On the Reliability of Chow Type Test for Parameter Constancy in Multivariate Dynamic Models [1]

by

Bertrand Candelon and Helmut Lütkepohl[2]
Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät,
Humboldt-Universität, Spandauer Str. 1, D-10178 Berlin, GERMANY

Abstract

The small sample properties of two types of Chow tests are investigated in the context of multiple time series models. It is found that the tests may have substantially distorted size if the sample size is not large relative to the number of parameters in the model under study. In particular the tests reject far too often in this situation. It is shown that bootstrap versions of the tests have much better properties in this respect. In other words, the bootstrap can be used to size-adjust the tests.

Key Words: Bootstrap, vector autoregressive process, vector error correction model, stability tests

JEL classification: C32, E41, E43

[1]Financial support was provided by the DFG, Sonderforschungsbereich 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213).

[2]Corresponding author: Helmut Lütkepohl, Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität, Spandauer Str. 1, D-10178 Berlin, GERMANY, tel. +49-30-2093 5718, fax. +49-30-2093 5712, email: luetke@wiwi.hu-berlin.de

1 Introduction

In econometric studies based on time series data, it is common practice to test for structural change during the observation period. Chow tests are standard tools for this purpose. This study serves to point out problems related to using these tests for checking the structural stability of vector autoregressive (VAR) models. It will be demonstrated by simulation that in the small sample situations where these tests are often applied, the distributions of the test statistics under the stability hypothesis may be substantially different from the assumed asymptotic χ^2 distributions. In some cases the distortions are so large that the tests become useless for practical purposes. We will consider bootstrap modifications of the tests which turn out to be much more reliable in small samples.

The observation that Chow tests may have distorted distributions relative to the asymptotic χ^2 or approximate F distributions in dynamic models is not new (see, e.g., Diebold & Chen (1992) for related results and discussions). However, in systems of equations the problem becomes so dramatic that we find it worthwhile to demonstrate and analyze it for this class of models. Because we are considering a small sample problem, the dependence on the process parameters makes it difficult to obtain general results. Therefore, to demonstrate the relevance for applied work, we consider empirical models from the literature for which the tests have been applied and we demonstrate the small sample distortions within the context of these models by using simulations based on estimated models. In particular we use a bivariate system of Danish interest rates from Engsted & Tanggaard (1994) and Engsted & Nyholm (2000) and a five-dimensional German monetary macro system from Juselius (1996, 1998) to illustrate the problems of standard Chow tests. A bootstrap modification is shown to have superior properties for the example models. In fact, for the example models considered, the bootstrap versions of the tests have roughly correct size even in relatively small samples.

This study is organized as follows. In the next section, the test versions which we will consider are summarized formally and a bootstrap modification is presented. The small sample performance of the tests is analyzed in Sec. 3 on the basis of the example models from the literature and conclusions are drawn in the final section.

2 The Chow Tests

Given a set of n time series variables $y_t = (y_{1t}, \ldots, y_{nt})'$, the basic VAR($p$) model considered in the following has the form

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \qquad (t = 1, \ldots, T), \tag{1}$$

where ν is a constant $(n \times 1)$ vector, the A_i are $(n \times n)$ coefficient matrices and $u_t = (u_{1t}, \ldots, u_{nt})'$ is an unobservable zero mean white noise process with time invariant positive definite covariance matrix Σ, i.e., $u_t \sim iid(0, \Sigma)$. It is straightforward to introduce further deterministic terms such as seasonal dummy variables or polynomial trend terms in the model or include further exogenous variables. We use the simple model form (1) mainly for convenience because it simplifies the presentation of the tests. Moreover, if the system contains integrated and cointegrated variables it may be preferable to analyze the model in error correction form. Again such an extension is straightforward and will not be considered here in detail. In any case, if the unrestricted levels form of the model in (1) has time varying coefficients,the corresponding error correction version cannot be time invariant. Hence, a stability test may in fact be performed on the levels form. Clearly, the model (1) is in reduced form because all right-hand side variables are predetermined or deterministic and no instantaneous relations are modeled. This is sufficient for the purposes of stability tests because the structural form cannot be stable if the reduced form is unstable. Therefore stability tests may be based on the latter.

