TESTING FOR WHITE NOISE IN
TIME SERIES MODELS

Ph.H. FRANSES

REPORT 9043/A

TESTING FOR WHITE NOISE IN TIME SERIES MODELS

by

Philip Hans FRANSES

Econometric Institute, Erasmus University Rotterdam
P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands

ABSTRACT

The new pure significance test for white noise proposed in the present paper is based on the estimated R^2 of an ARMA model fitted to residuals. A small empirical size and power investigation is carried out, and the latter seems to indicate that this test meets its purpose more than the portmanteau test.

CONTENTS

ACKNOWLEDGEMENTS

1. INTRODUCTION

Define a white noise process $\{\varepsilon_t\}$ as a sequence of uncorrelated variables with zero mean and constant variance, i.e. $E\{\varepsilon_t\}=0$, $E\{\varepsilon_t^2\}=\sigma_\varepsilon^2$ and $E\{\varepsilon_s\varepsilon_t\}=0$, for $s\neq t$. Consider the class of linear models

$$y_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u} \qquad (1.1)$$

where $\{y_t\}$ is a discrete time series with zero mean, $t=1,2,..,n$, and g is a linear function which transforms the white noise process into the series y_t. Define the backward shift operator B by $B^k y_t = y_{t-k}$, and with the use of

$$G(z) = \sum_{u=0}^{\infty} g_u z^u$$

and that $|z|<1$, where z might be complex valued, (1.1) can be written as

$$y_t = G(B)\varepsilon_t \qquad (1.2)$$

This model however involves an infinite number of parameters, so to fit it to data additional assumptions have to be made. One such assumption is that $G(z)$ can be approximated by a rational function of the form

$$G(z) = (1 + \alpha_1 z + ... + \alpha_q z^q)/(1 + \theta_1 z + ... + \theta_p z^p)$$

which reduces (1.2) to a mixed autoregressive moving average model of order (p,q), or ARMA(p,q), to be denoted as

$$\theta_p(B)y_t = \alpha_q(B)\varepsilon_t \qquad (1.3)$$

Suppose one fits (1.3) to real data and one obtains estimated residuals u_t, then it seems natural to check the assumption of white noise. Other inadequacies, such as significant correlation between u_t and y_{t-i} for some

1

i, and nonlinearity in the residuals, are also testable features. Most tests however make use of a white noise error process assumption. Although constant variance in u_t should be tested too, the focus in this paper is on testing the hypothesis of $E\{u_s u_t\}=0$ for $s \neq t$.

A well-known portmanteau statistic has been developed by Box and Pierce [1970]. It is a pure significance test in the sense that there is no explicit formulation of an alternative hypothesis (see for a survey Godfrey and Tremayne [1988]). Denoting the τ-th sample autocorrelation as

$$r_\tau(u) = \sum_{t=\tau+1}^{n} u_t u_{t-\tau} \Big/ \sum_{t=1}^{n} u_t^2 \tag{1.4}$$

then the Box–Pierce test (BP) is defined as

$$BP(m) = n.\sum_{\tau=1}^{m} r_\tau^2(u) \tag{1.5}$$

which is asymptotically $\chi^2(m-p-q)$ distributed if m/n is small and m is moderately large, often taken to be 20 or 30. In Davies, Triggs and Newbold [1977] it has been shown that the actual fitted sample distribution of BP can differ from the predicted χ^2 distribution, and hence the BP has been modified in Ljung and Box [1978] to the often applied portmanteau statistic

$$LB(m) = n(n+2) \sum_{\tau=1}^{m} (n-\tau)^{-1} r_\tau^2(u) \tag{1.6}$$

which has the same asymptotic distribution as the BP test statistic. Both test statistics have been extensively used in practice, and also several power investigations have been carried out (cf. Davies and Newbold [1977], Clarke and Godolphin [1982], Hall and McAleer [1989]). From these studies it emerged that both test statistics may lack power, although the LB test is not always performing badly. Because of these results several new tests have been proposed.

