24 June 2003

A General Test for Time Dependence in Parameters

Ralf Becker
Queensland University of Technology

Walter Enders*
University of Alabama

Stan Hurn
Queensland University of Technology

Abstract

A new test for time-dependent parameters is proposed. The *Trig*-test is based on a trigonometric expansion to approximate the unknown functional form of the variation in the parameters concerned. It is shown to have the correct empirical size and excellent power to detect structural breaks and stochastic parameter variation. The appropriate use of the *Trig*-test is demonstrated by testing for structural breaks in the U.S. inflation rate. The test detects a statistically significant increase in the U.S. inflation rate beginning in the early 1970s and lasting through to the early 1980s.

Keywords: Time-varying parameters, Fourier-series approximation, Nuisance parameters, Bootstrap, Empirical size and power

JEL Classification: C51, C52, G12

* **Corresponding author**: Department of Economics, Finance & Legal Studies, University of Alabama, Tuscalooosa, AL 35487, wenders@cba.ua.edu.

In this paper, we propose a simple, yet powerful, test for time dependence in parameters. The null hypothesis is a linear model with time-invariant parameters; under the alternative hypothesis, the model is linear in variables but has time-varying coefficients. The central idea of the test is to use a trigonometric expansion to approximate the unknown functional form of a time-varying regression coefficient. The flexibility of this approximation means that the test is capable of detecting time-variation in coefficients that arise from a number of different sources, namely, structural breaks, seasonality or stochastic variation. Rigorous implementation of the testing strategy relies on bootstrapping to deal with the presence of an unidentified parameter under the null hypothesis. A version of the test based on OLS is also available that is very simple to implement but remains effective in detecting time variation in parameters.

1. A test based on a trigonometric approximation

Consider a model that is linear in the weakly stationary variable w_t but has a time-varying coefficient $b(t)$:

$$y_t = w_t b(t) + e_t \; ; \qquad e_t \sim N\left(0, s_e^2\right). \tag{1}$$

Since any absolutely-integrable function $\hat{a}(t)$ can be approximated by means of a Fourier-series expansion of appropriate order, we parameterize equation (1) as:

$$y_t = w_t \left[b_0 + b_1 \sin\left(\frac{2fpt}{T}\right) + b_2 \cos\left(\frac{2fpt}{T}\right) \right] + e_t \tag{2}$$

where b_0, b_1 and b_2 are constant parameters and f is the frequency. Although many frequencies may be needed to mimic the actual behavior of the time-varying coefficient, for testing purposes, it seems reasonable to approximate the expansion by only one frequency. After all, if $\hat{a}_1 = \hat{a}_2 = 0$ cannot be rejected for all frequencies, the null hypothesis of time invariance is rejected. The proposed test, which we call the *Trig*-test, is the test of the null hypothesis $b_1 = b_2 = 0$ by means of a likelihood-ratio test, $LR(f)$.

The implementation of the *Trig*-test is not entirely straightforward because the frequency, f, is unidentified under the null hypothesis. As in Davies (1987), Andrews and Ploberger (1994), and Stinchcombe and White (1998), the appropriate way to proceed is to consider the discrete set Γ containing G frequencies as its elements and the related G tests of the null hypothesis, H_0: $b_1 = b_2 = 0$. Each test is constructed on the basis of a different frequency $f_i \in \Gamma$, $i = 1, \ldots G$. The information in these G values may then be used to compute the sup-norm (LR_{sup}) , unweighted average (LR_{ave}) and exponentially weighted average (LR_{exp}) tests suggested by Andrews and Ploberger (1994):

$$LR_{\text{sup}} = \sup_{f_i \in \Gamma} LR(f_i);\; LR_{\text{ave}} = \frac{1}{G}\sum_{f_i \in \Gamma} LR(f_i);\text{ and } LR_{\text{exp}} = \ln\left[\frac{1}{G}\sum_{f_i \in \Gamma}\exp\left(\frac{LR(f_i)}{2}\right)\right] \tag{3}$$

Obtaining critical values for these test statistics is non-trivial and bootstrapping using the approach suggested by Hansen (1999) is used here. *First*, J replications of the data, y_t^*, may be generated using the general scheme:

$$y_t^* = w_t \hat{b}_0 + e_t^* \tag{4}$$

where e_t^* is resampled (with replacement) from the empirical distribution of the rescaled and centered residuals obtained by estimating the model under the null hypothesis. Should w_t contain lags of the dependent variable, as is the case in both the simulations and the empirical examples to follow, the realizations, y_t^*, will need to be generated recursively. *Second*, for each bootstrap sample, y_t^*, a test statistic LR^j is computed. The proportion of the J bootstrapped test statistics, which exceed the LR test statistic computed from the observed data, is then an estimate of the p-value of the test. The empirical results obtained in the Monte Carlo exercises provided in the next section justify the conjecture that this bootstrapping procedure produces consistent inference.

