Estimating Structural Changes in Linear Simultaneous Equations

HUANG, Weihong and ZHANG, Yang
Division of Economics
School of Humanities and Social Sciences
Nanyang Technological University, Singapore

Abstract

Tests and estimation for structural changes in single equation models, particularly the analysis of covariance and the Chow tests, are well known to econometricians and widely used. This paper demonstrates that analogous estimation can also be constructed in simultaneous equation models when equations are estimated by common estimator like 2SLS and 3SLS. In the present paper, we discuss the problem of estimating structural changes, both equation by equation and globally, in simultaneous equations model. We consider the case of possible multiple switching of the parameters at unknown sample points and investigate the simultaneous estimation of multiple structural changing points along with the regression coefficients within subdomains. A recursive segmentation method will be used which is based on the principle of dynamic programming and allow global minimizers to be obtained using a number of sums of squared residuals.

Key words: Structural changes, simultaneous equations, recursive segmentation

1. Introduction

Before an econometric model can be used to draw inference about economic phenomena, it is of great importance to assess the adequacy of its specification. To test structural change, or parameter constancy as the simplest case, is particularly appropriate for dynamic econometric models in that policy prescriptions might be quite different with the presence of structural changes.

Simultaneous equation model has been widely used in econometric literatures; nevertheless there are only a few results available on analyzing and testing the stability of the coefficients in structural equations. Andrews and Fair (1988) discusses the problem in a general setting, while more concrete situations are studied in Lo and Newey (1985) and Erlat (1983). The former work extended Chow's (1960) tests to simultaneous equations and proposed a simple Wald test, composed of two-stage lease-squares (2SLS) estimator and the estimate of its covariance matrix. Erlat (1983) advocated an exact F test for the cases when there are inadequate degrees of freedom. It also constituted an extension of Gile's (1981) result, where CUSUM and CUSUM of squares tests for parameter stability in a single structural equation are developed. However, it is a common drawback of all the tests mentioned above that the switching point is assumed to be known a priori, which is often not the case in applied research. Their studies, in addition, only examine the structural change of one single equation belonging to a simultaneous equation model rather than considering the entire system as a whole.

The single-equation counterpart in this agenda has been discussed by several authors (Chong 1995; Bai 1997, etc). In many applications, the number of structural breaks as well as the locations of break points are taken as unknown parameters and a theory of least squares estimation has been developed. Operationally, the whole sample is split at each possible breakpoint, the other parameters are estimated by OLS and the sum of squared errors (SSE) is calculated. The least squares breakpoint estimation is the value that minimizes the full sample SSE. Independently in an early study, Huang, Liu and

Zhang (1985) considered multiple structural changes in a linear model estimated by least-squares and proposed an information criterion for the selection of the number of changes. Their study discussed the problem in a more general framework, where the form of the model is not fixed between various segments and other estimation criteria besides OLS are allowed. However, there has not been a corresponding development of tests and estimation method for changes in the coefficients of structural equations in simultaneous-equation models. The purpose of this paper is to investigate this issue in simultaneous equations context.

The outline of the paper is as follows. Section 2 presents the Recursive Segmentation (RS) method which is able to detect the unknown changing points in simultaneous equations. Section 3 tackles the problem of determining number of the changes where information criterion is applied. Concluding remarks are provided in section 4.

2 Recursive Segmentation

The basis of the method, for specialized cases, is documented by Fisher (1958) and Guthery (1974). However, thorough treatment and description of the main idea in the econometrics literature seems still sparse. We hope to fill a small part of that breach.

The standard linear simultaneous equations model is considered first, where all identities are assumed to have been substituted out of the system of equations:

$YB + X\Gamma = U$

where Y is the $N \times M$ matrix of jointly dependent variables, X is the $N \times K$ matrix of predetermined variables, and U is a $N \times M$ matrix of the structural disturbances of the system. Thus the model consists of M equations and N observations. We have assume B is nonsingular, $rk(X) = K$, and that all equations satisfy the rank condition for identification. Lastly, the orthogonality condition applies between the predetermined variables and structural errors, and the second order moment matrices of the current predetermined and endogenous variables are assumed to have nonsingular probability

limits. The structural errors are assumed to be independent and identically distributed. Then structural change is said to be present within the range of the index i if

$$Y(i)B_1 + Z(i)\Gamma_1 = U_1(i) \qquad i \in I_1$$

$$Y(i)B_2 + Z(i)\Gamma_2 = U_2(i) \qquad i \in I_2$$

.

.

$$Y(i)B_l + Z(i)\Gamma_l = U_l(i) \qquad i \in I_l \tag{2.1}$$

Here $I_1 \cup I_2 \cup ... \cup I_l = I$ (I is the domain of i in the set of samples.) Usually, we have $I_s \cap I_t = \varnothing \quad (s \neq t)$ The index variable i, in time series data, corresponds to time or observation.

Now with the index or partitioning variable identified, the inferential problem confronting us involves three parts: (1) the specification of the number of changes in the model, l; (2) the detection of the change point $\{i_s\}$, or the boundaries of intervals over which each of the model pieces applies; (3) the estimation of the model parameters within each subdomain. If l and the $\{i_s\}$ were specified, step 3 would simply consist of applying the classical theory, interval by interval. Summing the residual sums of squares for the various intervals yields an overall index of the quality of fit of the segmented model. With l fixed, the $\{i_s\}$ may be estimated by minimizing this index. Further minimization of the index to estimate l will base on information criterion for model selection problems.