Nested and nonnested hypothesis tests, as proposed in Godfrey [1979] and McAleer et al. [1988] respectively, specify a specific alternative model of type (1.3) and test whether the residuals indicate a possible way to modify the originally specified model. As expected, the empirical powers of these tests can be higher than the portmanteau tests, although the sizes vary across the different generating processes (Hall and McAleer [1989]). In some cases the Lagrange Multiplier (LM) test, advocated in e.g. Godfrey [1979], can be calculated as n times the R^2 of an auxiliary regression, which depends on the alternative hypothesis. Furthermore, it has been demonstrated that often the LM test and the BP statistic are equivalent (Newbold [1980]). A modification to the portmanteau test statistic has been proposed by Godolphin [1980]. This test is however not easy to calculate, although its power seems to be rather satisfying (see Clarke and Godolphin [1982]).

Yet another approach is to fit ARMA(p_u, q_u) models to the residuals u_t, and see whether p_u and/or q_u are unequal to zero, and thus rejecting white noise. In Pukkila and Krishnaiah [1988] several autoregressive order determination criteria, such as the AIC and BIC, are used to test the hypotheses that u_t is generated by an AR(0) process versus an AR(k) process. If the criteria are consistent, and the process u_t follows a white noise process, then the AR(0) model should be chosen. The power of this test seems to be promising, although further research might be needed.

The new pure significance test for white noise proposed in this paper is based on the R^2 of an ARMA model fitted to residuals, and it will be given in the following section. In section 3, a small size and power investigation will be carried out. In section 4, conclusions and suggestions for further research will be given.

3

2. A NEW TEST

Define the squared multiple correlation coefficient R^2 as the proportion of the variability of y_t explained by a stationary invertible ARMA model as in (1.3) as

$$R^2 = 1 - \sum_t \varepsilon_t^2 / \sum_t y_t^2$$

which is estimated by

$$\hat{R}^2 = 1 - \sum_t u_t^2 / \sum_t y_t^2$$

where $\{u_t\}$ is the sequence of estimated residuals. In Hosking [1979] it is derived that this $\hat{R}^2$ is asymptotically distributed as

$$\hat{R}^2 \sim N\{R^2, \; 4(\sigma_\varepsilon^2/\sigma_y^2)\sum_{\tau=1}^{\infty} \rho_\tau^2/n\} \tag{2.1}$$

where $\sigma_y^2 = \text{var}(y_t)$, and $\rho_\tau = E(y_t y_{t-\tau})/E(y_t^2)$.

This expression gives an opportunity to construct a test for eventual white noise properties of $\{u_t\}$. In case $\{u_t\}$ is not white noise, there is some information in past values of u_t which might explain present u_t, and hence one can fit an ARMA(p_u, q_u) model to $\{u_t\}$ with p_u and/or q_u unequal to zero. Defining the squared multiple correlation coefficients R_u^2 and $\hat{R}_u^2$ of that model as

$$R_u^2 = 1 - \sum_t \varepsilon_t^2 / \sum_t u_t^2$$

and

$$\hat{R}_u^2 = 1 - \sum_t \hat{\varepsilon}_t^2 / \sum_t u_t^2$$

where $\{\hat{\varepsilon}_t\}$ is the sequence of estimated residuals, it is clear that the

4

estimated $\hat{R}_u^2$ is asymptotically distributed as

$$\hat{R}_u^2 \sim N\{R_u^2,\ 4(\sigma_\varepsilon^2/\sigma_u^2)\textstyle\sum_{\tau=1}^{\infty}\rho_\tau^2(u)/n\} \tag{2.2}$$

where $\sigma_u^2=\mathrm{var}(u_t)$, and $\rho_\tau(u)=E(u_t u_{t-\tau})/E(u_t^2)$. In case $\{u_t\}$ is white noise, the proportion of the variability of u_t explained by the model would be zero, so a natural null hypothesis is

$$R_u^2 = 1 - \sigma_\varepsilon^2/\sigma_u^2 = 0$$

Under this H_0 the expression in (2.2) reduces to

$$\hat{R}_u^2 \sim N\{0,\ 4\textstyle\sum_{\tau=1}^{\infty}\rho_\tau^2(u)/n\} \tag{2.3}$$

Estimating the $\rho_\tau(u)$ by $r_\tau(u)$ (see (1.4)), and replacing ∞ by m, (2.3) gives the test statistic

$$Q = \tfrac{1}{2}\, n\hat{R}_u^2/\sqrt{(n\textstyle\sum_{\tau=1}^{m} r_\tau^2(u))} \tag{2.4}$$

which has an asymptotic $N(0,1)$ distribution.

