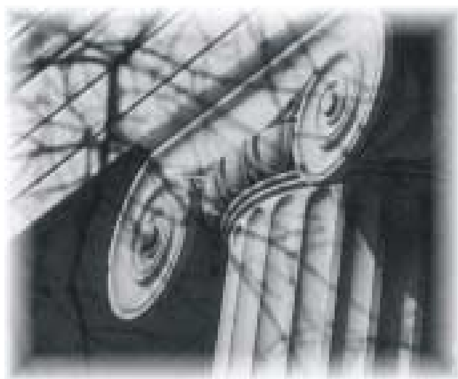
THE UNIVERSITY OF ALABAMA

ECONOMICS, FINANCE AND LEGAL STUDIES

WORKING PAPER SERIES



Working Paper No. 01-02-05

A General Test For Time-dependence in Parameters

Ralf Becker, Walter Enders and Stan Hurn

February 4, 2002

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

We propose a new test based on a Fourier series to approximate the unknown form of a nonlinear time-series model. The test has good size and power properties to detect structural breaks, seasonal parameters and random coefficients. Moreover, it has reasonable power to discriminate between nonlinearity in variables and nonlinearity in parameters. We use the test to show that U.S. inflation is appropriately estimated with a time-varying intercept that jumps in the late 1960's, peaks in the early 1980's and then begins to decline. German income and consumption data is used to illustrate the ability of the test to suggest the form of the nonlinearity.

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.

A General Test for Time Dependence in Parameters

1. Introduction

This paper proposes a test for time dependence in parameters that can approximate the unknown functional form of a time-varying regression coefficient. The null hypothesis is a linear model with time-invariant parameters and the alternative hypothesis is a model that is linear in variables but has time-varying coefficients. The issue is important since time-varying coefficients may result from structural breaks, seasonality or stochastic parameter variation. The proposed test is capable of detecting each of the alternative hypotheses. Rigorous implementation of this testing strategy has to deal with the presence of an unidentified parameter under the null hypothesis (see Davies, 1987, Andrews and Ploberger, 1994, Hansen 1996 and Stinchcombe and White, 1998) but this problem is easily handled in a bootstrapping environment. It is also shown that a version of the test based on OLS is very simple to implement.

The paper is structured as follows. Section 2 reviews some popular tests for time-varying parameters. Section 3 introduces our trigonometric testing approach and Section 4 discusses the relevant literature concerning inference in the presence of a parameter that is unidentified under the null hypothesis. In Section 5, a number of Monte Carlo simulation exercises are performed to document the empirical performance of our proposed test. Since many software packages do not readily perform bootstrapping, Section 6 presents a simplified version of the test. To preview our main findings, the test has the correct size and good power to detect structural breaks, seasonality and stochastic parameter variation. It is interesting that the test has poor power to detect nonlinearity in variables. As such, it can help identify the form of the nonlinearity. Moreover, the use of a Fourier approximation to 'mimic' the time-varying parameter can also help identify the nature of the nonlinearity.

Section 7 presents two examples that illustrate how the method works in practice. The U.S. inflation rate, as measured by the *CPI*, is shown to have a time-dependent intercept term. In addition, we compare the ability of our procedure to detect seasonal coefficients to that of Lütkepohl and Herwartz (1996) by re-examining their data on German per-capita income and consumption. Section 8 contains brief concluding comments.

2. Testing for parameter variation

Perhaps, the most common reason given for time variation in the parameters of an econometric model is that of structural breaks (see Perron, 1989, Clements and Hendry, 1999). Let b_{it} be the i-th element of the (k x 1) parameter vector $\hat{\mathbf{a}}_t$ from a classical regression model. If there is a single breakpoint at $t = t$, we can write:

$$b_{it} = \begin{cases} b_{i1} \text{ for } t = 1 \ldots t \\ b_{i2} \text{ for } t = t + 1 \ldots T \end{cases} \tag{1}$$

If the date of the breakpoint t is known, it is straightforward to test for a change in the regression coefficients using a Chow test. One variant of the Chow test splits the sample into two periods and uses an F-test to determine whether the coefficients in each period are distinct. Alternatively, a dummy variable may be employed and a simple t-test can be used to detect time variation in a particular coefficient b_{it}. An attractive feature of the Chow test is that it is possible to investigate multiple breaks at known breakpoints in a similar way.

If the breakpoint is not known, the testing strategy is more complex since, under the null hypothesis, t is an unidentified nuisance parameter. The optimal test turns out to be a weighted average of F-statistics, called $F(t)$, which can be used to test the null hypothesis of parameter constancy. While it is straightforward to calculate the sample value of $F(t)$, the theoretical distribution is nonstandard so that it is necessary to use the asymptotic critical values tabulated by Andrews and Ploberger (1994). Note that is necessary to use a different set of critical values when testing for m breakpoints and the situation will be even more complex should m be left unspecified.

The CUSUM and CUSUM[2] tests were also designed to detect structural change. These tests involve the cumulated sums of standardized one-step-ahead prediction errors. If the sum of these errors is statistically significant, it is possible to conclude that there is a structural break. Brown *et al.* (1975) describe how the significance of the resultant test statistics can be evaluated.

A second reason for parameter variation is the presence of seasonality. If it is suspected that b_{it} is subject to seasonal variation of periodicity S, we can write:

$$b_{it} = \begin{cases} b_{i1} \text{ for } t = 1, 1+S, 1+2S\ldots \\ b_{i2} \text{ otherwise} \end{cases} \tag{2}$$

The usual testing procedure is to define a seasonal dummy variable, d_t, which takes the value 1 for observations 1, $1+S$, $1+2S$, etc. Then, premultiply the i-th variable in the regression by d_t and evaluate the significance of seasonal variation in b_{it} by means of a standard t-test. Of course, more general types of seasonal variation can be accommodated within this simple framework. It is instructive to note that tests for seasonal parameter variation can also be achieved by using trigonometric functions instead of dummy variables. Harvey (1989) and Herwartz (1995) show that, in certain circumstances, trigonometric functions can provide a very parsimonious representation of seasonal variation.

A third type of time dependence in parameters allows b_{it} to vary stochastically:[1]

$$b_{it} = a_1 + a_2 b_{it-1} + e_t \tag{3}$$

Modeling parameter variation of this sort is a signal-extraction problem that is amenable to the Kalman filter (Harvey, 1989). Two early tests to detect stochastic parameter variation are due to Watson and Engle (1985) and Nyblom (1989). The Watson-Engle test is

---

[1] Throughout, the symbol ε_t is used to denote a zero-mean identically and independently distributed (*i.i.d.*) error term.

designed to test the time invariance of one particular element b_{it} of the (k x 1) parameter vector $\hat{\mathbf{a}}$ while the Nyblom test evaluates the time-invariance of the entire parameter vector. Note that the tests have different alternative hypotheses. The alternative for the Watson-Engle test is that b_{it} is a stationary AR(1) process such that $e_t \sim N\left(0,\ s_e^2\right)$. Of course, the hypothesis of a constant value for b_{it} is equivalent to a test $s_e^2 = 0$.[2] The alternative hypothesis for the Nyblom test is that the parameter vector $\mathbf{b}$ is a martingale process.

All the tests outlined above are concerned with parameter variation in the time domain. Another general class of tests is concerned with detecting parameter variation in the frequency domain using band spectrum regressions (Engle, 1974). In essence, the mechanics of the test resembles that of the Chow test. Consider the null hypothesis:

$$\mathbf{y} = \mathbf{X} \quad + \hat{\mathbf{a}} e \qquad e \sim N\left(0, s_e^2\right) \tag{4}$$

where: under the null hypothesis, the parameter vector $\mathbf{b}$ is constant.