There is a special case where bootstrapping is not required. Davies (1987) calculates the asymptotic distribution for the null hypothesis $b_1 = b_2 = 0$ when w_t is a vector of 1's, i.e, when: $y_t = b_0 + b_1\sin(2f\pi t/T) + b_2\cos(2f\pi t/T) + e_t$. As shown by Davies (1987), the critical values depend only on the range of frequencies used in the test. This result suggests that the *Trig*-test can be implemented by OLS as follows. Estimate (2) by OLS for each value of f in Γ. Let f* be the frequency which yields the smallest residual sum of squares, *RSS** and let b_0^*, b_1^* and b_2^* be the coefficients associated with this frequency. The *F*-test, F_{trig}^{OLS} of $b_1^* = b_2^* = 0$ is given by:

$$F_{trig}^{OLS} = \frac{(RSS_r - RSS^*)/2}{RSS^*/(T - k + 1)} \tag{5}$$

where RSS_r is the residual sum of squares with the restriction imposed. Although the distribution of the test statistic is invariant under the null hypothesis, it does not follow a standard *F*-distribution. Davies (1987) provides a useful approximation to the asymptotic distribution, but it is also straightforward to tabulate critical values by simulation as demonstrated by Ludlow and Enders (2000).

2. Empirical performance of the test

Although the test is simple to implement, its usefulness will depend on its empirical size and power. While optimal tests may be available if the reason for the time-varying coefficients is known, the *Trig*-test will be shown to have the correct size and good power properties when the applied researcher has no information about the nature of the time variation.

2.1 Empirical size

Consider the two linear autoregressive processes given by

$$\begin{aligned} y_t &= 0.6y_{t-1} + e_t \quad e_t \sim N(0,1), \\ y_t &= 0.9y_{t-4} + e_t \quad e_t \sim N(0,0.04). \end{aligned} \tag{6}$$

The lengths of the simulated processes are T = 50, 100 and 200, respectively. The autoregressive coefficients of both these models and the constant of the AR[4] model are tested for time invariance by means of the *Trig*-test. All three LR tests are applied and the choice of frequencies is given by Γ = [1, (T/2)-1]. The results reported in Table 1 show that the *Trig*-test has the correct empirical size regardless of the mapping utilized and the length of the process (as one would expect for a bootstrapped significance level).

2.2 Power

The power of the test to detect seasonal parameter instability is not reported. Clearly the *Trig*-test has excellent power in this case because inclusion of the relevant trigonometric terms is identical to the use of seasonal dummy variables. In this section, therefore, the focus is on *Trig*-test's power to detect structural breaks and stochastic parameter variation.

To investigate the power of the test to detect structural breaks, the example provided by Clements and Hendry (1999)

$$y_t = a_0 + a_{1t}x_t + e_t\,;\; e_t \sim N(0, 1) \tag{7}$$

is used, with six different specifications for α_{1t}:

$$SB1:\; a_{1t} = \begin{bmatrix} 1 & t \le 40 \\ 1.5 & t > 40 \end{bmatrix} \qquad SB2:\; a_{1t} = \begin{bmatrix} 1 & t \le 50 \\ 1.5 & t > 50 \end{bmatrix}$$

$$SB3:\; a_{1t} = \begin{bmatrix} 1 & t \le 20, t > 40 \\ 1.5 & 20 < t \le 40 \end{bmatrix} \qquad SB4:\; a_{1t} = \begin{bmatrix} 1 & t \le 40, t > 55 \\ 1.5 & 40 < t \le 55 \end{bmatrix}$$

$$SB5:\; a_{1t} = \begin{bmatrix} 1 & t \le 20 \\ 1.5 & 20 < t \le 40 \\ 0.5 & t > 40 \end{bmatrix} \qquad SB6:\; a_{1t} = \begin{bmatrix} 1 & t \le 40 \\ 1.5 & 40 < t \le 55 \\ 0.5 & t > 55 \end{bmatrix}$$

In all the models, a_0, is set to zero, the values for x_t are drawn from a normal distribution with mean and variance equal to unity and a sample size T = 60.

