

# On the Reliability of Chow Type Tests for Parameter Constancy in Multivariate Dynamic Models

Citation for published version (APA):

Candelon, B., & Lütkepohl, H. (2001). On the Reliability of Chow Type Tests for Parameter Constancy in Multivariate Dynamic Models. *Economics Letters*, 73, 155-160. [https://doi.org/10.1016/S0165-1765\(01\)00478-5](https://doi.org/10.1016/S0165-1765(01)00478-5)

## Document status and date:

Published: 01/01/2001

## DOI:

[10.1016/S0165-1765\(01\)00478-5](https://doi.org/10.1016/S0165-1765(01)00478-5)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



ELSEVIER

Economics Letters 73 (2001) 155–160

**economics  
letters**

www.elsevier.com/locate/econbase

# On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models

Bertrand Candelon, Helmut Lütkepohl\*

*Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität, Spandauer Str. 1, D-10178 Berlin, Germany*

Received 17 November 2000; accepted 19 April 2001

---

## Abstract

The small sample properties of two types of Chow tests are investigated in the context of multiple time series systems. 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. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords:* Bootstrap; Vector autoregressive process; Vector error correction model; Stability tests

*JEL classification:* C32; E41; E43

---

## 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  $\chi^2$  distributions. In some cases the distortions are so large that the tests become useless for practical purposes. We will consider bootstrap versions 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  $\chi^2$  or approximate  $F$  distributions in dynamic models is not new (see, e.g. Diebold and Chen (1992) for

---

\*Corresponding author. Tel.: +49-30-2093-5718; fax: +49-30-2093-5712.

*E-mail address:* luetke@wiwi.hu-berlin.de (H. Lütkepohl).

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 an empirical model from the literature for which Chow tests have been applied and we demonstrate the small sample distortions within the context of this model by using simulations based on an estimated model. A bootstrap modification of the Chow tests is shown to have superior properties for the example model. In fact, the bootstrap versions of the tests have roughly correct size even in relatively small samples.

## 2. The Chow tests

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

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad (t = 1, \dots, T) \quad (1)$$

where  $\nu$  is a constant ( $n \times 1$ ) vector, the  $A_i$  are ( $n \times n$ ) coefficient matrices and  $u_t = (u_{1t}, \dots, u_{nt})'$  is an unobservable zero mean white noise process with time invariant positive definite covariance matrix  $\Sigma$ , 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 (SS) tests and break-point (BP) tests (see Doornik and 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 SS test statistics can be written as follows:$

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

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  $\lambda_{SS}$  is large. The BP test statistic has the form:

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

where

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

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  $\lambda_{BP}$ . This test is included in the software package ‘PcFiml’ and is therefore used occasionally in empirical work.

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  $\lambda_{SS}^*$  or  $\lambda_{BP}^*$  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  $\lambda_{crit}^*$ , 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  $\lambda$ .

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 (2001) for details). Of course, the theoretical result does not necessarily guarantee a satisfactory performance of the bootstrap in small samples although Horowitz (2001) 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

To get realistic parameter values we base our simulations on a small German system of the monetary sector as considered by Juselius (1996,1998) who found a structural break 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)<sup>1</sup> 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.

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.<sup>2</sup> Simulation results are given in Table 1 based on 10 000 sets of simulated time series. For the bootstrap tests, 1000 replications are performed using 1000 of the generated time series with  $M = 1000$  bootstrap draws in each replication. Massive size distortions are observed for both original tests when used with  $\chi^2$  and  $F$  critical values for  $\lambda_{SS}$  and  $\lambda_{BP}$ , respectively. Even with  $T = 300$  observations the SS tests have unacceptable rejection frequencies in excess 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 macroeconomic 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 considered 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 Fig. 1. 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. It is not clear where the break has occurred, however. The  $P$ -values of the bootstrap-BP test are far

Table 1

Relative rejection frequencies of tests for structural change for DGP based on German data with nominal significance level 5%

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

<sup>1</sup>The data are available on <http://ise.wiwi.hu-berlin.de/oekonometrie/index.html>.

<sup>2</sup>The precise model used in the simulations is available upon request.

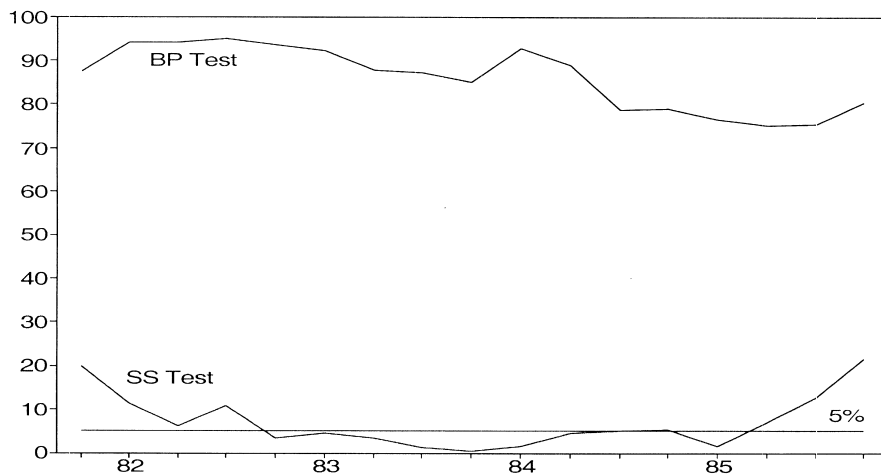


Fig. 1. German monetary system: *P*-values of bootstrap tests for structural change.

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.

#### 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 a small German macroeconomic system that has 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.

#### Acknowledgements

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

**References**

- Diebold, X.D., Chen, C., 1992. Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures. University of Pennsylvania, mimeo.
- Doornik, J.A., Hendry, D.F., 1997. Modelling Dynamic Systems using PcFiml 9.0 for Windows. International Thomson Business Press, London.
- Horowitz, J.L., 2001. The bootstrap. In: Heckman, J.J., Leamer, E.E. (Eds.). Handbook of Econometrics, Vol. 5. North-Holland, Amsterdam.
- 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.