_h|) - n \right) \quad (10)$$

■ Exact MLE

Exact mle for μ given β

$$\hat{\mu} = \frac{\mathbf{1}'_n \Gamma_n^{-1} z}{\mathbf{1}'_n \Gamma_n^{-1} \mathbf{1}_n} \quad (11)$$

This would require $O(n^2)$ flops. The algorithm developed below solves this with $O(n)$ flops. We make use a symbolic method for computing the inverse covariance matrix of an ARMA process given by Zinde-Walsh (1988, *Econometric Theory*).

Zinde-Walsh, V. (1988). Some exact formulae for autoregressive moving average processes. *Econometric Theory*, 4, 384-402.

Efficiency of the Sample Mean

$$\frac{\text{Var}(\hat{\mu})}{\text{Var}(\bar{z})} = \frac{n^2}{(\mathbf{1}' \Gamma_n \mathbf{1}) (\mathbf{1}' \Gamma_n^{-1} \mathbf{1})} \quad (12)$$

	-0.975	-0.9	-0.5	0.	0.5	0.9	0.975
-0.975	1.000	1.000	0.980	0.840	0.367	0.015	0.001
-0.9	1.000	1.000	0.996	0.959	0.711	0.061	0.006
-0.5	0.999	0.999	1.000	0.997	0.950	0.298	0.034
0.	0.998	0.998	0.999	1.000	0.990	0.578	0.098
0.5	0.993	0.993	0.994	0.997	1.000	0.840	0.259
0.9	0.959	0.959	0.960	0.963	0.971	1.000	0.802
0.975	0.892	0.892	0.893	0.895	0.901	0.944	1.000

Table verified by simulation. Concluded software working.

Improved Portmanteau Test

Lin Jen-Wen and McLeod, A.I. (2006). Improved Pena-Rodriguez Portmanteau Test. *Computational Statistics and Data Analysis*, 51, 1731-1738.

Lin, Jen-Wen and McLeod (2008). Portmanteau Tests for ARMA Models with Infinite Variance. *Journal of Time Series Analysis*. To appear.

Our test statistic is based on the previous work of Peña and Rodríguez (2002).

Peña, D. and Rodríguez, J. (2002). A powerful portmanteau test of lack of fit for time series. *Journal of the American Statistical Association* 97, 601-610.

The residual autocorrelations,

$$r_k = \frac{\sum_{t=1}^{n-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^n \hat{a}_t^2}$$

for $k=1,2,\dots$, where $\hat{a}_t = \hat{\theta}^{-1}(B) \hat{\phi}(B) X_t$.

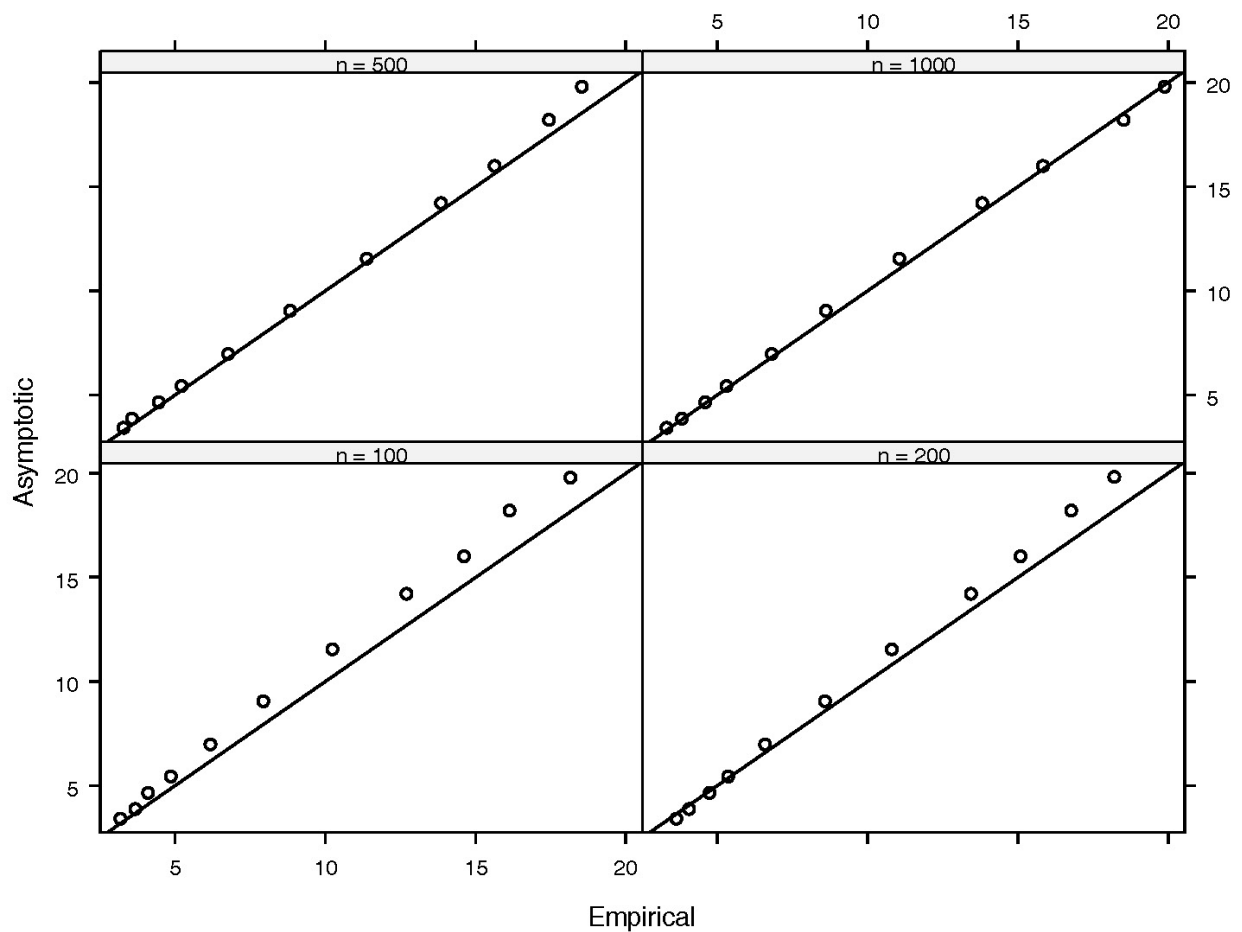
This definition ensures the nonnegative definiteness of the sample autocorrelation function (McLeod and Jiménez,1984).