The discrete Fourier transform of $\mathbf{y}$ is created by pre-multiplying equation (4) with the (T x T) matrix $\mathbf{F}$ with typical element given by:

$$f_{mn} = \exp\left(2p\,i(m-n)(n-1)/T\right), \tag{5}$$

to yield:

$$\mathbf{Fy} = \mathbf{FX}\hat{\mathbf{a}} + \mathbf{F}e \qquad \mathbf{F}e \sim N\left(0, s_e^2\mathbf{FF}'\right). \tag{6}$$

Since $\mathbf{F}$ is an orthogonal matrix, $\mathbf{F}\mathbf{F}' = \mathbf{I}$. Hence, equation (6) may be rewritten as:

$$\mathbf{y}^* = \mathbf{X}^*\hat{\mathbf{a}} + e^* \qquad e \sim N\left(0, s_e^2\mathbf{I}\right). \tag{7}$$

If a Chow breakpoint test is applied to equation (7), it is possible to evaluate whether the parameter vector is constant across different frequency ranges (Engle, 1974). For

---

[2] It transpires that the value of the test statistic depends critically on the value of a_2, which is in fact unidentified under the null hypothesis. The range of values over which estimate the test statistic, however, follows naturally given the stationarity assumption. Watson and Engle (1985) recognize that the resultant test statistic has a nonstandard distribution and describe an upper bound for the critical value under the validity of the null hypothesis.

example, one might test for seasonality by excluding seasonal frequencies, or, alternatively, one could attempt to differentiate between contributions from high and low frequencies. This general approach has recently been developed further by Tan and Ashley (1999), who allow **b** in equation (4) to take on a number of different values.

The following section takes up the idea of testing for parameter variation using a test based on a trigonometric expansion with a single frequency to capture time dependence in regression parameter. As such, the test remains within the time domain. The important contribution, however, is that by treating the frequency of the trigonometric approximation as an unidentified parameter, the test is able to detect parameter variation due to any of the three major reasons for time invariance discussed above.

3. A test based on a trigonometric approximation

Consider the simple null hypothesis given by

$$y_t = x_t b + e_t \qquad e_t \sim N\left(0, s_e^2\right) \tag{8}$$

where x_t is assumed to be weakly stationary, while the alternative hypothesis specifies a model which is linear in variables but with a time-varying coefficient:

$$y_t = x_t b_t + e_t \qquad e_t \sim N\left(0, s_e^2\right). \tag{9}$$

Instead of specifying the form of the nonlinearity, our test makes use of the fact that any absolutely-integrable function b_t can be approximated by means of the Fourier-series expansion:

$$b_t = b_0 + \sum_{j=1}^{M}\left[b_{1j}\sin\left(\frac{2f_j p\, t}{T}\right) + b_{2j}\cos\left(\frac{2f_j p\, t}{T}\right)\right] + R \tag{10}$$

where b_0, b_{1j} and b_{2j} are constant parameters to be estimated, the f_j are the M frequencies in the expansion, T is the number of usable observations and R is a remainder term. With a sufficiently large $M,$ the unknown functional form of b_t may be approximated arbitrarily closely. The usefulness of this general approach derives from the fact that the linear model is the special case where all $b_{1j} = b_{2j} = 0$.

There are two immediate questions to be addressed in employing equation (10) as the basis of a testing strategy, namely the choice of M and the choice of the associated frequencies. For testing purposes, it appears reasonable to restrict M to 1; it stands to reason that if $b_{1j} = b_{2j} = 0$ can be rejected for just one frequency, the null hypothesis of time invariance is rejected. With this restriction imposed, the model under the alternative hypothesis in equation (9) now becomes:[3]

$$y_t = x_t\left[b_0 + b_1 \sin\left(\frac{2fp\,t}{T}\right) + b_2 \cos\left(\frac{2fp\,t}{T}\right)\right] + e_t \qquad e_t \sim N\left(0, s_e^2\right). \tag{11}$$

It is possible to use grid search methods to obtain estimates of the parameters b_0, b_1, b_2 and f. Note, however, that the frequency, f, is unidentified under the null hypothesis $b_1 = b_2 = 0$ so test statistics involving these parameters will not have a standard distribution.

Before proceeding to discuss the problem of dealing with the unidentified parameter in detail, it is useful to consider how the range of the frequency parameter will affect the power of the test. Should the parameter of interest follow a seasonal pattern with periodicity S (where $1 < S < T/2$) it is straightforward to see that this will be captured by the trigonometric expansion with frequency $f = T/S$.

The situation is less obvious when there are structural breaks or random coefficients. Panel *a* of Figure 1 illustrates a one-time structural break occurring halfway through a sample of 100 observations. It is also shown that a scaled sine function of frequency 1 can approximate this structural break. Panel *a* also illustrates the point that optimizing the fit of

[3] Although we do not pursue the issue here, it is conceptually straightforward to expand this specification to allow for more than one parameter being time dependent if it is assumed that all the relevant parameters can be expanded with trigonometric terms of the same frequency. Similarly, it may be that some of the coefficients in the regression are constant even under the alternative hypothesis.

the sine function by choice of the frequency improves the approximation considerably. In this case, a frequency of 0.8242 (found via a simple grid search) best approximates the structural break. More generally, frequencies less than 1 can be useful in the approximation of structural breaks and should be considered when using our test to detect breaks. Panel *b* illustrates the approximation of a temporary structural break using a single integer frequency. It should be apparent that, for integer frequencies, the approximation fits best when the break is near the center of the sample.

Panel *c* of Figure 1 illustrates the situation when b_t is generated as $0.5b_{t-1} + \varepsilon_t$. Here a relatively large frequency integer frequency ($f = 8$) best characterizes the nature of the process. In Panel *d*, b_t is a unit-root process; as in the case of a structural break, a low frequency best mimics the stochastic trend. Hence, when considering stochastic parameters, a simple strategy is to include a wide range of possible frequencies. Extending the range of frequencies, however, comes at a cost in the form of reducing the power of the test. This effect is well illustrated by Davies (1987).

4. The distribution of the test statistic

The appropriate method for dealing with the fact that the frequency, f, is unidentified under the null hypothesis is to use the methods developed by Davies (1987), Andrews and Plogerger (1994) and Hansen (1996). Consider the discrete set Γ containing G frequencies as its elements. One therefore has to consider G tests of the null hypothesis, H_0: $b_1 = b_2 = 0$, each constructed on the basis of a different frequency $f_j \in \Gamma$, $j = 1, \ldots G$. For each fixed frequency, the model under both the null and alternative hypotheses is easy to estimate. As such, for each frequency f_j, it is convenient to use a likelihood ratio test:

$$LR(f_j) = -2\left(l_T - l_T^{f_j}\right) \tag{12}$$

where l_T is the log-likelihood of the restricted model (*i.e.*, the linear model) and $l_T^{f_j}$ is the log-likelihood of the model which includes the trigonometric terms with frequency f_j.

It is now necessary to distil the information contained in these G values into one test statistic. This may be achieved by a mapping $f\left(LR(f_1), LR(f_2)...LR(f_G)\right)$ which maps $R^G \rightarrow R^1$. Three variations of $f\left(\bullet\right)$ have been proposed, namely, the sup-norm, unweighted average and exponentially weighted average (Davies, 1987, Andrews and Ploberger, 1994). The resultant test statistics are given by:

$$\begin{aligned} LR_{\text{sup}} &= \sup_{f_i \in \Gamma} LR(f_i) \\ LR_{\text{ave}} &= \frac{1}{G}\sum_{f_i \in \Gamma} LR(f_i) \\ LR_{\text{exp}} &= \ln\left[\frac{1}{G}\sum_{f_i \in \Gamma} \exp\left(\frac{LR(f_i)}{2}\right)\right]. \end{aligned} \tag{13}$$

The LR_{ave} and LR_{exp} tests are likely to be most useful when time variation in the parameter of interest is not exactly of a trigonometric type. Here, several frequencies might indicate deviations from the null hypothesis but any one frequency alone might not trigger a clear rejection. In these situations, averaging strategies may be superior in detecting deviations from the null hypothesis.

To ease the exposition slightly, let LR_f refer to the general case of a test statistic constructed using one of the three mappings. Note that the distribution of LR_f under the null hypothesis depends on four factors, namely, the set Γ, the mapping $f\left(\bullet\right)$, the distribution of each individual $LR(f_j)$ and the covariance between $LR(f_i)$ and $LR(f_j)$ for $i \neq j$ which may be denoted Σ. In general, the construction of critical values for this distribution is non-trivial (Andrews and Ploberger, 1994) and bootstrapping becomes the only realistic alternative.

