Structural breaks and multivariate forecasting
James H. Stock

Maurice J.G. Bun, University of Amsterdam*

1 Introduction

In his lectures James Stock covers two main topics, i.e. structural breaks and dynamic factor models. The two topics are both in the field of time series econometrics, but are different of nature. Therefore, in this report a clear distinction between the two topics is made like in the lectures. The main topic in section 1 is testing on structural breaks. First two common models for structural breaks, i.e. the change point model and the time varying parameter model, are discussed. Second, the attention is focused on testing on structural breaks in the change point model, both when the break date known or unknown. Third, estimation of break dates is considered shortly. Section 2 is about dynamic factor models and discusses the estimation methods for this type of model. With a relatively small number of economic time series application of maximum likelihood using the Kalman filter is feasible. However, with many series things are more difficult.

2 Structural Breaks

2.1 Model

Central in the discussion of tests on structural breaks is the following regression model

$$y_t = x_t'\beta_t + \varepsilon_t, \qquad t = 1, ..., T \tag{1}$$

*Tinbergen Institute & Faculty of Economics and Econometrics, University of Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, Email: mbun@fee.uva.nl.

where y_t is regressed on a $K \times 1$ vector of stationary regressors x_t and ε_t is a white noise process with variance σ^2. We allow the vector x_t to contain predetermined variables, i.e. lagged values of y_t are possibly included as regressors.

When there are no structural breaks, i.e. $\beta_t = \beta$ for all t, the ordinary least squares estimator of β and σ^2 is consistent and efficient. However, problems arise when β_t does change over time. Changes in regression relationships over time can have several origins. First, changes in economic behaviour arise as a result of policy changes, i.e. the well-known Lucas critique. Second, there can be regime changes like the switch from floating to fixed exchange rates. Other examples are the reunification of West and East Germany in 1990 or the creation of the European Monetary Union.

To parametrize structural breaks several models are possible for the unknown parametervector β, i.e.

$$\begin{aligned} \beta_t &= \beta, \qquad t \leq k \\ \beta_t &= \beta + \delta, \qquad t > k \end{aligned} \tag{2}$$

where k is the date where the break occurs or

$$\beta_t = \beta_{t-1} + \eta_t, \qquad E\left[\eta_t \eta_t'\right] = \tau^2 G, \tag{3}$$

where η_t is serially uncorrelated and uncorrelated with ε_t. Model (1) with (2) is called the change point model and model (1) with (3) is the time varying parameter (TVP) model. Note that (2) is a special case of (3), i.e. through setting $\tau^2 = 0$, $\eta_t = \delta$ for $t = k + 1$ and $\eta_t = 0$ for $t \neq k + 1$. Also the no break model results from (3) when $\tau^2 = 0$ and $\eta_t = 0$ for all t.

2.2 Testing

There are numerous testing procedures for detecting structural breaks depending on the model for β and the status of the break date. If (2) holds, then the model becomes

$$\begin{aligned} y_t &= x_t'\beta + \varepsilon_t, \qquad t = 1, ..., k \\ y_t &= x_t'(\beta + \delta) + \varepsilon_t, \qquad t = k + 1, ..., T \end{aligned} \tag{4}$$

which can be written more compactly as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & O \\ O & X_2 \end{bmatrix} \begin{bmatrix} \beta \\ \beta + \delta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

In case of k known, the null of no structural break, i.e. $H_0 : \delta = 0$ can be tested with a Wald, LR or LM test, which are asymptotically equivalent and χ^2_K distributed under H_0. Assuming that $\varepsilon \sim N(0, \sigma^2 I_T)$ the loglikelihood function for model is

$$\log L(\beta, \delta, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{T}{2\sigma^2}(\varepsilon_1'\varepsilon_1 + \varepsilon_2'\varepsilon_2) \tag{5}$$

which is dependent of β and δ through the last term in (5). If σ^2 is known the following holds

$$\begin{aligned} W &= LR = LM \\ &= \frac{1}{\sigma^2}\left[\tilde{\varepsilon}'\tilde{\varepsilon} - (\hat{\varepsilon}_1'\hat{\varepsilon}_1 + \hat{\varepsilon}_2'\hat{\varepsilon}_2)\right] \end{aligned}$$

where $\tilde{\varepsilon}'\tilde{\varepsilon}$ and $(\hat{\varepsilon}_1'\hat{\varepsilon}_1 + \hat{\varepsilon}_2'\hat{\varepsilon}_2)$ are the restricted and unrestricted residual sum of squares. When σ^2 is unknown and has to be estimated also, the three test statistics are different in finite samples, but still asymptotically equal and χ^2_K distributed. For example, the Wald statistic becomes

$$W = \frac{\tilde{\varepsilon}'\tilde{\varepsilon} - (\hat{\varepsilon}_1'\hat{\varepsilon}_1 + \hat{\varepsilon}_2'\hat{\varepsilon}_2)}{(\hat{\varepsilon}_1'\hat{\varepsilon}_1 + \hat{\varepsilon}_2'\hat{\varepsilon}_2)/(T - 2K)} \tag{6}$$

Alternatively, a Chow (1960) F-test statistic can be used

$$F = \frac{1}{K}W \tag{7}$$

which in case of strict exogenous regressors has an exact F-distribution with K and $T-2K$ degrees of freedom. With predetermined regressors, no exact result is available. However, often critical values from the F-distribution are still used as an approximation, because K times the F-test statistic is asymptotically χ^2_K distributed.