$$\tilde{r}_k^2 = \frac{(n+2)r_k^2}{n-k} \quad \text{BAD IDEA!}$$

$$\hat{\mathbb{R}}_m = \begin{pmatrix} 1 & \hat{r}_1 & \cdots & \hat{r}_m \\ \hat{r}_1 & 1 & \cdots & \hat{r}_{m-1} \\ \vdots & \cdots & \ddots & \vdots \\ \hat{r}_m & \cdots & \hat{r}_1 & 1 \end{pmatrix}$$

$$\hat{D} = n - n \cdot (\text{Det } \hat{\mathbb{R}}_m)^{1/m}$$

Rather work with approximations to the asymptotic distribution as in Peña and Rodríguez (2002), we computed the "exact" asymptotic distribution by method of (Imphov, 1961). By simulation, we found the small distribution did not agree very well with the asymptotic even for unless n was large.



■ Monte - Carlo Test Procedure

1. After fitting model, obtain \hat{D}_{obs}
2. Select the number of Monte-Carlo simulations, N . Typically $100 \leq N \leq 1000$.
3. Simulate the model using the estimated parameters obtained in Step 1 and obtain \hat{D} for the simulated series.
4. Repeat Step 3, N times and obtain $k = \# \{ \hat{D} \geq \hat{D}_{\text{obs}} \}$.
5. The P-value for the test is $(k + 1)/(N + 1)$.

Even with $N = 1000$ this algorithm usually takes less than a minute when $n \leq 200$ and the ARMA model is not complicated. For ARMA case, an R package, **PRTest**, is available from:
<http://www.stats.uwo.ca/faculty/aim/2007/LinMcLeod/>

Medical Time Series

■ Introduction to Medical Time Series

Matthews (1998) pointed out that many clinical trials give overly optimistic results. Since Matthews (1998) it has been found that indeed many clinical trials were indeed flawed but not because of the statistical methods. One major problem in many clinical trials has to do with removing cases where the drug does not work or perform well. This cheating has been hard to detect in the past since the complete data were not made available for public scrutiny. Some major medical journals have announced that new standards will be put in place to audit the clinical trial databases and the U.S. Congress is discussing whether government regulation may be required.

It may happen that a new medical procedure such as heart-valve surgery or a new prescription medicine is really clinically important and useful. But then it can be mis-applied. In the U.S., for example, it is well known that surgeons perform far more heart-valve surgery than is warranted because the operation is so lucrative. Another possibility is that the circumstances under which a new medical procedure is beneficial are complicated and practicing physicians may err in prescribing the medicine and in some situations this may lead to increased morbidity and mortality as in the case study reported by Juurlink et al. (2004).

The new field of study, medical time series, examines how changes in medical procedures affect globally the whole relevant population after they have been implemented. This method is a very important adjunct to clinical trials.