The solution to the problem of determining the significance of the test statistic may be overcome by using the bootstrapping approach suggested by Hansen (1999) as follows. [4] *First*, n replications of the data, y_t^*, may be generated using the general scheme:

$$y_t^* = x_t \hat{b} + e_t^* \tag{14}$$

where the bootstrapped residuals, e_t^*, are resampled (with replacement) from the empirical distribution of residuals obtained by estimating the model under the null hypothesis [*i.e.,* the residuals from equation (8)].[5] Should x_t contain lags of the independent variable, as is the case in the simulations and the empirical examples to follow, the realizations y_t^* will need to be generated recursively. *Second*, for each of the n bootstrap samples, y_t^*, the sample statistic LR_f^n is computed. The estimate of the *prob*-value of the test is the proportion of the bootstrapped test statistics exceeding the LR_f test statistic computed from the observed data.

This parametric bootstrap is known to deliver consistent inference in regression problems (Bose, 1988, Horowitz, 1997) under the assumption of homoskedastic residuals and the *LR* test statistic is asymptotically pivotal (Li and Maddala, 1996).[6] Here, however, the test statistic is a mapping of correlated *LR* statistics and the theoretical consistency of the bootstrap in this environment is, as yet, unchartered water. However, for integer frequencies, $\sin(2\pi f_i t/T)$ and $\cos(2\pi f_i t/T)$ are orthogonal to each other *and* to the series $\sin(2\pi f_j t/T)$ and $\cos(2\pi f_i t/T)$, $i \neq j$. Hence, intuition suggests that the correlations between the *LR* statistics should be small. The empirical results from the various Monte Carlo exercises provided in the next section justify this conjecture.

---

[4] The test for the significant frequencies could be accomplished in the frequency domain, using the approach of Tan and Ashley (1999), for example. There seems little advantage in pursuing this avenue, however, seeing that the bootstrapping needs to be done in the time domain.

[5] Note the resampling is done by drawing from the rescaled and centered residuals (Mammen, 1993, Flachaire, 1999, Bergström, 1999).

[6] Although the assumption of homoskedastic residuals is maintained in this particular testing strategy, there are well-known techniques to address this problems arising from potential heteroskedasticity (Horowitz, 1997).

5. Empirical performance of the test

The previous section has asserted that a test based on a trigonometric expansion of a potentially time-varying regression parameter is capable of detecting departures from the null hypothesis of time invariance. This assertion is now investigated empirically in the form of a range of simulation experiments that are designed to demonstrate the following points. *First*, the test, which will be referred to as the *Trig*-test for ease of reference, has the correct statistical size under the null hypothesis of a linear time-invariant autoregressive process. *Second*, the test has satisfactory power properties for the alternative hypotheses of time-varying coefficients. *Third*, the test can distinguish between specifications involving time-varying coefficients and nonlinearity in variables.

5.1 Empirical size

With the exception of the section dealing with structural breaks, the null hypothesis in the simulations experiments will be an autoregressive process of some predetermined order. This section therefore simulates two linear autoregressive processes with constant parameters to evaluate the size of the test. Consider:

$$y_t = 0.6 y_{t-1} + e_t \quad e_t \sim N(0,1). \tag{15}$$

$$y_t = 0.9 y_{t-4} + e_t \quad e_t \sim N(0,0.04). \tag{16}$$

The first process is a standard AR(1) model. The second process, adapted from Lütkepohl and Herwartz (1996), is an autoregressive AR[4] process often used to capture seasonality.[7] 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

---

[7] Square brackets denote that the fourth-order process has the parameters on the first three lags set to zero, to distinguish it from the traditional AR(4) notation.

tested for time invariance by means of the *Trig*-test.[8] All three mappings are applied and the choice of frequencies is $\Gamma = [\ 1, (\ usable\ observations/2 - 1)\ ]$. This range is fairly large, but the size of the test should be correct for any choice of Γ, as the primary influence of the range of Γ is on the power of the test rather than on the size.

The results reported in Table 1 clearly illustrate the excellent size properties of the proposed *Trig*-test. Regardless of the mapping utilized and the length of the process (as one would expect for a bootstrapped significance level), the actual empirical size is very close to the nominal size.

5.2 Power against structural breaks

In order to evaluate the *Trig*-test's ability to detect structural breaks, its performance will be compared against that of the optimal test for one breakpoint proposed by Andrews, Lee and Ploberger (1996) – ALP hereafter – and the CUSUM and CUSUM^2 tests (Brown *et al*., 1975). Although the ALP test is optimal for a single break, Figure 1 suggests that the *Trig*-test might do well in detecting multiple breaks. For this reason the ensuing simulation experiment will entail models with one and two changes in parameter values. The basic processes used in this simulation are those used by Clements and Hendry (1999b)

$$y_t = b_0 + b_{1t} x_t + e_t \qquad e_t \sim N(0,1)\,, \tag{17}$$

with six different specifications for b_{1t}:

$$SB1:\ b_{1t} = \begin{bmatrix} 1 & t \le 40 \\ 1.5 & t > 40 \end{bmatrix}$$

$$SB2:\ b_{1_t} = \begin{bmatrix} 1 & t \le 50 \\ 1.5 & t > 50 \end{bmatrix}$$

$$SB3:\ b_{1_t} = \begin{bmatrix} 1 & t \le 20, t > 40 \\ 1.5 & 20 < t \le 40 \end{bmatrix}$$

---

[8] In the DGP for (16), the value of the intercept is zero. In generating Table 1, we estimated AR[4] models with and without intercept terms. The results are virtually identical.

$$SB4: \; b_{1t} = \begin{cases} 1 & t \le 40, t > 55 \\ 1.5 & 40 < t \le 55 \end{cases}$$

$$SB5: \; b_{1t} = \begin{cases} 1 & t \le 20 \\ 1.5 & 20 < t \le 40 \\ 0.5 & t > 40 \end{cases}$$

$$SB6: \; b_{1t} = \begin{cases} 1 & t \le 40 \\ 1.5 & 40 < t \le 55 \\ 0.5 & t > 55 \end{cases}$$

In all the models, b_0, is set to zero, the values for x_t are drawn from a normal distribution with mean and variance equal to unity and the length of the simulated series T = 60. The ALP test examines all possible breakpoints occurring within the middle 90% range of the data [$4 \le t \le 56$]. The *Trig*-test includes frequencies in the range [1/512, 6] in steps of 1/512. Although all three mappings of the LR_f tests were computed, the performance of each was quite similar; as a result only those for LR_{exp} are reported.

Before discussing the results reported in Table 2, it is useful to point out some issues that are noted by Clements and Hendry (1999b). The scale of the random variable x_t matters as it transpires that the power of tests for structural breaks improves as the mean of x_t increases.[9] They also demonstrate how the power increases with increasing the size of the break and how detection of the break is made more difficult as the breakpoint moves to either end of the sample. 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, both the ALP and *Trig*-tests experience severe reduction in power as the breakpoint shifts from observation 40 to 55. Also note the relatively poor performance of the CUSUM and CUSUM^2 tests relative to the ALP and *Trig*-tests.

Turning now to a comparison between the ALP and *Trig*-tests, it appears that the former has a slight advantage in terms of power when there is only one changepoint (models

---

[9] The results reported here relate only to x_t with mean of 1, but this tendency for power to increase with increased mean of x_t was confirmed in additional simulations.

SB1 and SB2). It should be emphasized, however, that this advantage is in the neighborhood of 10%. Differences arise when two-breakpoint models (SB3 to SB6) are generated. 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 improved power of the *Trig*-test can be substantial; at the 1% and 5% levels, the *Trig*-test has more than twice the power of ALP test to detect breaks in the form of SB3. The ALP test is superior if the breaks are late in the sample and asymmetric (as in SB6).[10]

5.3 Power against seasonal variation

The *Trig*-test is expected to have power against seasonal parameter variation as the inclusion of the relevant trigonometric terms is identical to the use of seasonal dummy variables. An important difference is that the *Trig*-test does not require the frequency to be fixed to fit a certain seasonal pattern.