Several features of this test Q immediately become clear, such as that the denominator in (2.4) is the convential BP test statistic as in (1.5). Secondly, in the BP test one only has to choose the m, but for Q an additional choice of (p_u,q_u) has to be made. One can argue that this choice should be such that the $\hat{R}_u^2$ is maximized. However, in the statistic Q the numerator and the denominator are not independent, and this might seriously bias the empirical distribution. So, it is decided to let empirical size investigations determine an optimal combination of the free parameters in (2.4). Note that the Q-test is a one-sided test, so the 5% and 10% critical values are 1.645 and 1.282, respectively.

5

3. SIZE AND POWER

In the following experiments p_u is taken to be 5 and q_u is set equal to 0. The number of observations is restricted to 100 and 200, and all experiments are based on 1000 replications. Consider first table 1, where the empirical size of the test Q is calculated for some values of m, by fitting AR(5) models to observations drawn from a N(0,1) distribution.

Table 1. Empirical size versus nominal size of 5% and 10%,

nominal size	n	$m=5$	$m=8$	$m=10$	$m=20$
0.05	100	.078	.028	.015	.001
	200	.062	.023	.015	.002
0.10	100	.300	.172	.112	.021
	200	.265	.141	.096	.024

From this table it can be seen that the empirical size often underestimates the nominal size of 5% , but that for $m=10$ the nominal size of 10% is well approximated. Hence, in the following power experiments the test Q is used with $p_u=5$, $q_u=0$ and $m=10$.

These power investigations have been done within the framework used in Davies and Newbold [1977] and Clarke and Godolphin [1982], i.e. some ARMA(2,2) processes have been generated for y_t, after which AR(1) or AR(4) models have been fitted. The residuals of these misspecified models are tested for serial dependence with the Q test above, and also with the *LB* test to make direct comparison of the powers possible.

Table 2. Empirical power of the Q and LB test, n=100

Series[1]	Parameter values[2]				A[3]	Q[4]		LB[5]	
	θ_1	θ_2	α_1	α_2	p	5%	10%	5%	10%
1	0	0	0.2	−0.4	4	.004	.016	.087	.134
2	−0.3	0	−0.75	0	4	.003	.040	.095	.152
3	−0.9	0	0.25	0	1	.186	**.428**	.239	.340
4	0	0	0.6	0.4	1	.130	**.417**	.268	.366
5	−0.8	0	−0.2	−0.4	1	.207	**.491**	.251	.351
6	−0.8	0.4	0.8	0	4	.149	**.356**	.196	.301
8	−0.4	0	0.2	0.4	1	.347	**.641**	.397	.511
9	0	0	0.2	−0.4	1	.798	**.929**	.601	.720
10	0	0	0.9	0.8	1	.897	**.970**	.783	.870
11	−0.6	0	0.75	0	1	.997	**1.00**	.842	.911

[1]   The number of the series corresponds to the models as displayed in table 1 and 2 in Clarke and Godolphin [1982]. Series 7 has not been used for the practical problems they were confronted with (ibid,p.149).

[2]   True model is

$$y_t + \theta_1 y_{t-1} + \theta_2 y_{t-2} = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}$$

with $\varepsilon_t \sim N(0,1)$, $y_0=0$ and the first 50 observations are deleted.

[3]   An autoregressive model of order p is fitted to the observations generated by the true model (see footnote 2).

[4]   The Q-test is a one-sided test, so the 5% and 10% critical values are taken to be 1.645 and 1.282, respectively. Bold values indicate that for that case the power of Q exceeds or is equal to the power of LB.

[5]   The LB test statistic is calculated for $m=20$, because for this value the size of the test seems reasonable (cf. Clarke and Godolphin [1982]).