We consider two versions of Chow tests, sample-split tests and break-point tests (see Doornik & Hendry (1997)). It is assumed that a structural break may have occurred in period T_B. The model under consideration is estimated from the full sample of T observations and from the first T_1 and the last T_2 observations, where $T_1 < T_B$ and $T_2 \leq T - T_B$. The resulting residuals are denoted by $\hat{u}_t$, $\hat{u}_t^{(1)}$ and $\hat{u}_t^{(2)}$, respectively. Moreover, define $\hat{\Sigma} = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$, $\hat{\Sigma}_{1,2} = T_1^{-1} \sum_{t=1}^{T_1} \hat{u}_t \hat{u}_t' + T_2^{-1} \sum_{t=T-T_2+1}^{T} \hat{u}_t \hat{u}_t'$, $\hat{\Sigma}_{(1)} = T_1^{-1} \sum_{t=1}^{T_1} \hat{u}_t^{(1)} \hat{u}_t^{(1)'}$ and $\hat{\Sigma}_{(2)} = T_2^{-1} \sum_{t=T-T_2+1}^{T} \hat{u}_t^{(2)} \hat{u}_t^{(2)'}$. Using this notation, the two test statistics can be written as follows:

Sample-Split (SS) Test

$$\lambda_{SS} = (T_1 + T_2) \log\det \hat{\Sigma}_{1,2} - T_1 \log\det \hat{\Sigma}_{(1)} - T_2 \log\det \hat{\Sigma}_{(2)} \approx \chi^2(k), \tag{2}$$

where k is the number of restrictions imposed by assuming a constant coefficient model for the full sample period, that is, k is the difference between the sum of the number of coefficients estimated in the first and last subperiods and the number of coefficients in the full sample model. The parameter constancy hypothesis is rejected if the value of the test statistic λ_{SS} is large.

Break-Point (BP) Test

$$\lambda_{BP} = \frac{1-(1-R_r^2)^{1/s}}{(1-R_r^2)^{1/s}} \cdot \frac{Ns-q}{nk} \approx F(nk, Ns-q), \tag{3}$$

where

$$s = \left(\frac{n^2k^2-4}{n^2+k^2-5}\right)^{1/2}, \quad q = \frac{nk}{2}+1, \quad N = T-k_1-k-(n-k+1)/2$$

with k_1 being the number of regressors in the restricted, stable model and

$$R_r^2 = 1-\left(\frac{T_1}{T}\right)^n |\hat{\Sigma}_{(1)}|(|\hat{\Sigma}|)^{-1}.$$

Again the null hypothesis of parameter constancy is rejected for large values of λ_{BP}. This test is included in the software package *PcFiml* and is therefore used occasionally in empirical work (e.g., Beyer (1998)).

Bootstrap Tests

Bootstrap versions of the tests are obtained by estimating the model of interest, denoting the estimation residuals by $\hat{u}_t$, computing centered residuals $\hat{u}_1 - \bar{u}_., \dots, \hat{u}_T - \bar{u}_.$, where $\bar{u}_. = T^{-1}\sum \hat{u}_t$, and generating bootstrap residuals $u_1^*, \dots, u_T^*$ by randomly drawing with replacement from the centered residuals. These quantities are then used to compute bootstrap time series recursively starting from given presample values $y_{-p+1}, \dots, y_0$. The model of interest is then reestimated with and without stability restriction and a bootstrap version of the statistic of interest, say λ_{SS}^* or λ_{BP}^* is computed. Repeating these steps a large number of times (say M times), a critical value is then obtained as the relevant percentage point, say λ_{crit}^*, from the empirical distribution of the bootstrap test statistic and the stability hypothesis is rejected if $\lambda > \lambda_{crit}^*$. Alternatively, the p-value of the test may be estimated as the fraction of times the value of the bootstrap statistic exceeds λ.