Three different processes with quarterly parameter variation are simulated. The first process, SP1, is similar to an AR(1), but with the distinguishing feature that the parameter switches every fourth period;

$$y_t = b_s y_{t-1} + e_t \qquad e_t \sim N(0,1)$$
$$b_s = \begin{cases} 0.9 & s = 1,2,3 \\ -0.2 & s = 4 \end{cases} \tag{18}$$

The second process, SP2, is used by Lütkepohl and Herwartz (1996) to illustrate the use of a feasible generalized least-squares method to model time-varying parameters;

$$y_t = b_{0t} + b_{1t} y_{t-1} + e_t \qquad e_t \sim N(0,1)$$
$$b_{0t} = \{-0.6, 0.3, -0.9, 0.8\} \text{ for } s = 1,2,3,4 \tag{19}$$
$$b_{1t} = \{-0.4, 0.7, -0.3, 0.2\} \text{ for } s = 1,2,3,4.$$

The final process, SP3, is similar to SP2 but has been altered slightly to display a symmetrical seasonal pattern:

[10] Note that the ALP test is not an optimal test for multiple breakpoints.

$$
\begin{aligned}
&y_t = b_{0t} + a_{1t} y_{t-1} + e_t \qquad e_t \sim N(0,1) \\
&b_{0t} = \{-0.6, 0.3, 0.6, -0.3\} \text{ for } s = 1,2,3,4 \\
&b_{1t} = \{-0.4, 0.7, 0.4, -0.7\} \text{ for } s = 1,2,3,4.
\end{aligned} \tag{20}
$$

The autoregressive parameter is tested for time variation in all three processes. In addition, the constant term in SP2 and SP3 is also tested. The frequency set used for the *Trig*-test is Γ = [1, 49] with a step size of 1. In testing for seasonality using dummy variables, statistical significance is assessed by means of a t-test if only one coefficient is involved and by means of an F-test if both the constant and the autoregressive parameter are being tested. In a sense, our head-to-head comparison of the *Trig*-test is unfair since the dummy variables used to test for seasonality correspond to the actual seasonal pattern. In contrast, we do not force the frequency of the *Trig*-test to equal *T/S*.

Several conclusions may be drawn from the results reported in Table 3. *First*, of the three versions of the *Trig*-test the LR_{ave} has the lowest power. This accords with our intuition since averaging across all frequencies will dilute the detection of the strong single seasonal frequency. Of course, the frequency associated with the LR_{sup} test will provide useful information concerning the nature of the dominant seasonal frequency. *Second*, when the seasonal process is reasonably simple, as in SP1, the power of the *Trig*-test is indistinguishable from the power of the t-test using seasonal dummy variables. In other words, not knowing the exact periodicity of the seasonal pattern is, for all practical purposes, not penalised in these situations. *Third*, for more complex seasonal patterns (SP2 and SP3), not knowing the exact periodicity causes a reduction in power, particularly if the pattern of seasonality is asymmetric (SP2). Fourth, it transpires that the *Trig*-test detects periodicity in the autoregressive coefficients more easily than periodicity in the constant term.

In summary, therefore, it appears that the *Trig*-test is a useful tool in testing for seasonality in the behavior of the coefficients of regression equations. It reliably detects seasonal parameter variation even when the periodicity is not imposed *a priori* and whenever it is found to reject the null hypothesis, the frequency of the LR_{sup} test is always that of the simulated seasonality.

5.4 Power against stochastic parameter variation

Testing for stochastic parameter variation is not as common as testing for seasonality or for structural breaks. The only suitable tests implemented in most standard software packages are the CUSUM and CUSUM^2 tests. Although the Watson and Engle (1985) and Nyblom (1989) tests have not attracted much attention in empirical work, we include them in our comparisons.

The power of the *Trig*-test to detect stochastic parameter variation is tested using two data generating processes corresponding to those of the Watson-Engle and Nyblom tests. The former assumes a stationary autoregressive process (SPV1) for the time-varying parameter while the latter assumes that the time-varying parameter is a martingale (SPV2). The actual models used in the simulations are:

$$\begin{aligned} y_t &= b_t y_{t-1} + e_t \qquad & e_t &\sim N(0,1) \\ b_t &= 0.3 + 0.5 b_{t-1} + n_t \qquad & n_t &\sim N(0,0.25) \end{aligned} \tag{21}$$

and

$$\begin{aligned} y_t &= b_t y_{t-1} + e_t \qquad & e_t &\sim N(0,1) \\ b_t &= b_{t-1} + n_t \qquad & n_t &\sim N(0,0.25). \end{aligned} \tag{22}$$

The power results for the five tests against these alternatives are presented in Table 4. Once again, the power of the CUSUM test against both simulated DGPs 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 both processes, but the performance of the Nyblom test is poor by comparison, even against the Martingale alternative.

Since the sample size is 100, in implementing the three versions of the *Trig*-test, we used a frequency range Γ = [1, 49]. This choice yields power results which are particularly impressive for the LR_{ave} mapping, indicating that averaging across different values for the unidentified parameter, f, is very valuable in this context. These results are strongest demonstration yet of the usefulness of the *Trig*-test in detecting time dependence in parameters.

5.5 Power against nonlinearity in variables

If a test for nonlinearity is to be useful, it has to be able to help identify the form of the nonlinearity. Barnett *et al*. (1997) and Ashley and Patterson (2000) demonstrate how difficult it is to detect the form of a nonlinear model. One advantage of the *Trig*-test is that the form of the nonlinearity can be visualized by a graphical representation of the approximated function. Moreover, the test seems to be able to discriminate between nonlinearity in coefficients versus nonlinearity in variables. The following bilinear (BL), threshold autoregressive (TAR) and logistic smooth-transition autoregressive (LSTAR) processes that are nonlinear in variables have been used by Lee *et al.* (1993), Teräsverta *et al.* (1993) and Dahl (1998):

$$\text{BL: } y_t = 0.7 y_{t-1}\varepsilon_{t-2} + \varepsilon_t \qquad (23)$$

$$\text{TAR: } y_t = b_t y_{t-1} + \varepsilon_t \qquad (24)$$

$$b_t = \begin{cases} 0.9 \text{for} |y_{t-1}| \leq 1 \\ -0.3 \text{for} |y_{t-1}| > 1 \end{cases}$$

$$\text{LSTAR: } y_t = 0.02F_t + (1.8 - 0.9F_t)\, y_{t-1} + (-1.06 + 0.795F_t) y_{t-2} + \varepsilon_t \qquad (25)$$

$$F_t = [\, 1 + \exp(100(y_{t-1} - 0.02))\,]^{-1}$$

Table 4 reports the power of the LR_{exp} test against a number of commonly uses tests for nonlinearity in variables, namely the Tsay test, the V23 test, the LSTAR test and the CUSUM and $CUSUM^2$ tests.

Notice that the LR_{exp} test does have some power against the bilinear model. This should not be too surprising since the BL model can be interpreted an AR(1) model such that the degree of autoregressive decay has a mean of 0.7 and a variance equal to the variance of ε_{t-1}. Therefore, the coefficient can be viewed as time-varying. However, as hoped, the power of the *Trig*-test to detect the TAR and LSTAR models is very limited. Hence, the test can be useful in identifying the nature of the nonlinearity.

6. A simple implementation of the test

The previous section has emphasised that the significance of the test statistic is correctly evaluated by means of bootstrapping because the distribution of this statistic is sensitive to the particular data set being considered, or in other words, is not invariant under the null hypothesis. In one special case, established by Davis (1987), the distribution of the test statistic is invariant to nuisance parameters under the null hypothesis. Using the notation from equation (12), an invariant distribution under the null hypothesis is obtained when $E[LR(f_i), LR(f_j)] = 0$ for $i \neq j$. Consider

$$y_t = b_1 \sin(2f_j\pi t/T) + b_2 \cos(2f_j\pi t/T) + \varepsilon_t \quad (26)$$

where ε_t is independently normally distributed.