The power of the *Trig*-test will be evaluated with reference to the optimal test for one breakpoint, proposed by Andrews, Lee and Ploberger (ALP) (1996), and the test for multiple breakpoints due to Bai and Perron (BP) (1998, 2003). The ALP test examines all possible breakpoints occurring within the middle 90% of the data [$4 \le t \le 56$]. BP propose a number of tests for multiple structural breaks. Since the *Trig*-test is designed for use when the number of breaks under the alternative hypothesis is unknown, its power is compared to that of the *UDmax* and the *WDmax* tests.[1] The *Trig*-test includes frequencies in the range [(1/512), 6] in steps of 1/512. We use a maximum frequency of 6 since structural breaks are likely to be a low frequency phenomenon. Although all three mappings of the *LR* tests were computed, the performance of each was quite similar; as a result only those for LR_{exp} are reported.

The results reported in Table 2 confirm that the power of tests for structural breaks deteriorates, *ceteris paribus*, as the breakpoint moves towards the end of the sample. In the single breakpoint model, the ALP, BP and *Trig*-tests experience severe reduction in power as the breakpoint shifts from observation 40 to 50. Also note the relatively poor performance of the CUSUM and CUSUM^2 tests relative to the ALP, BP and *Trig*-tests. When there is only one breakpoint (models SB1 and SB2), the ALP test has a slight advantage over the *Trig*-test and a moderate advantage over the BP tests. Differences arise for the two-breakpoint models (SB3 to SB6). The *Trig*-test performs better than the ALP test for those processes that have breakpoints close to the middle of the sample (SB3, SB4 and SB5). The ALP test is superior if the breaks are late in the sample and asymmetric (as in SB6). The ALP test and the *Trig*-test outperform the two versions of the BP test considered here.

The power of the *Trig*-test to detect stochastic parameter variation is evaluated using three data-generating processes. The first is a stationary autoregressive process (SPV1) for the time-varying parameter (Watson and Engle, 1985); the second (SPV2) specifies the time-varying parameter as a martingale (Nyblom, 1989); and the third (SPV3) uses the bilinear specification of Lee *et*. *al*. (1993). The formal definitions are:

SPV1: $$y_t = b_t y_{t-1} + e_t;\ b_t = 0.3 + 0.5b_{t-1} + v_t\ ;\ e_t \sim N(0, 1)\ ;\ v_t \sim N(0, 0.25) \qquad (8)$$

---

[1] Of course, the sequential tests for multiple breaks developed by BP are likely to have more power than the *UDmax* and *Wdmax* tests. The code for all their tests is available at the *Journal of Applied Econometrics'* Data Archive and at www.econ.bu.edu/perron.

SPV2: $y_t = b_t y_{t-1} + e_t; b_t = 0.3 + 0.5b_{t-1} + v_t$; $e_t \sim$ N(0, 1) ; $v_t \sim$ N(0, 0.25) (9)

SPV3: $y_t = b_t y_{t-1} + e_t$; $b_t = 0.7e_{t-2}$; $e_t \sim$ N(0, 1) (10)

The power results for the five tests against these alternatives are presented in Table 3. Once again, the power of the CUSUM test against the three processes is disappointing. The CUSUM^2 test fares better, especially when the alternative is the martingale parameter process. The Watson-Engle test has satisfactory power against SPV1 and SPV2 but is poor against SPV3. The power of the Nyblom test is low in all three cases (even against the martingale alternative for which it was designed). In implementing the three versions of the *Trig*-test, we used a frequency range Γ = [1, 49]. The performance of the LR_{ave} version of the *Trig*-test is impressive, indicating that averaging across different values for the unidentified parameter, f, is very valuable in this context. These results are the strongest demonstration yet of the usefulness of the *Trig*-test in detecting time dependence in parameters.

3. Empirical Illustration: U.S. Inflation

Bai and Perron (2003) present an example of a break in the U.K. inflation rate. In order to illustrate the OLS-based *Trig*-test, monthly values of the U.S. *CPI* (seasonally adjusted) were obtained from the website of the Federal Reserve Bank of St. Louis for the period 1947:1 to 2001:11. As shown in Figure 1, inflation rates during the 1970's were substantially higher than those prevailing in other periods. Let y_t denote the logarithmic change in the U.S. *CPI*, the following augmented Dickey-Fuller test (with t-statistics in parentheses) shows that the unit-root hypothesis can be rejected for the sample:[2]

$$\Delta y_t = \underset{(3.00)}{0.591} - \underset{(-4.22)}{0.171} y_{t-1} - \sum_{i=1}^{11} b_i \, \Delta y_{t-i} + \varepsilon_t \tag{11}$$

[2] The *AIC* select the 12-lag specification while the *SBC* selects a model with 11-lagged changes. The essential results are virtually identical using either specification.