In this particular case the bootstrap can be justified by asymptotic theory if suitable regularity conditions hold, because the statistic is a continuous function of sample moments and the statistic is asymptotically pivotal (see Horowitz (1999) for details). Of course, the theoretical result does not necessarily guarantee a satisfactory performance of the bootstrap in small samples although Horowitz (1999) reports examples of related cases where the bootstrap worked very well in small samples. In the next section we will use simulations to explore the small sample properties of the tests reviewed in this section, including the bootstrap versions.

3 Monte Carlo Simulations

Because simulations of bootstrap methods are very computer intensive, we first use small systems (bivariate models with respectively one and three lags) in the Monte Carlo study based on two Danish interest rates. To illustrate the impact of increasing the dimension of the system under consideration we then also perform simulations for a model based on a five-dimensional German macro system.

3.1 Danish Interest Rates

The first example is based on monthly Danish interest rate data for 1976(1) to 1991(12) so that we have $T = 192$ observations.[3] They were also analyzed by Engsted & Tanggaard (1994) and Engsted & Nyholm (2000). Specifically the series consist of 1-month (r_t) and 3-month (R_t) Danish zero-coupon bond yields. In the original paper, the authors consider a system including the spread $S_t = R_t - r_t$ and the changes in the short rate $\Delta r_t = r_t - r_{t-1}$. Therefore we also use a model for these variables. Based on a Markov regime switching analysis, Engsted & Nyholm (2000) found a structural break in the system around August 1983. Because they use a VAR(1) model we also consider that model in part of our analysis. In addition we report results for simulations based on a VAR(3) model for the two series to study the impact of the model size on the properties of the tests.

For analyzing the sizes of the Chow tests we use the data to fit constant coefficient VAR(1) and VAR(3) models for the full period. The resulting estimated processes are

$$\begin{bmatrix} S_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} 0.0272 \\ -0.2145 \end{bmatrix} + \begin{bmatrix} 0.8045 & 0.0120 \\ 1.5175 & 0.0470 \end{bmatrix} \begin{bmatrix} S_{t-1} \\ \Delta r_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \qquad (4)$$

with covariance matrix

$$\hat{\Sigma} = \begin{bmatrix} 0.0107 & \cdot \\ -0.0687 & 0.9433 \end{bmatrix}$$

[3] The data are available from C. Tanggaard's homepage: http://www.hda.dk/cat/Datafiles/dkrente.txt.

and

$$\begin{bmatrix} S_t \\ \Delta r_t \end{bmatrix} = \begin{bmatrix} 0.0194 \\ -0.2052 \end{bmatrix} + \begin{bmatrix} 0.5168 & -0.0136 \\ 2.1301 & 0.0232 \end{bmatrix} \begin{bmatrix} S_{t-1} \\ \Delta r_{t-1} \end{bmatrix} + \begin{bmatrix} 0.2371 & -0.0031 \\ -0.0611 & 0.0839 \end{bmatrix} \begin{bmatrix} S_{t-2} \\ \Delta r_{t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} -0.0136 & -0.0067 \\ 0.0232 & -0.0288 \end{bmatrix} \begin{bmatrix} S_{t-3} \\ \Delta r_{t-3} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

with covariance matrix

$$\hat{\Sigma} = \begin{bmatrix} 0.0100 & \cdot \\ -0.0652 & 0.8864 \end{bmatrix}.$$

We simulated 10,000 sets of time series with different lengths (T=60 and 100 plus presample values) using these processes. We then applied the standard tests for checking different break dates. We also applied the bootstrap tests to 1,000 of the generated time series using $M = 1,000$ bootstrap draws in each of the 1,000 applications of the tests. In Table 1 rejection frequencies for the tests based on a nominal significance level of 5% are presented for a range of different break dates. Here $T_1 = T_B$ and $T_2 = T - T_1 - p - 1$.