In these circumstances, a simple OLS procedure may be used to generate a test for $b_1 = b_2 = 0$. Estimate equation (26) by OLS for each value of f_j in the restricted domain. Denote by f^* 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 that frequency value. Denote the F-statistic for the null hypothesis $b_1^* = b_2^* = 0$ as F_{trig}^{OLS}. Hence:

$$F_{trig}^{OLS} = \frac{(RSS_r - RSS^*)/2}{RSS^*/(T-k+1)} \quad (27)$$

where RSS_r is the residual sum of squares with the restriction imposed.

Since the F_{trig}^{OLS} statistic is calculated for the frequency that minimizes the residual sum of squares, this is equivalent to using the sup-norm mapping of the previous section. Note, however, the distribution of the test statistic under the null hypothesis, although invariant, does not follow a standard F-distribution. Davies (1987) approximation to the asymptotic distribution or the tabulated critical values by Ludlow and Enders (2000) can be used to perform the test.[11] For convenience, Table 5 reports critical values for standard *prob*-values and sample sizes using a maximum frequency of $T/2 - 1$.

It can be shown that these critical values are invariant to the present of an intercept term. A natural extension is to extend equation (26) to allow for lagged dependent variables. Towards this end, we generated critical values for the following three autoregressive models:

$$\text{AR(1)}: \; y_t = b + b_0 y_{t-1} + \varepsilon_t \qquad b_0 \in [\, 0.1, 0.9 \,] \tag{28}$$

$$\text{AR(2)}: \; y_t = b + b_0 y_{t-1} - 0.3 y_{t-2} + \varepsilon_t \qquad b_0 \in [\, 0.1, 0.9 \,] \tag{29}$$

$$\text{AR[4]}: \; y_t = b + b_0 y_{t-4} - 0.3 y_{t-2} + \varepsilon_t \qquad b_0 \in [\, 0.1, 0.9 \,] \tag{30}$$

[11] Davis (1987) reparameterizes (26) such that:

$$E_{t-1}(x_t) = a_1 \sin[\,(t - 0.5T - 0.5)\theta\,] + b_1 \cos[\,(t - 0.5T - 0.5)\theta\,]$$

where: the values of $\{x_t\}$ are zero-mean, unit-variance *i.i.d* normally distributed random variables with a period of oscillation equal to $2\pi/k$ (since $\theta = 2\pi k/T$). In practice, Davies recommends demeaning and standardizing the sample to obtain the $\{x_t\}$ sequence. For the possible values of θ in the range $[\, L, U \,]$ where $0 \le L < U \le \pi$, construct:

$$S(q) = \left[\sum_{t=1}^{T} x_t \sin[(t - 0.5T - 0.5)q\,]\right]^2 / v_1 + \left[\sum_{t=1}^{T} x_t \cos[(t - 0.5T - 0.5)q\,]\right]^2 / v_2$$

where: $v_1 = 0.5T - 0.5\sin(T\theta)/\sin(\theta)$ and $v_2 = 0.5T + 0.5\sin(T\theta)/\sin(\theta)$.

If fractional frequencies are used, Davies shows that:

$$prob\ [\ sup\ \{\ S(q): L \le q \le U\ \} > u\ ]$$

can be approximated by:

$$Tu^{0.5} e^{-0.5u} (\,U - L\,)/(24\pi)^{0.5} + e^{-0.5u}$$

If only integer frequencies are used, Davies shows that the approximation to use is:

$$1 - (\,1 - e^{-0.5u})0.5T^{(U-L)/\pi}$$

Each model was simulated for the parameter ranges shown and the coefficients were tested for time variation. For example, for each potential frequency in the interval $[1, T/2 - 1]$, the AR(1) specification was estimated as:

$$y_t = b + [\ b_0 + b_1\sin(2\pi f_j t/T) + b_2\cos(2\pi f_j t/T)\ ]\ y_{t-1} + \varepsilon_t \tag{31}$$

Now, let b_1^* and b_2^* be the estimates of b_1 and b_2 associated with the integer frequency yielding the smallest residual sum of squares. Sample values of F statistic for the hypothesis $b_1^* = b_2^* = 0$ using 50000 repetitions were tabulated. Note that the AR(2) model allows the test to be conducted on both α_1 and α_2. The results for the 5% significance level are presented in Table 6.

Since the effect of changing the parameters of the underlying linear model are small, the OLS-based procedure can be a useful approximation to the bootstrapped critical values. The key to this result is the restriction of the domain of the frequency, f_j, to only integer values between 1 and $T/2$. As the approximate distribution changes very little with changes in the parameters, the critical values tabulated in Ludlow and Enders (2000), where variations the sample size for given Γ are considered, can be used.

7. Empirical Illustrations

7.1 The OLS Test: A Structural Break in the Inflation Rate

It is well known that inflation rates, measured by the *CPI*, act as long-memory processes. For example, Hall (1986) presents evidence that the U.K. *CPI* is $I(2)$, and Hatanaka (1996) presents similar evidence for the U.S. More recently, Baillie, Han and Kwon (2002) review a number of papers indicating that U.S. inflation is fractionally integrated. Clements and Mizon (1991) argue that structural breaks can explain such findings; a break in a time-series can cause it to behave like a unit-root process.

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 (http://www.stls.frb.org/fred/index.html) for the 1947:1 to 2001:11 period. As shown in Panel *a* of Figure 2, inflation rates during the 1970's were substantially above those prevailing in other periods. If we 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 our long sample:[12]

$$\text{Ä}y_t = \underset{(3.00)}{0.591} - \underset{(-4.22)}{0.171}y_{t-1} - \sum_{i=1}^{11} b_i \, \text{Ä}y_{t-i} + e_t \qquad (32)$$

Rewriting the equation in levels:

$$y_t = \underset{(3.10)}{0.591} + \underset{(7.70)}{0.301}y_{t-1} + \underset{(3.96)}{0.162}y_{t-2} + \underset{(1.82)}{0.074}y_{t-3} + \ldots - \underset{(-4.56)}{0.170}y_{t-12} + e_t \qquad (33)$$

The key point to note is that standard diagnostic checks of the residual series $\{e_t\}$ 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

Moreover, the Ljung-Box *Q*-statistics (with *prob*-values in parentheses) are:

$Q(4) = 0.901$ (*prob* = 0.924) $Q(8) = 3.023$ (*prob* = 0.933)
$Q(12) = 8.280$ (*prob* = 0.763) $Q(16) = 21.062$ (*prob* = 0.176)
$Q(20) = 23.801$ (*prob* = 0.251) $Q(24) = 41.568$ (*prob* = 0.014).

The 24-th autocorrelation is -0.148 and so that $Q(24)$ has a low *prob*-value. However, all other autocorrelations are quite low and, since the data is seasonally adjusted, most researchers would ignore this value of ρ_{24}.

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

To use the OLS version of the *Trig*-test, let $\{e_t\}$ denote the regression residuals from equation (33). Since we are searching for a small number of possible breaks, we divided the interval $\Gamma = [0\ ,\ 3]$ into 512 steps and performed the test using Davies parameterization. The "best" fitting frequency was found to 1.00195 and the sample value of test statistic is: 11.72. This value is significant at the 2.5% level. Similarly, we obtained the same frequency from OLS regressions of the form:

$$e_t = a_0 + a_1 \sin(2p\ f_j t / T) + a_2 \cos(2p\ f_j t / T) + e_t \tag{34}$$

Using f_j = 1.00195, we constructed the variables sin[2π (1.00195) t/T] and cos[2π (1.00195) t/T].[13] Next, we estimated the regression equation:

$$\underset{(4.85)}{y_t = 1.12} \underset{(-2.54)}{- 0.486} \sin(2p\ f_j t / T) \underset{(-4.04)}{- 0.795} \cos(2p\ f_j t / T) + \sum_{i=1}^{12} b_i\ y_{t-i} + e_t \tag{35}$$

Panel *b* of Figure 2 shows the time path of the intercept 1.12 - 0.886sin($2\pi f_j t/T$) - 0.795 cos($2\pi f_j t/T$). It should be clear the time-varying intercept captures the behaviour of the inflation rate during the 1970's. As the inflation rate jumped in the late 1960's (see Panel *a*), the time-varying intercept rises above 1.12 (see Panel *b*). Similarly, as the inflation rate falls from its early-1980's peak, the time-varying intercept in shown in Panel *b* begins its decline.