When the break date k is unknown, which is likely to be the case in a lot of datasets, things are more difficult. Quandt (1960) proposes to minimize the likelihood ratio over k, i.e. choose k such that

$$\lambda_k = \frac{L(\beta, 0, \sigma^2)}{L(\beta, \delta, \sigma^2, k)} \tag{8}$$

is minimal, where $L(\beta, 0, \sigma^2)$ and $L(\beta, \delta, \sigma^2, k)$ are the restricted and unrestricted likelihood functions. It is obvious that the restricted likelihood is not depending k, because under the restrictions of the null hypothesis there is no break. Minimizing (8) is the same as maximizing

$$LR_k = -2\log\lambda_k \tag{9}$$

and the resulting test statistic is called the Quandt Likelihood Ratio (QLR) test statistic

$$QLR = \max_{k=k_0,\ldots,k_1} LR_k \tag{10}$$

The asymptotic distribution of the statistic in (10) is not χ^2_K, but a functional of Brownian motions, i.e.

$$QLR \Rightarrow \sup_{\lambda \epsilon [\lambda_0, \lambda_1]} \frac{B'(\lambda)B(\lambda)}{\lambda(1-\lambda)} \tag{11}$$

where $\Rightarrow$ means convergence in distribution, $\lambda_i = \lim_{T\to\infty} \frac{k_i}{T}$, $i = 1, 2$, $B(\lambda) = W(\lambda) - \lambda W(1)$ and $W(\lambda)$ a standard K-dimensional Brownian motion. A sketch of the proof can be found in Stock (1994) with references to more detailed proofs. The proof uses the expression of the Wald test statistic in (6) as starting point for the derivation. First, the asymptotic behavior of the numerator is established with the help of Functional Central Limit Theory (FCLT). Second, the denominator in (6) converges to σ^2. Third, with the help of the Continuous Mapping Theorem (CMT) the asymptotic distribution of the maximum of a series of χ^2 statistics is derived, which is the result in (11). Note that for fixed λ, i.e. known break date k, the asymptotic distribution in (11) is again χ^2_K like the standard Chow-type test statistics.

Because the asymptotic distribution of the QLR statistic is a functional of Brownian motions, critical values can be simulated on the computer. A realisation of the asymptotic distribution is generated through simulating a realisation of a standard Brownian motion, calculating the functional in (11) over a range of prespecified break dates and take the supremum. Repeating this exercise many times one gets an approximation of the asymptotic distribution and critical values are set through taking appropriate quantiles. For example, the tables in Andrews (1993) show critical values for various λ_0, λ_1 and K. These critical values are considerably higher than corresponding χ^2 critical values. Therefore, the χ^2 distribution is a poor approximation to the asymptotic distribution of the QLR statistic, i.e. using χ^2 critical values leads to overrejection of the null hypothesis even when T is large.

Other functions than the maximum of a series test statistics, i.e. the average instead of the supremum, could be exploited. Also the analysis could in principle be extended to more than one break, but maybe the TVP model in (3) could be preferable in this case. In the TVP model the null hypothesis of no break can be formulated as $H_0 : \tau^2 = 0$. Using FCLT and CMT limiting distributions for these tests can be derived, which will be again functionals of standard Brownian motions. Therefore, like in the case of the QLR statistic critical values can be calculated by Monte Carlo simulation.

Because the asymptotic distribution of the QLR statistic is known, asymptotically it has no size distortions. As far as asymptotic power is concerned, several local alternatives can be considered, i.e.

$$y_t = x_t'\beta + \frac{1}{\sqrt{T}} x_t'\delta I_{t>k} + \varepsilon_t, \qquad t = 1, ..., T$$

where $I_{t>k}$ is the indicator function, which is one for $t > k$ and zero else, or

$$\begin{aligned} y_t &= x_t'\beta_t + \varepsilon_t \\ \beta_t &= \beta_{t-1} + \frac{1}{T} H \eta_t \\ \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix} &\sim IID \begin{bmatrix} I_K & 0 \\ 0 & \sigma^2 \end{bmatrix} \end{aligned}$$

The local alternative in both the change point and the random walk model is indexed with the sample size T, because otherwise for large T every test would have power one. It is possible to derive the limiting distribution of the QLR test statistic under these local alternatives, which consists of a null part and a local alternative piece. With these analytical expressions asymptotic power comparisons between test statistics can be made.

As far as finite sample properties of the QLR and other test statisitcs are concerned, these have to be established mainly by Monte Carlo experiments. To derive analytical results is extremely difficult. Several studies are concerned with finite sample size and power comparisons. In general, using asymptotical critical values most tests tend to overreject the null hypothesis, i.e. actual size exceeds nominal size in finite samples.