Obviously, for moderate sample sizes the size distortions of the original version of the SS test is unacceptable even if the VAR(1) model with relatively few parameters is considered. The test performance deteriorates further if a higher order VAR(3) model is used in the simulations. In contrast the BP test and both bootstrap versions result in rejection frequencies which are much better in line with the desired significance level, although λ_{BP} rejects somewhat too often if a break close to the beginning of the sample is tested (see $T_B/T = 0.2$). Both bootstrap test versions have almost ideal size properties even for a sample size as small as $T = 60$ and a VAR order of 3. For the original test versions much larger sample sizes are necessary for getting such a result. The figures in Table 1 show that $T = 100$ is generally not sufficient. For larger samples (more than 1,000 observations) we obtain sizes close to 5%.[4] Clearly, based on these results, the original SS test is useless for small sample situations because it is unable to control the size at least approximately even for a very simple DGP whereas the bootstrap versions of both tests and to some extend also the original BP test have better properties.

In the light of these results we have used the bootstrap versions of the tests on the original data. Given the previous simulation results, one would expect that in this situation these versions of the Chow tests are fairly reliable in terms of size. We report the estimated p-values of the tests for a series of hypothesized break dates in Figure 1. Clearly, the SS test

[4]Results are available from authors upon request.

rejects at a 5% level of significance at almost all break dates, that is, the estimated p-values are clearly less than 5% except at the very beginning and at the very end of the sampling period. In contrast, the BP test version has estimated p-values well above 5% and, hence, if that test version is applied, the stability hypothesis cannot be rejected. These results indicate that the two types of tests have quite different power properties because one of the tests finds a break whereas the other one does not. Note, however, that the tests give no indication of where the break has occurred or whether there is perhaps more than one break.

To explore the power of the tests we have simulated the VAR(1) DGP (4) with a shift of 0.05 in the intercept term of each equation at a fixed date. Then, we have applied the SS and BP tests using various different null hypotheses not always corresponding to the actual break date. Thereby we hope to gain insights in the ability of the tests not only to detect a break in some period but also into their ability of finding the actual break date. The rejection frequencies of all the tests are presented in Table 2. The figures in the table are again based on 10,000 replications for the original tests and on 1,000 replications of the bootstrap tests for which we use again $M = 1,000$ bootstrap draws within each of the 1,000 replications of the experiment. Notice, that the original tests are based on χ^2 and F critical values and, hence, suffer from size distortion.

Obviously, the tests have very different power properties. Given that no size adjustment has been made, the slightly larger power of the original SS tests is not surprising. In fact, it is perhaps a bit surprising that the bootstrap version has almost as much power as the original version although it has a much lower rejection frequency under the null hypothesis. Thus this test has a much better ability to discriminate between stable and unstable processes than the original counterpart. Notice, however, that none of the two versions of the SS test is very well able to detect the true break date. In other words, if there is a break both tests tend to reject the stability null hypothesis even if the break date is misspecified under the null. In contrast, the BP tests tend to reject the stability hypothesis only if the break date specified in the null hypothesis lies before the actual break date. This behaviour of the test is quite plausible because this test can be interpreted as a comparison of the predictions from the model fitted to the sample before the hypothesized break date with the actual observations from the period after the break. Clearly, as long as the break occurs after the hypothesized break date, the predictions will differ markedly from the observed values. If the break has occurred before the hypothesized date, however, the forecasts will be poor with large uncertainty related to them and, hence, the test has trouble to detect that the actual observations are from some other DGP. Both tests tend to have smaller power for breaks towards the beginning of the sample than for breaks in the middle or at the end.

In summary, the power results in Table 2 indicate that the bootstrap tests have similar or even better power than the original versions of the tests in many situations even without correction for the size distortions of the original tests. Whereas the SS tests are not able to clearly indicate the break point, the BP tests may give at least an indication where the break may have occurred. Thus, the SS tests are suitable only for a global stability analysis.