7.2 German per-capita income

In order to demonstrate the use of the *Trig*-test in its correct bootstrapped form, seasonally unadjusted quarterly per-capita income and consumption data for Germany from 1960.I to 1988.IV (116) observations are used.[14] Lütkepohl and Herwartz (1996) use these data over the identical time period to illustrate a modeling strategy known as flexible

---

[13] Clearly, the use of integer frequencies only yielded f_j = 1.0.

[14] The authors wish to thank Hiltrud Nehls of the Rhine-Westphalia Institute for Economic Research for providing the data.

generalized least squares (FGLS); as such, we can provide a useful benchmark against which to compare the results of the *Trig*-test.

The first log difference of the quarterly per-capita income data is plotted in Figure 3. There appears to be a strong seasonal pattern, with growth rates being negative in the first quarter and highest in the fourth quarter. The modeling problem posed by these data is that this pattern is capable of being generated by a variety of models. For example, a model with time-varying autoregressive parameters, an AR[4] model with constant parameters or a model with a time-varying intercept are all examples of data generating processes capable of producing the seasonal pattern in the observed income data.

If we estimate a standard AR(4) model, we obtain:

$$\underset{}{y_t} = \underset{(3.07)}{0.012} - \underset{(-3.28)}{0.181} y_{t-1} - \underset{(-3.35)}{0.183} y_{t-2} - \underset{(-3.51)}{0.192} y_{t-3} + \underset{(14.60)}{0.803} y_{t-4} + e_t \tag{36}$$

The point estimates suggest a strong seasonal pattern. However, an analysis of the residuals shows that the model is unsatisfactory. The correlogram of the residuals 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}
-0.047	0.075	0.342	-0.289	0.079	0.053	-0.072	-0.040	-0.030	0.126	-0.071	0.033

This conclusion concurs with that reached by Lütkepohl and Herwartz (1996). Using FGLS, they find that that the intercept displays a seasonal pattern. Given the pattern in the data, we apply the *Trig*-test to the intercept but leave the frequency unspecified. The frequency selected by the LR_{sup} statistic is 34.77 and presence of the trigonometric coefficients is significant that the 5% level. In a sense, we agree with Lütkepohl and Herwartz (1996) in finding a time-varying intercept. An interesting aspect of our results, however, is that the frequency $f = 34.77$ is too high to indicate quarterly seasonal behavior. It may be that the time dependence in the constant term is due to seasonality plus some other neglected influence on the behavior of per-capita income.

7.3 German per-capita consumption

The first log difference of German per-capita consumption is plotted in Figure 4 with a marked seasonal pattern once again in evidence. Just as for the income data, the null hypothesis of time invariance in the parameters of an AR(1) model is strongly rejected (*prob*-values of 0.00) with seasonal frequencies of 28 and 29 respectively recorded by the LR_{sup} tests on the intercept and autoregressive parameter. When testing the coefficients of an AR(4) model for time constancy, however, the results are markedly different to those obtained for per-capita income. In this instance, the null hypothesis is strongly rejected for the first, second and fourth autoregressive coefficients. This result is different from that reported by Lütkepohl and Herwartz (1996) who argue that the first and second autoregressive coefficients may be treated as constant, but the third and fourth should be modeled as time-varying.

It is also interesting to note that the frequency, 10, chosen by the sup-norm mapping with 111 usable observations in the model, hints at cycles in the autoregressive parameters of approximately 3 years. Following up on this information, when the AIC and SIC are used to find the optimal AR lag structure for a linear autoregressive model for these data, both criteria select lag 12. While it is not the aim of this paper to pursue the optimal time-series model for the data, the information obtained from the *Trig* testing strategy suggests that a low order periodic AR model, which can generate autocorrelation patterns at multiples of the number of seasons, is a suitable starting point for a more substantive investigation.

8. Conclusion

The framework presented in this paper develops a straightforward way to approach the problem 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 in parameters due to either structural breaks, seasonality or stochastic parameter variation. The only econometric problem encountered is that of an unidentified parameter under the null hypothesis. As a result, the test statistic has a non-standard distribution and it is necessary to assess the significance of the statistic by bootstrapping. Nevertheless, the test is implemented in the time domain and the critical values can be well-approximated by OLS.

The Monte Carlo experiments reported in the paper provide strong empirical evidence that the *Trig*-test has the correct size and good power against all the alternative models considered. The *Trig*-test detects one-time structural breaks almost as reliably as the Andrews, Lee and Ploberger test and can have better power than the test in cases with multiple breaks. In the context of seasonal models, it is demonstrated that the use of prior information about the periodicity of the seasonality results in tests that are more powerful than the *Trig*-test. In the absence of this information, the *Trig*-test can be a valuable tool for the detection of seasonality. The test also proved to have better power against stochastic parameter processes than other commonly used tests. This includes tests that were specifically designed to detect the type of process used in the simulation. The test was used to detect structural breaks in the U.S. *CPI* and seasonality in German income and consumption data. Our *Trig*-test captures the sharp rise in U.S. inflation rates in the 1970’s. Moreover, the test indicates that the parameter instability in the German income is likely due to one or more breaks while that in the consumption data is likely due to seasonality.

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.

Ashley, R.A. and Patterson D.M. (2000) "Nonlinear model specification/diagnostics: Insights from a battery of nonlinearity tests" Department of Economics Working Paper E99-05: Virginia Polytechnic University.

Baillie, R. T., Han, Y.W. and Kwon, T. (2002) "Further Long Memory Properties of Inflationary Shocks" *Southern Economic Journal*, 68, 496 – 510.

Barnett W.A., Gallant A.R., Hinich M.J., Jungeilges J.A., Kaplan D.T. and Jensen M.J. (1997) "A single-blind controlled competition among tests for nonlinearity and chaos" *Journal of Econometrics*, 82, 157-192.

Bergström P. (1999) "Bootstrap methods and applic ations in econometrics - a brief survey" mimeo, Department of Economics, Uppsala University.

Bose A. (1988) "Edgeworth correction by bootstrap an autoregressions" *The Annals of Statistics*, 16, 1709-1722.

Brown R.L., Durbin J. and Evans, J.M. (1975) "Techniques for testing the constancy of regression relationships over time" (with discussion) *Journal of the Royal Statistical Society* B, 37, 149-192.

Clements, M.P. and Hendry D.H. (1999) *Forecasting Non-stationary Economic Time Series*. MIT Press: Cambridge.

Clements, M.P. and Hendry D.H. (1999b) "On winning forecasting competitions in economics." *Spanish Economic Review* 1. 123 – 160.

Clements, M. and Mizon, G.E. (1991). "Empirical Analysis of Macroeconomic Time Series." *European Economic Review* 35. 887-932.

Dahl, C.M. (1999) "An investigation of tests for linearity and the accuracy of flexible nonlinear inference", Working Paper 1999-8, Department of Economics, University of Aarhus.

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

Engle, R. F. (1974) "Band Spectrum Regression" *International Economic Review*, 15, 1-11.

Flachaire E. (1999) "A better way to bootstrap pairs", *Economics Letters*, 64, 257-262.

Gallant A. R. and Souza, G. (1991) "On the asymptotic normality of fourier flexible form estimates" *Journal of Econometrics*, 50, 329-53.

Hall, S.G. (1986). 'An Application of the Engle-Granger Two-Step Estimation Procedure to United Kingdom Aggregate Wage Data.' *Oxford Bulletin of Economics and Statistics*, 48. 229-39.