The key point to note is that standard diagnostic checks of the residuals indicate that the model is adequate. If ρ_i denotes the residual autocorrelation for lag *i,* the correlogram is:

ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}
-0.006	0.012	0.033	-0.008	0.022	0.052	-0.019	0.011	0.027	-0.012	-0.039	0.074

To test the stationary inflation series for possible structural breaks, the OLS version of the *Trig*-test is implemented with the trigonometric expansion applied to the constant term. The resultant test regression is estimated for all $f_j \in \Gamma = [(1/512)\,, 3]$ in steps of 1/512. The "best" fitting frequency is $f^* = 1.178$ yielding the regression model

$$\begin{array}{l} y_t = \underset{(5.17)}{1.33} - \underset{(-3.92)}{0.900}\ \sin(2p\, f^* t/T) - \underset{(-2.31)}{0.402}\ \cos(2p\, f^* t/T) + \sum_{i=1}^{12} b_i\, y_{t-i} + e_t \end{array} \tag{12}$$

The computed value of the *Trig*-test statistic is 9.911. Hence, the null hypothesis of linearity is rejected using the conservative 1% critical value of 9.17 (Ludlow and Enders, 2000). Figure 1 superimposes a scaled version of the time-varying intercept on the actual U.S. inflation rate. It is clear the time-varying intercept captures the behaviour of the inflation rate during the 1970's.

For comparative purposes, the BP testing procedure was applied to the U.S. inflation data. Both the *UDmax* and *Wdmax* tests are unable to reject the hypothesis of parameter constancy. The test statistics recorded were 6.0634 (critical value at 10% is 7.46) and 7.3611 (critical value at 10% is 8.20) respectively. This result is consistent with the simulation evidence presented above indicating that these tests have lower power than the *Trig*-test. However, the BP *supF* test statistic for 2 structural breaks, given the presence of 1 break, is 12.3871 which is significant at the 1% level and the number of breaks chosen by the BIC is also 2. The dates for these two breaks are 1972:01 and 1980:09 which correspond nicely with the pattern revealed by the time-varying intercept in Figure 1.

4. Conclusion

The framework presented in this paper develops a straightforward way of testing regression parameters for time dependence. A test based on a trigonometric expansion of the coefficient believed to be time varying is shown to capture the effects of variation due to either structural breaks or stochastic parameters. A rigorous implementation of the test uses bootstrapping to deal with the fact that the frequency of the trigonometric expansion is unidentified under the null hypothesis. A simple OLS version of the test, however, performs just as well. The Monte Carlo experiments reported in the paper provide empirical evidence that the *Trig*-test has the correct size and good power against all the alternative models considered. The *Trig*-test detects structural breaks as reliably as the ALP and BP tests. The test also proved to have better power against stochastic parameter processes than tests specifically designed for this purpose. The *Trig*-test detects the presence of structural breaks in the U.S. inflation rate. Specifically the test detects an increase in the U.S. inflation rate beginning in the early 1970s and lasting through to the early 1980s.

References

Andrews D.W.K., Lee, I. and Ploberger, W. (1996) "Optimal changepoint test for normal linear regression" *Journal of Econometrics*, 70, 9-38.

Andrews D.W.K. and Ploberger W. (1994) "Optimal tests when a nuisance parameter is present only under the alternative" *Econometrica*, 62, 1383-1414.

Bai, J. and Pierre P. (1998) "Estimating and testing linear models with multiple structural changes." *Econometrica,* 66, 47-78.

Bai, J. and Pierre P. (2003) "Computation and analysis of multiple structural change models." *Journal of Applied Econometrics,* 18, 1-22.

Clements M.P. and Hendry D.F. (1999) "On winning forecasting competitions" *Spanish Economic Review*, 1, 123-60.

Davies R.B. (1987) "Hypothesis testing when a nuisance parameter is only identified under the alternative" *Biometrika*, 47, 33-43.

Hansen B. (1999) "Testing for linearity" *Journal of Economic Surveys*, 13, 551-576.

Lee, T., White H. and Granger C.W.J. (1993) "Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests." *Journal of Econometrics*, 56, 269-290.

Ludlow J. and Enders W. (2000) "Estimating non-linear ARMA models using fourier coefficients" *International Journal of Forecasting*, 16, 333-347.

Nyblom, J. (1989) "Testing for the constancy of parameters over time" *Journal of the American Statistical Association,* 84, 223-30.

Stinchcombe M.B. and White H. (1998) "Consistent specification testing with nuisance parameters present only under the alternative" *Econometric Theory*, 14, 295-325.

Watson, M. and Engle R. F. (1985) "Testing for regression coefficient stability with a stationary AR(1) alternative" *Review of Economics and Statistics*, 67 (2), 341-46.