These results can explain the fact that the SS test rejects for many different break points for the example system in Figure 1. They cannot explain well the differences in the p-values of the two types of tests, however. Of course, one possible explanation may be that the break is of a different nature than the one in our simulation where a simple shift in the constant term is used. Clearly, changes in the other coefficients of the DGP such as the residual covariance or the VAR coefficients are possible as well and the tests may have different power against such changes.

3.2 German Monetary System

Small systems of the monetary sectors of three European countries were considered by Juselius (1996, 1998). Here we will focus on her model for Germany because she found a structural break for this country using a sample-split test. The following variables are considered: m_t is the logarithm of M3, y_t denotes the logarithm of real Gross Domestic Product (GDP), p_t is the log GDP deflator, r_t denotes the yearly rate on private bank deposits and R_t is the yearly rate of the effective yield on bonds in circulation. The set of variables used in the study is then $y_t = (y, m-p, \Delta p, R, r)'_t$. In addition to these stochastic variables the following deterministic terms are included in Juselius' model: a constant, a linear time trend, seasonal dummies, a step dummy for the German monetary unification which has the value 1 starting in the third quarter of 1990 and is zero elsewhere, an impulse dummy for the second quarter of 1990 and an extra set of seasonal dummies for the period after the German monetary unification. Juselius uses quarterly, seasonally unadjusted data for the period 1975(3) - 1994(4)[5] and she fits a vector error correction model with two lags of the differenced variables and a cointegrating rank of 2. She diagnoses a structural break in 1983 and supports this finding with a sample-split test among other means. Clearly, given the simulation results based on the Danish interest rate system, size distortions are expected in a system of the present size.

[5] The data are available on http://ise.wiwi.hu-berlin.de/oekonometrie/index.html.

Therefore we have also simulated the stochastic part of Juselius' model obtained by fitting a constant coefficient model to the full sample including a constant but deleting the other deterministic terms. For this purpose we used an unrestricted VAR(3) model.[6] Simulation results are given in Table 3 based on the same number of replications as in the Danish interest rates case (i.e., 10,000 sets of time series are simulated, 1,000 bootstrap tests are performed from 1,000 of the generated time series using $M = 1,000$ bootstrap draws each time). Massive size distortions are observed for both original tests when used with χ^2 and F critical values for λ_{SS} and λ_{BP}, respectively. Even with $T = 300$, observations the SS tests have unacceptable rejection frequencies in access of 20%. Although the BP test is somewhat better for the larger sample size, its performance is clearly unacceptable for the smaller sample of $T = 76$. Thus, it is clear that both original tests cannot be recommended for typical macroeconometric applications, for example. Again, the performance of the bootstrap versions is much better. Their rejection frequencies are close to the ideal 5% even for the smaller sample size. Thus, the bootstrap may be used as a possibility for size adjusting the tests.

Obviously, given our simulation results, it is clear that the SS test used by Juselius (1998) in checking the stability of the system may be spurious and due to the massive size distortion of the test. Therefore we have applied the bootstrap versions of the tests to the model fitted to the original data, including this time all the deterministic terms from Juselius' original study. In this exercise we also used an unrestricted VAR(3) model. Because such a model is even less restricted than the vector error correction model used by Juselius, structural instability due to model misspecification should be less likely in this model than in the one used by Juselius. The p-values of the bootstrap versions of the tests for different break dates are presented in Figure 2. It turns out that the p-values of the bootstrap-SS test are below 5% in the mid 80s so that there is some indication of a break in the sampling period. Given the results of the power simulations for the bivariate VAR(1) model it is not clear where the break has occurred, however. Again the bootstrap-BP test is not very helpful in dating the break because its p-values are far away from any common significance level of a usual test. In other words, based on this test, rejection of the stability hypothesis is not possible in any period. One possible explanation for this result may be obtained from the power study of the previous subsection. It was found that the power of the tests is low for breaks close to the sample beginning and the BP test does not have power for breaks that have occurred

[6] The precise model used in the simulations is available upon request.

before the date specified in the null hypothesis. Thus, if the break has occurred early on, e.g., at the beginning of the 1980s, the p-values in Figure 2 are not implausible.