Hansen B. (1996) "Inference when a nuisance parameter is not identified under the null hypothesis" *Econometrica*, 64, 413-430.

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

Harvey A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press: Cambridge.

Hatanaka, Michio (1996). *Time-Series Based Econometrics*. (Oxford University Press: Oxford).

Hendry D.F. (1995) *Dynamic Econometrics*, Oxford University Press: Oxford.

Herwartz H. (1995) Analyse saisonaler Zeitreihen mit Hilfe periodischer Zeitreihenmodelle, Verlag Josef Eul: Bergisch-Gladbach.

Horowitz, J.L. (1997) "The bootstrap" forthcoming in *Handbook of Econometrics*, Volume V, Elsevier Science: Amsterdam.

Lee, T., White, H. and Granger, C.W.J., (1993) "Testing for neglected nonlinearity in time series models", *Journal of Econometrics*, 56, 269-290.

Li H. and Maddala G.S (1996) "Bootstrapping time series models" *Econometric Reviews*, 15, 115-195 (incl. discussion).

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

Lütkepohl H. and Herwartz H. (1996) "Specification of varying coefficient time series models via generalized flexible least squares" *Journal of Econometrics*, 70, 261-290.

Mammen E. (1993) "Bootstrap and wild bootstrap for high dimensional linear models", *The Annals of Statistics*, 21, 255-285.

Nyblom, J. (1989) "Testing for the Constancy of Parameters over Time" *Journal of the American Statistical Association,* 84, 223-30.

Perron, P. (1989) "The Great Crash, the oil price shock and the unit root hypothesis" *Econometrica*, 57, 1361-1401.

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

Tan H.B. and Ashley R. (1999) "Detection and modeling of regressioon parameter variation across frequencies with an application to testing the permanent income hypothesis" *Macroeconomic Dynamics*, 3, 69-83.

Teräsvirta T., Lin C.F. and Granger C.W.J. (1993) "Power of the neural network linearity test" *Journal of Time Series Analysis*, 14, 209-220.

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 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

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

Table 3: Power against seasonal parameter variation.

	p-value	SP1	SP2 (a_0)	SP2 (a_1)	SP3(a_0)	SP3 (a_1)
LR_{sup}	*0.01*	0.922	0.118	0.206	0.341	0.955
	0.05	0.965	0.230	0.344	0.521	0.978
	0.10	0.980	0.317	0.445	0.607	0.986
LR_{ave}	*0.01*	0.605	0.000	0.041	0.002	0.459
	0.05	0.812	0.000	0.153	0.006	0.716
	0.10	0.881	0.000	0.261	0.012	0.824
LR_{exp}	*0.01*	0.925	0.114	0.207	0.335	0.954
	0.05	0.967	0.227	0.347	0.517	0.978
	0.10	0.983	0.304	0.451	0.600	0.987
t-test	*0.01*	0.942	0.983	0.777	0.727	0.899
	0.05	0.964	0.997	0.904	0.895	0.967
	0.10	0.975	1.000	0.942	0.947	0.985

Sample size = 100. Power computed in 1000 repetitions with 400 bootstraps per repetition for the *Trig*-test.

Table 4: 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

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

Table 5: Critical Values for the F_{Trig}^{OLS} statistic: $a_{0t} = a_t$

	T = 50	T = 100	T = 250	T = 1000
10%	5.81	6.37	7.17	8.53
5%	6.71	7.19	7.94	9.25
1%	8.87	9.09	9.72	10.95

Integer search for a maximum frequency equal to the integer value of $T/2$ -1

Table 6: The effect of Autoregressive Processes on the Critical Values for the F_{Trig}^{OLS} statistic

Parameter Value	AR(1) H_0: $a_{1t} = a_1$	AR(2) H_0: $a_{1t} = a_1$	AR(2) H_0: $a_{2t} = a_2$	AR[4] H_0: $a_{4t} = a_4$
0.1	7.206	7.282	7.280	7.265
0.2	7.206	7.320	7.274	7.265
0.3	7.248	7.322	7.277	7.254
0.4	7.237	7.304	7.273	7.249
0.5	7.236	7.305	7.265	7.255
0.6	7.273	7.331	7.289	7.269
0.7	7.284	7.337	7.299	7.271
0.8	7.292	7.346	7.308	7.287
0.9	7.278	7.314	7.303	7.314

The table reports the 5% critical values for T = 100 using an integer search for a maximum frequency equal to the integer value of $T/2$ –1. The entries are to be compared to the value 7.19 reported in Table 5.

Figure 1: Four Fourier Approximations

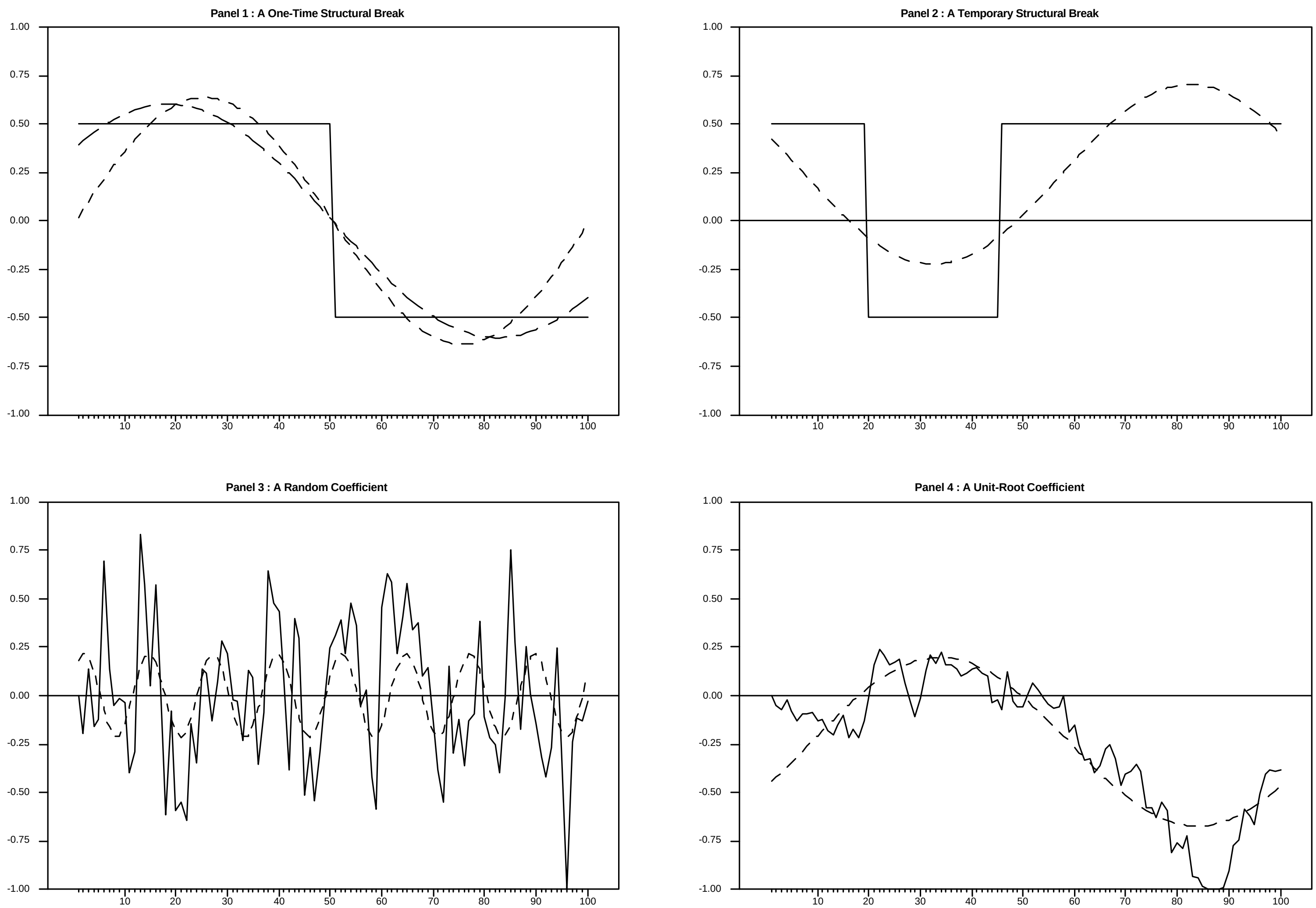


Figure 2: A Structual Break in U.S. Inflation?