2.3 Estimation of break date

Another related subject to structural breaks is the estimation of break dates. Considering the change point model

$$y_t = x_t'\beta + x_t'\delta I_{t>k_0} + \varepsilon_t, \qquad t = 1, ..., T \tag{12}$$

a natural estimator for the true break date k_0 is

$$\hat{k} = \arg \min_{\lambda_0 < \frac{k}{T} < \lambda_1} SSR(k)$$

where $SSR(k)$ is the sum of squared OLS residuals of the regression (12) for a specific value of k. For fixed nonzero δ the estimator $\hat{\lambda} = \frac{\hat{k}}{T}$ is a consistent estimator of $\lambda_0 = \lim_{T\to\infty} \frac{k_0}{T}$, the fraction of the sample where the break occurs. However, $\hat{k}$ is not a consistent estimator for the break date k_0 itself. Only if the break magnitude δ is indexed with the sample size T an asymptotic distribution is available for $\hat{k}$, which is a functional of a two-sided Brownian motion. With this asymptotic distribution in hand, confidence intervals can be made for the location of the break.

3 Dynamic factor models

This type of model can be used for a variety of purposes. One of the goals can be to detect common movements in several macroeconomic time series. The extracted common factors represent then the state of the economy, i.e. they summarize the state of the economy in only a few trend series. A dynamic factor model can be written as

$$y_t = \Gamma(L) f_t + u_t, \qquad t = 1, ..., T, \tag{13}$$

where the $n \times 1$ vector of observables $y_t = (y_{1t}, ..., y_{nt})'$ is decomposed in two elements, i.e. k unobserved common factors $f_t = (f_{1t}, ..., f_{kt})'$ and an unobserved element specific idiosyncratic shock $u_t = (u_{1t}, ..., u_{nt})'$. The factors f_t have an $n \times k$ matrix lag polynomial $\Gamma(L)$ as loadings. Furthermore, both u_t anf f_t follow an AR process, i.e.

$$\begin{aligned} d_i(L) u_{it} &= \varepsilon_{it}, \qquad i = 1, ..., n, \\ \beta_j(L) f_{jt} &= \eta_{jt}, \qquad j = 1, ..., k \end{aligned} \tag{14}$$

where ε_{it} and η_{jt} are serial uncorrelated and uncorrelated with each other and have finite variance $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{\eta_j}$ respectively. Also we assume stationarity of the common factors. This is not a limiting assumption as such, because residuals from cointegrating relations can be elements of f_t.

The primary interest of the analysis is to make forecasts about the vector of common factors f_t. To predict the common factors it is necessary to get estimates of the unknown parameters in $\Gamma(L)$, $d_i(L)$ and $\beta_j(L)$ first. When n is relatively small the unknown parameters in the dynamic factor model can be estimated with maximum likelihood and forecasts for the state vector f_t can be made using the Kalman filter. First the model in (13) and (14) is rewritten in state space form. The state equation describes the development in the unobserved elements f_t and u_t, while the measurement or observable equation links the observed variables y_t to f_t and u_t. In general notation, the state space form is

$$\begin{aligned} \xi_{t+1} &= F\xi_t + v_t \\ y_t &= H'\xi_t + w_t \end{aligned} \tag{15}$$

where v_t and w_t are serial uncorrelated with

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}\right]$$

The elements of ξ_t are the elements of f_t and u_t plus lagged values depending on the order of the lag polynomials $\Gamma(L)$, $d_i(L)$ and $\beta_j(L)$. The matrices F, H, Q and R consist of the unknown parameters in $\Gamma(L)$, $d_i(L)$ and $\beta_j(L)$ and $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{\eta_j}$. The likelihood function for model (15) can be construcuted and the unknown parameters in F, H, Q and R can be estimated with maximum likelihood. Having estimates for the unknown parameters using the Kalman filter one can forecast values for the state vector f_t.

For larger n things are more difficult, because the asymptotic theory available is valid for large T and fixed n. For both T and n large, brute application of the techniques above is not to be preferred. Most alternatives proposed in the literature include some asymptotic theory for T and n growing at a certain rate, but they differ significantly in the treatment of the problem and the estimation techniques used.

References

Andrews, D.W.K. (1993), Tests for parameter instability and structural change with unknown change point, *Econometrica* **61**, 821-856.

Chow, G.C. (1960), Tests of equality between sets of coefficients in two linear regressions, *Econometrica* **28**, 591-605.

Diebold, F.X. and C. Chen (1996), Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures, *Journal of Econometrics* **70**, 221-241.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, New Yersey.

Quandt, R.E. (1960), Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of the American Statistical Association* **55**, 324-330.

Stock, J.H. (1994), Unit Roots, Structural Breaks and Trends, chapter 46 in *Handbook of Econometrics*, volume IV, Elsevier Science, Amsterdam.

Stock, J.H. and M.W. Watson (1991), A Probability Model of the Coincident Economic Indicators, chapter 4 in K. Lahiri and G.H. Moore (eds), *Leading economic indicators: New approaches and forecasting records*, Cambridge University Press, New York.