4 Conclusions

In conclusion, we have demonstrated in this study that standard Chow tests may have massive size distortions in situations where they are often applied in practice. More precisely, they reject the stability hypothesis far too often for multivariate dynamic models with many parameters relative to the number of available observations. We have illustrated the problem using two small macroeconomic systems that have been considered previously in the literature. The size distortions of the tests are so large that in our view the tests are useless for applied work unless a size adjustment is made. Of course, this also means that the original tests should not be used recursively as is sometimes done in applied work. A size adjustment may be based on the bootstrap and we have pointed out how that may be done. Although the computational burden is quite reasonable if a single hypothesis is tested, a recursive procedure may quickly become expensive in terms of computer time. Therefore, instead of performing a fully recursive check for each sample point one may want to focus on a few individual time periods throughout the sample period.

In a very limited power simulation we found that the sample-split test is not suitable for detecting the actual break date. If it rejects the stability hypothesis this indicates that a break may have occurred somewhere in the sample. It is not clear where it has occurred. The break-point test may be more suitable for dating the break. It may have considerably more power if the break date specified in the null hypothesis is before the actual break date than if it is after the actual break. This property of the test may help in getting some indication of the break point.

References

Beyer, A. (1998), Modelling Money Demand in Germany, *Journal of Applied Econometrics*, 13, 57-76.

Diebold, X.D. & C. Chen (1992), *Testing structural Stability with Endogenous Breakpoint: A Size Comparison of Analytic and Bootstrap Procedures*, Mimeo University of Pennsylvania.

Doornik, J.A. & D.F. Hendry (1997), *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*, London: International Thomson Business Press.

Engsted, T. & K. Nyholm (2000), Regime Shifts in the Danish Term Structure of Interest Rates, *Empirical Economics*, 25, 1-13.

Engsted, T. & C. Tanggaard (1994), A Cointegration Analysis of Danish zero-coupon Bond Yields, *Applied Financial Economics*, 4, 265-278.

Horowitz, J.L. (1999), The Bootstrap, in *Handbook of Econometrics*, Vol. 5.

Juselius, K. (1996), An Empirical Analysis of the Changing Role of the German Bundesbank after 1983, *Oxford Bulletin of Economics and Statistics*, 58, 791-819.

Juselius, K. (1998), Changing Monetary Transmission Mechanisms within the EU, *Empirical Economics*, 23, 455-481.

Table 1. Relative Rejection Frequencies of Tests for Structural Change Based on Bivariate Model for a Nominal Significance Level of 5%

VAR order	Test	T_B/T ($T=60$) 0.2	0.4	0.5	0.6	0.7	T_B/T ($T=100$) 0.2	0.4	0.5	0.6	0.7
1	λ_{SS}	17.89	12.13	12.07	12.27	19.42	12.15	9.43	8.40	9.39	12.73
	λ_{SS}^*	5.20	4.20	4.00	4.40	6.20	5.00	5.80	4.80	5.20	6.20
	λ_{BP}	11.13	7.85	7.49	7.07	5.92	9.18	6.45	6.56	5.98	5.67
	λ_{BP}^*	6.60	6.00	4.80	4.80	5.00	5.80	4.80	5.20	4.00	5.80
3	λ_{SS}	53.35	24.17	23.85	28.78	51.70	23.46	13.14	12.87	14.60	18.24
	λ_{SS}^*	6.60	4.80	4.00	4.40	3.80	6.20	5.60	6.80	5.40	5.20
	λ_{BP}	12.26	7.86	7.07	6.62	6.17	8.73	6.72	6.18	5.94	5.75
	λ_{BP}^*	5.60	5.00	6.20	5.20	6.40	4.60	7.60	7.40	4.40	5.20