■ References

Matthews, Robert (1998). *The Great Health Hoax*. Copyright 1998 THE SUNDAY TELEGRAPH (UK). Available:
<http://www.sepp.org/Archive/contro/controversies/healthhoax.html>

Juurlink, David N., Mamdani, Muhammad M., Lee, Douglas S., Kopp, Alexander, Austin, Peter C., Laupacis, Andreas, Redelmeier, Donald A. (2004). Rates of Hyperkalemia after Publication of the Randomized Aldactone Evaluation Study *N Engl J Med* 2004 351: 543-551.

■ Intervention Analysis

Introduced in a paper by Box and Tiao (1976). Discussed in Hipel and McLeod (1994). Software freely available by McLeod and Hipel (2008). Power computation discussed by McLeod and Vingilis (2005).

$$z_t = \mu + I_t + N_t,$$

where μ denotes the mean level, the term I_t denotes the intervention effect and N_t which is the error term is modeled using a suitable ARMA(p,q) time series model,

$$N_t = \phi_1 N_{t-1} + \dots + \phi_p N_{t-p} + E_t - \theta_1 E_{t-1} + \dots + \theta_q E_{t-q},$$

where E_t denotes an independent error sequence. The simplest, which corresponds to the t -test in a non-time series setting, is the Step Intervention,

$$I_t = \omega_0 \mathfrak{S}_t^{(T)}$$

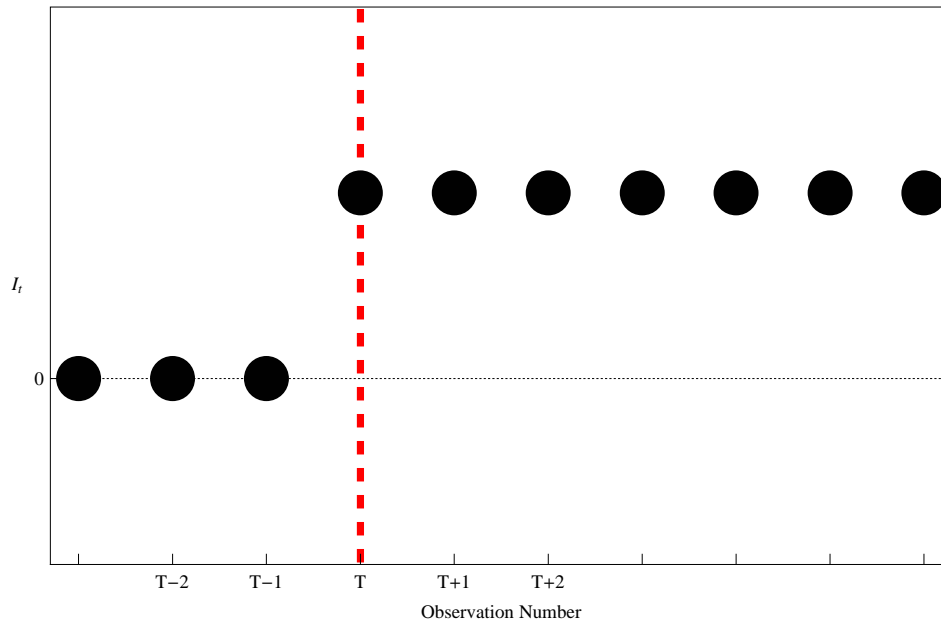
where $\mathfrak{S}_t^{(T)}$ is a step function,

$$\mathfrak{S}_t^{(T)} = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases}$$

The parameter ω_0 measures the change caused by the intervention and is estimated along with the ARMA time series component. The estimation procedure provides an estimate of ω and a confidence interval for the parameter.

■ Figure 1. Simple Step Intervention.

$$I_t = \omega_0 \mathfrak{S}_t^{(T)}$$



■ References

Hipel, K.W. and McLeod, A.I. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier. Now available online: <http://www.stats.uwo.ca/faculty/aim/1994Book/default.htm>

McLeod-Hipel Time Series Package (2008). <http://www.stats.uwo.ca/faculty/aim/epubs/mhts/default.htm>