Table 3. Empirical power of the Q and LB test, n=200

Series[1]	Parameter values[2] θ_1	θ_2	α_1	α_2	A[3] p	Q[4] 5%	10%	LB[5] 5%	10%
1	0	0	0.2	-0.4	4	.003	.025	.126	.194
2	-0.3	0	-0.75	0	4	.014	.076	.127	.201
3	-0.9	0	0.25	0	1	.446	**.746**	.448	.568
4	0	0	0.6	0.4	1	**.463**	**.738**	.454	.592
5	-0.8	0	-0.2	-0.4	1	**.530**	**.801**	.450	.558
6	-0.8	0.4	0.8	0	4	**.481**	**.735**	.449	.578
8	-0.4	0	0.2	0.4	1	**.787**	**.933**	.703	.817
9	0	0	0.2	-0.4	1	**.993**	**1.00**	.926	.965
10	0	0	0.9	0.8	1	**1.00**	**1.00**	.999	1.00
11	-0.6	0	0.75	0	1	**1.00**	**1.00**	1.00	1.00

[1] The number of the series corresponds to the models as displayed in table 1 and 2 in Clarke and Godolphin [1982]. Series 7 has not been used for the practical problems they were confronted with (ibid,p.149).

[2] True model is

$$y_t + \theta_1 y_{t-1} + \theta_2 y_{t-2} = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}$$

with $\varepsilon_t \sim N(0,1)$, $y_0=0$ and the first 50 observations are deleted.

[3] An autoregressive model of order p is fitted to the observations generated by the true model (see footnote 2).

[4] The Q-test is a one-sided test, so the 5% and 10% critical values are taken to be 1.645 and 1.282, respectively. Bold values indicate that for that case the power of Q exceeds or is equal to the power of LB.

[5] The LB test statistic is calculated for $m=20$, because for this value the size of the test seems reasonable (cf. Clarke and Godolphin [1982]).

From these tables it can be seen that often the power of the Q test exceeds or equals the power of the Ljung–Box test, especially in case $n=200$. However, for series 1 and 2 the power is extremely small. Writing the ARMA(2,2) models as autoregressive models and substituting the parameter values for series 1 and 2, results in models which are very close to AR(4) models, and hence low powers might not be unexpected.

4. DISCUSSION

The new pure significance test for white noise developed in this paper seems to meet its purpose, i.e. in a small experiment it often behaves more powerful than the well–known portmanteau statistic. Another feature of the test is that only one additional model has to be fitted, and that calculations can be done with most standard statistical programs. Furthermore, the test seems to be applicable to all kinds of residual processes, i.e. not only to those of univariate ARMA time series models. A further research topic is to look for values of p_u and q_u, which give reasonable size for m other than 10, and in which the test might be even more powerful. Another topic is to compare the test in power experiments with the approach advocated in Pukkila and Krishnaiah [1988].

9

REFERENCES

Box, G.E.P. and D.A. Pierce [1970], Distribution of residual autocorrelations in autoregressive integrated moving average time series models, Journal of the American Statitical Association, 65, 1509–1526.

Clarke, B.R. and E.J. Godolphin [1982], Comparative power studies for goodness of fit tests of time series models, Journal of Time Series Analysis, 3, 141–151.

Davies, N. and P. Newbold [1979], Some power studies of a portmanteau test of time series model specification, Biometrika, 66, 153–155.

Davies, N. C.M. Triggs and P. Newbold [1977], Significance levels of the Box–Pierce portmanteau statistic in finite samples, Biometrika, 64, 517–522.

Godolphin, E.J. [1980], A method for testing the order of an autoregressive moving average process, Biometrika, 67, 699–703.

Godfrey, L.G. [1979], Testing the adequacy of a time series model, Biometrika, 66, 67–72.

Godfrey, L.G. and A.R. Tremayne [1988], Checks of model adequacy for univariate time series models and their application to econometric relationships, Econometric Reviews, 7, 1–42.

Hall, A.D. and M. McAleer [1989], A Monte Carlo study of some tests of model adequacy in time series analysis, Journal of Business & Economic Statistics, 7, 95–106.

Hosking, J.R.M. [1979], The asymptotic distribution of R^2 for autoregressive-moving average time series models when parameters are estimated, Biometrika, 66, 157–158.

Ljung, G.M. and G.E.P. Box [1978], On a measure of lack of fit in time series models, Biometrika, 65, 297–303.

McAleer, M., McKenzie, C.R. and A.D. Hall [1988], Testing separate time series models, Journal of Time Series Analysis, 9, 169–189.

Newbold, P. [1980], The equivalence of two tests of time series model adequacy, Biometrika, 67, 463–465.

Pukkila, T.M. and P.R. Krishnaiah [1988], On the use of autoregressive order determination criteria in univariate white noise tests, IEEE Transactions on Acoustics, Speech, and Signal Processing, 36, 764–774.