Table 2. Power Analysis Based on Bivariate Model, Nominal Significance Level 5%

Break at T_B/T	Test	T_B/T ($T=60$) 0.2	0.4	0.5	0.6	0.7	T_B/T ($T=100$) 0.2	0.4	0.5	0.6	0.7
0.2	λ_{SS}	83.05	91.00	78.98	64.14	58.42	95.18	94.58	84.19	67.09	60.73
	λ_{SS}^*	56.65	77.82	64.04	44.00	24.01	90.21	90.64	79.60	57.07	37.43
0.4	λ_{SS}	65.17	96.65	88.88	75.50	71.00	76.59	99.75	96.81	87.31	82.73
	λ_{SS}^*	29.43	89.41	73.85	55.04	32.25	58.43	99.41	94.20	80.61	54.23
0.5	λ_{SS}	59.75	92.73	98.72	80.89	76.42	66.86	98.57	99.91	93.25	90.42
	λ_{SS}^*	30.62	83.45	96.43	60.21	40.22	50.00	98.00	100.00	88.00	62.00
0.6	λ_{SS}	54.84	88.72	97.31	99.55	95.42	60.01	95.76	99.54	99.96	97.73
	λ_{SS}^*	27.05	78.28	92.03	98.05	48.60	42.41	92.23	98.81	100.00	75.81
0.7	λ_{SS}	52.69	84.77	94.69	99.11	99.85	52.46	90.90	97.86	99.78	99.98
	λ_{SS}^*	24.42	67.21	87.42	96.48	99.02	33.61	85.20	96.82	99.67	100.00
0.2	λ_{BP}	50.44	0.04	0.07	0.11	0.28	52.16	0.01	0.03	0.03	0.13
	λ_{BP}^*	33.84	0.20	0.00	0.00	0.00	37.43	0.00	0.00	0.00	0.00
0.4	λ_{BP}	54.79	85.95	0.15	0.19	0.26	57.59	88.11	0.07	0.17	0.19
	λ_{BP}^*	36.22	80.64	0.20	0.20	0.40	44.82	84.41	0.40	0.00	0.20
0.5	λ_{BP}	56.00	87.91	95.19	0.35	0.41	58.67	89.06	96.41	0.34	0.44
	λ_{BP}^*	34.22	81.44	91.40	0.20	0.00	48.61	86.23	93.87	0.00	0.00
0.6	λ_{BP}	56.33	88.12	95.57	99.01	0.57	58.30	88.98	96.57	99.38	0.78
	λ_{BP}^*	36.88	84.20	94.63	98.69	0.80	48.81	87.83	97.22	99.27	0.61
0.7	λ_{BP}	55.41	87.36	95.58	98.95	99.89	57.24	88.25	96.10	99.19	99.92
	λ_{BP}^*	37.25	81.85	94.02	98.80	99.82	46.61	88.68	96.05	99.07	100.00

Note: Columns and lines indicate respectively the date of the fixed break and the date of the break detection.

Table 3. Relative Rejection Frequencies of Tests for Structural Change for DGP Based on German Data with Nominal Significance Level 5%

	T_B/T $(T = 76)$				T_B/T $(T = 300)$			
Test	0.3	0.4	0.5	0.6	0.3	0.4	0.5	0.6
λ_{SS}	100.00	99.49	99.27	99.97	27.07	20.09	19.92	21.35
λ_{SS}^*	5.20	4.40	6.60	5.40	4.60	5.80	4.40	4.80
λ_{BP}	44.11	17.97	12.53	9.86	6.66	6.66	5.53	5.72
λ_{BP}^*	5.30	5.00	6.20	5.40	5.40	5.60	7.00	5.60