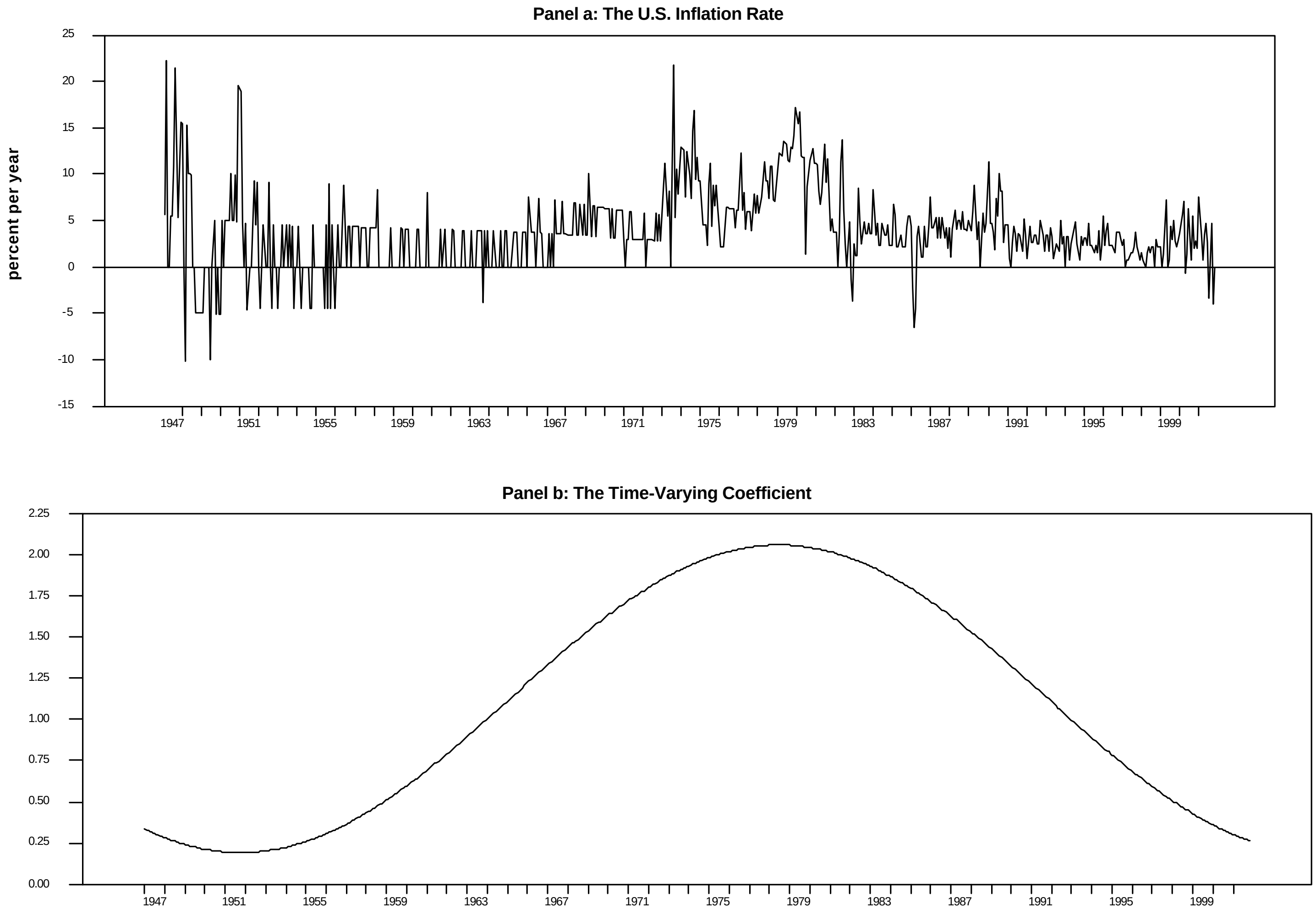


Figure 3:
Quarterly growth rate of German per-capita income 1960:2 to 1988:4

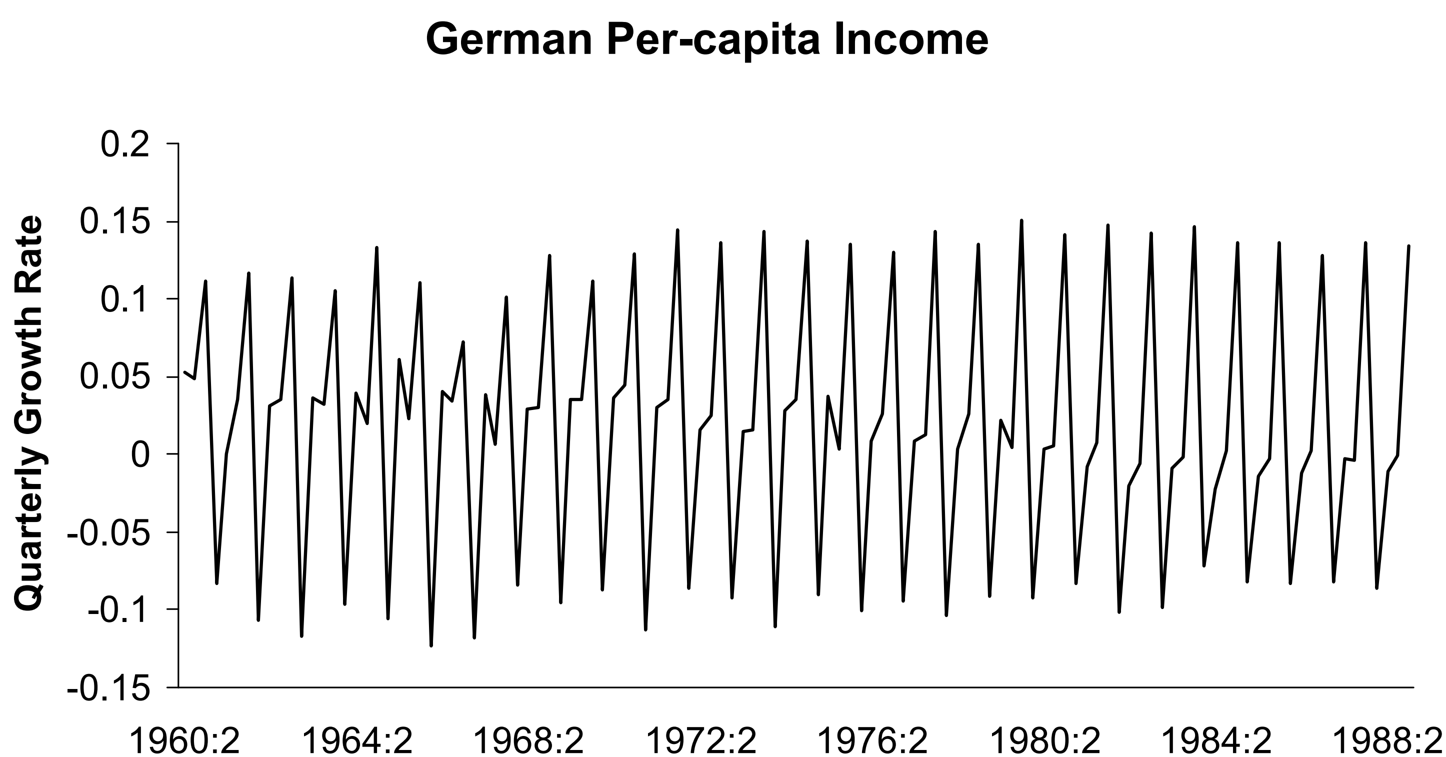


Figure 4:
Quarterly growth rate of German per-capita consumption 1960:2 to 1988:4