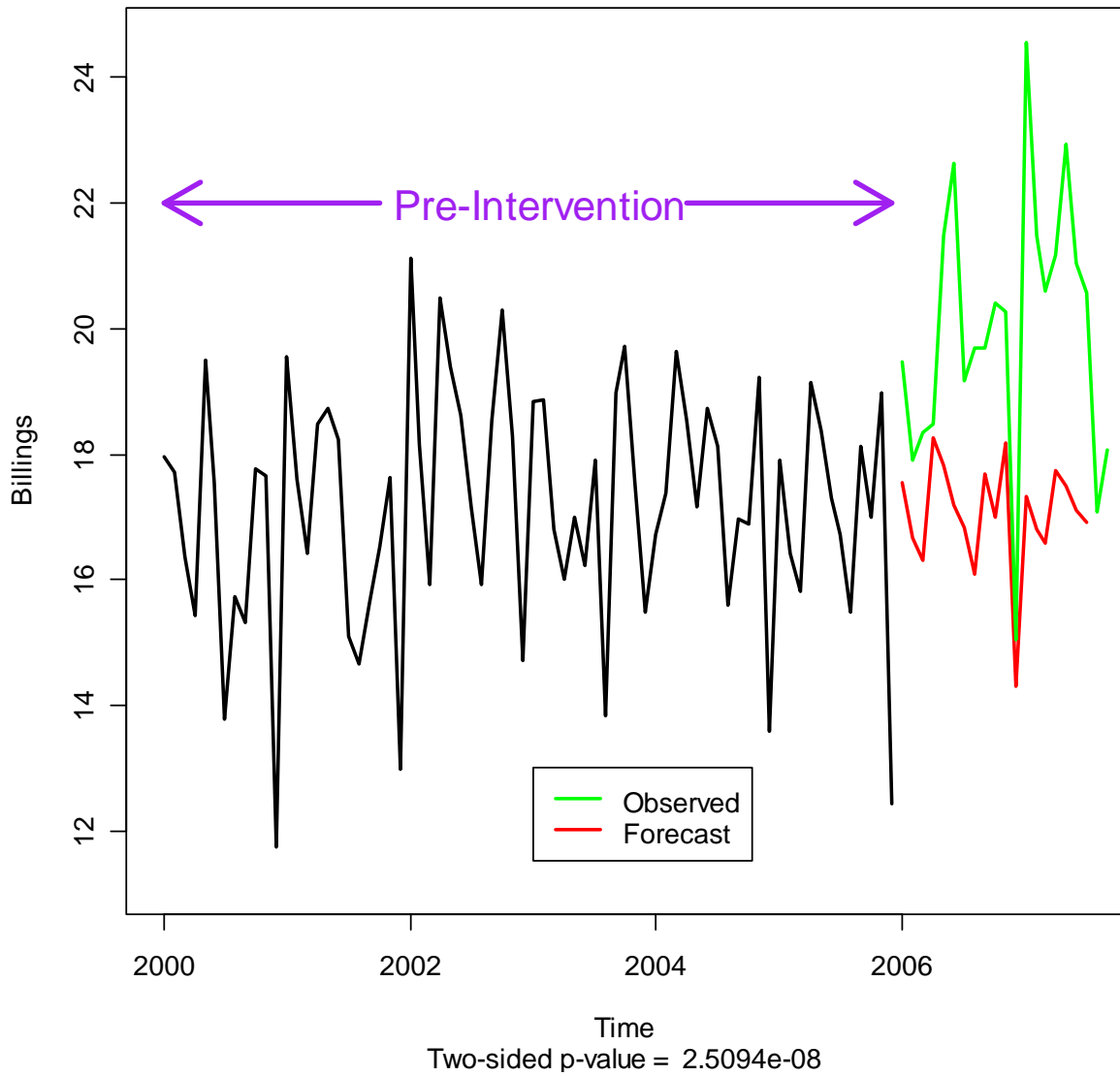
McLeod, A.I. and Vingilis, E.R. (2005), Power Computations for Intervention Analysis, *Technometrics*. 47, 174-180.

■ EGFR Reporting

Joint Research with Drs. Armit Garg and Arsh Jain

A change occurred in medical recommendations which was expected to change the referral rate to kidney specialists (nephrologists). The change in referrals was expected to occur starting January 2006. A simple step intervention model was fit and the change was found to significant on a two-sided test at around 10^{-8} . As shown in the plot below the change appears quite substantial.

EGFR Reporting: Column O



Burg Estimator

Zhang, Y. and McLeod, A.I. (2005), Computer Algebra Derivation of the Bias of Burg Estimators. *Journal of Time Series Analysis*.

Tjöstheim, D. & Paulsen, J. (1983), Bias of some commonly-used time series estimates. *Biometrika* 70, 389--399. Correction *Biometrika* 71, p. 656.

Percival, D. and Walder, A. T. (1993). *Spectral Analysis For Physical Applications*. Cambridge, University Press

For AR(1) the Burg estimator is

$$\hat{\phi} = \frac{2 \sum_{t=2}^n z_t z_{t-1}}{\sum_{t=2}^n (z_t^2 + z_{t-1}^2)}$$

and in general the Burg estimator is a linear estimator, that is, it is obtained by solving linear equations involving sums of cross-products.

Using computer algebra, Zhang and McLeod (2005) showed that the first-order asymptotic bias of the Burg estimator is the same as that for least-squares for AR(p), $p = 1, 2, 3$. Used Taylor series linearization to express estimates as linear function of sums of cross products.