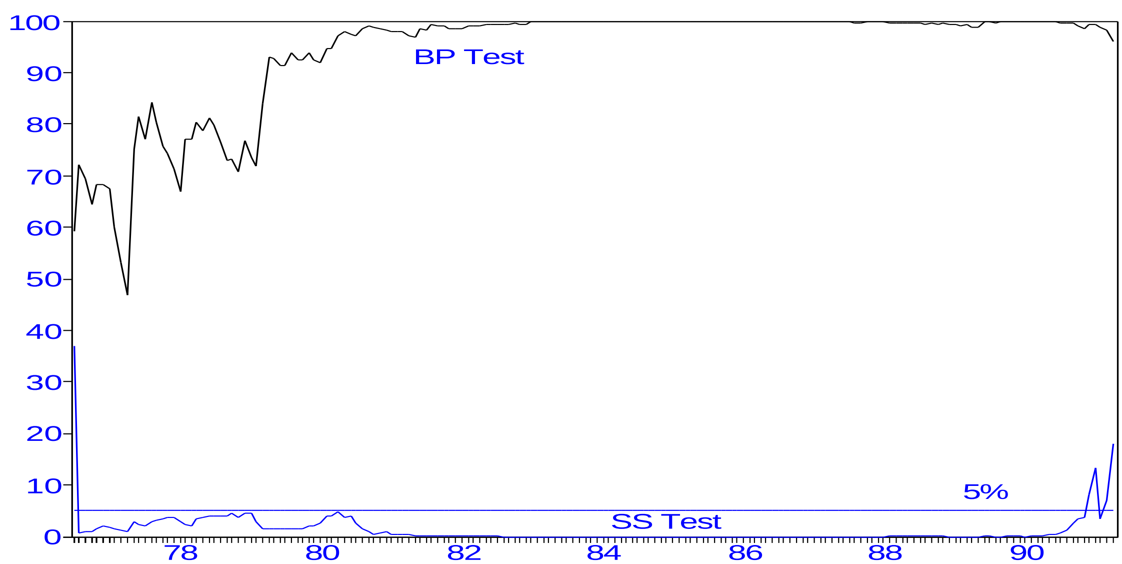

Figure 1. Danish Interest Rate System: p-values of Bootstrap Versions of the Tests for Structural Change

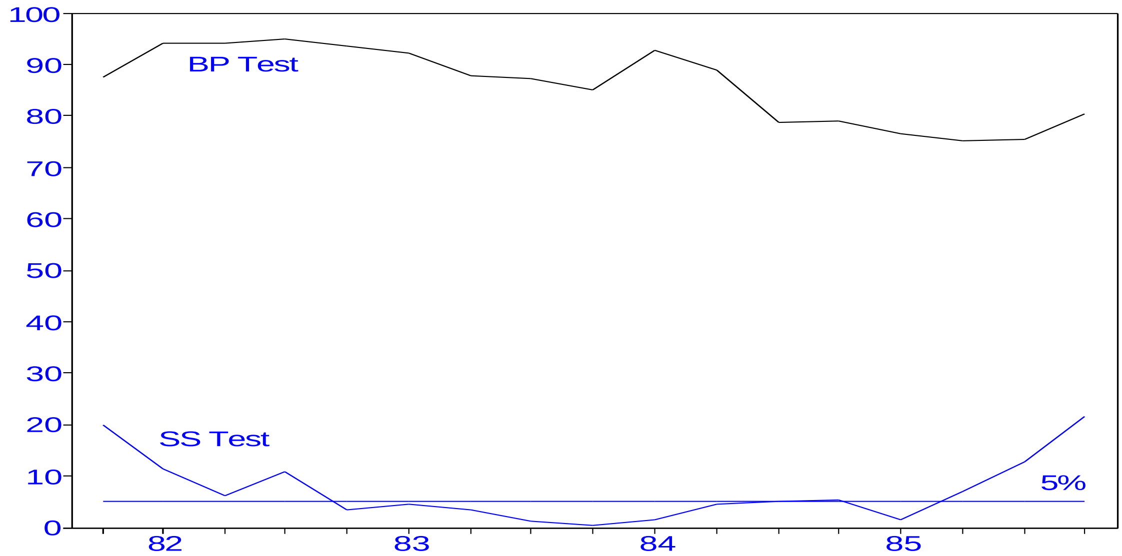

Figure 2. German Monetary System: p-values of Bootstrap Tests for Structural Change