Table 1: Empirical size of the *Trig*-test

	p-value	The AR(1) Model			AR[4] Constant Term			The AR[4] Coefficient		
		LR_{sup}	LR_{ave}	LR_{exp}	LR_{sup}	LR_{ave}	LR_{exp}	LR_{sup}	LR_{ave}	LR_{exp}
T=50	*0.01*	0.012	0.012	0.012	0.016	0.014	0.011	0.013	0.016	0.014
	0.05	0.048	0.055	0.049	0.065	0.058	0.051	0.055	0.064	0.055
	0.10	0.095	0.109	0.101	0.123	0.109	0.102	0.107	0.121	0.105
T=100	*0.01*	0.009	0.012	0.010	0.015	0.013	0.014	0.013	0.015	0.013
	0.05	0.054	0.047	0.051	0.051	0.051	0.060	0.055	0.049	0.051
	0.10	0.100	0.097	0.097	0.111	0.096	0.121	0.103	0.115	0.099
T=200	*0.01*	0.013	0.015	0.014	0.014	0.010	0.016	0.013	0.014	0.011
	0.05	0.056	0.060	0.058	0.061	0.050	0.064	0.055	0.062	0.053
	0.10	0.109	0.117	0.109	0.111	0.100	0.123	0.104	0.112	0.099

Size computed from 5000 repetitions with 400 bootstrap replications per repetition.

Table 2: Power of CUSUM, CUSUM^2, ALP, BP and *Trig*-test against structural breaks.

	p-value	SB1	SB2	SB3	SB4	SB5	SB6
CUSUM	*0.01*	0.015	0.006	0.012	0.014	0.007	0.008
	0.05	0.056	0.030	0.080	0.044	0.040	0.036
	0.10	0.116	0.084	0.157	0.101	0.099	0.082
CUSUM^2	*0.01*	0.032	0.036	0.008	0.026	0.052	0.056
	0.05	0.123	0.112	0.047	0.104	0.173	0.178
	0.10	0.193	0.191	0.098	0.169	0.278	0.273
ALP	*0.01*	0.320	0.174	0.069	0.121	0.633	0.105
	0.05	0.544	0.331	0.196	0.276	0.814	0.247
	0.10	0.675	0.450	0.323	0.375	0.885	0.364
LR_{exp}	*0.01*	0.286	0.140	0.198	0.121	0.684	0.065
	0.05	0.485	0.294	0.408	0.284	0.877	0.171
	0.10	0.615	0.407	0.527	0.383	0.918	0.270
Udmax	*0.01*	0.127	0.058	0.039	0.062	0.332	0.028
	0.05	0.276	0158	0.143	0.165	0.559	0.100
	0.10	0.381	0.241	0.221	0.245	0.675	0.187
WDmax	*0.01*	0.109	0.044	0.047	0.056	0.326	0.030
	0.05	0.246	0.156	0.151	0.157	0.551	0.108
	0.10	0.349	0.232	0.230	0.240	0.662	0.192

Power computed from 1000 repetitions with 400 bootstrap replications (ALP and *Trig*-test) per repetition.

Table 3: Power of tests for stochastic parameter variation

SPV1	**CUSUM**	**CUSUM^2**	**Watson-Engle**	**Nyblom**	**LR$_{sup}$**	**LR$_{ave}$**	**LR$_{exp}$**
0.01	0.060	0.533	0.818	0.052	0.723	0.831	0.747
0.05	0.123	0.653	0.914	0.128	0.838	0.928	0.876
0.10	0.207	0.723	0.954	0.215	0.892	0.960	0.912
SPV2	**CUSUM**	**CUSUM^2**	**Watson-Engle**	**Nyblom**	**LR$_{sup}$**	**LR$_{ave}$**	**LR$_{exp}$**
0.01	0.028	0.891	0.756	0.462	0.923	1.000	0.924
0.05	0.039	0.895	0.758	0.466	0.989	1.000	0.989
0.10	0.047	0.897	0.760	0.494	0.997	1.000	0.997
SPV3	**CUSUM**	**CUSUM^2**	**Watson-Engle**	**Nyblom**	**LR$_{sup}$**	**LR$_{ave}$**	**LR$_{exp}$**
0.01	0.028	0.501	0.051	0.118	0.600	0.808	0.634
0.05	0.066	0.643	0.136	0.228	0.776	0.918	0.824
0.10	0.102	0.741	0.252	0.301	0.843	0.944	0.887

Sample size = 100. Power computed in 1000 repetitions using 400 bootstrap replications in each for the Watson-Engle and *Trig*-tests

Figure 1: A Structural Break in U.S. Inflation?