In estimating the appropriate sample separation in simultaneous equation system, there are two approaches to analyze the timing and form of structural changes, either to estimate equation by equation separately using a limited information estimator, or globally consider joint estimation of the entire system. In this section, we first deal with the estimation of structural changes in one equation embedded in a system of

simultaneous equations. We have parallel, more complex results for system methods of detecting structural changes. The corresponding discussion follows.

2.1 Single Equation: Limited Information Method

We shall, without loss of any generality, consider the first equation in the system of simultaneous equations and write it as

$$y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1 \tag{2.1.1}$$

where y_1 and u_1 are $(N\times 1)$, $u \sim N(0,\sigma^2 I)$, Y_1 is $(N\times g_1)$, X_1 is $(N\times m_1)$ and the elements of β_1, γ_1 identified.

The reduced form corresponding to (2.1.1) is

$$Y_I = Z\Pi_I + V_I \tag{2.1.2}$$

where $Y_I = (y_1, Y_1)$, $Z = (X_1, X_2)$, $\Pi_I = (\pi_1, \Pi_1)$ and $V_I = (v_1, V_1)$. Z is the $(N\times G)$ matrix of non-stochastic exogenous variables in the complete system, Π_1 is a $G\times(m_1+1)$ matrix of reduced form coefficients, and V_I is a $G\times(m_1+1)$ matrix of reduced form disturbances whose rows are assumed to be normally and independently distributed with zero mean and covariance Ω_I.

Now comparing (2.1.1) and (2.1.2), we have $u_1 = v_1 - V_1\beta_1 = V_I\beta_1^0$ where $\beta_1^0 = (1, -\beta_1')'$. Thus, one may estimate u_1 by utilizing appropriate estimators for V_I and β_1^0. Then V_I may be estimated by applying OLS to (2.1.2) to yield $\hat{V}_I = Y_I - Z\hat{\Pi}_I$, since V_I is reduced form coefficient and OLS will give consistent estimation. Meanwhile, β_1^0 will be estimated from (2.1.1) using 2SLS. Thus, the appropriate estimator of u_1 would then be $u_1^* = \hat{V}_I \cdot \hat{\beta}_1^0$.

Since we know that $\hat{V}_I = M_Z Y_I$, where $M_Z = I - Z(Z'Z)^{-1}Z'$, it follows that u_1^* may be obtained directly as the residual vector of the unrestricted OLS regress of $y_1 - Y_1\hat{\beta}_1^1$ on Z,

that is, regressing $y^* = Y_1 \hat{\beta}_1^0$ on Z. Denoting the $G \times 1$ coefficient vector of said regression by δ and u_1^* would be expressed alternatively as

$$u_1^* = y^* - Z\hat{\delta} \tag{2.1.3}$$

As shown in Harvey and Phillips (1980), conditional on $\hat{\beta}_1^0$, u_1^* has the same distributional property as the OLS residuals from the general linear regression model with well-behaved disturbances.

Now we proceed to the structural change analysis. Suppose we have structural change model in the form of (2.1) and have divided the entire sample set into l segments, and let $1 = i_1 < i_2 < ... < i_l < N$ be the subscripts of the first data points of each segment. We denote such segmentation with N observation of l segments as $P(N, l) = \{i_1, ..., i_l\}$. We define target function, e, as the statistics that describe the overall goodness-of-fit of the model using certain estimation criteria. The value of the target function within a segment is called the diameter, denoted as d. Obviously, e is a function of d. To illustrate the segmentation procedure, the most commonly used criterion in regression, ordinary least square (OLS), is used in this section. This gives specific form to the target function and diameters and simplifies the discussion. Other methods, like maximum likelihood and minimax criteria may also be applied.

Following on the above definition and using OLS, the diameter of the segment from point i_s to $i_{s+1} - 1$, out of N observations, is defines as:

$$\begin{aligned} d\ (i_s, i_{s+1} - 1) &= \min_{\delta} \sum_{t=i_s}^{i_{s+1}-1} (y_t^* - Z_t \delta_t)^2 \\ &== \sum_{t=i_s}^{i_{s+1}-1} (y_t^* - Z_t \hat{\delta}_t)^2 \qquad\qquad s = 1, 2, ..., l \end{aligned} \tag{2.1.4}$$

where y_t^* is the tth element of y^* matrix. And e is defined as:

$$e[p(N, l)] = \min_{\delta^{(s)}} \{ \sum_{s=1}^{l} \sum_{t=i_s}^{i_{s+1}-1} (y_t^* - Z_t \delta_t)^2 \} \qquad s = 1, 2, ..., l \tag{2.1.5}$$

that when $\delta^{(s)}$ ($s=1, 2, \ldots, l$) are all different, from (2.1.5) we have:

$$e\,[p(N,l)] = \sum_{s=1}^{l} \sum_{t=i_s}^{i_{s+1}-1} (y_t^* - Z_t\hat{\delta}_t)^2 = \sum_{s=1}^{l} d\ (i_s, i_{s+1}-1) \tag{2.1.6}$$

Equation (2.1.6) shows that by the construction of the method, the target function $e[\,p(N,l)]$ can be decomposed into the sum of individual diameters. This is to say, given l, the overall optima with respect to δ can be achieved by optimizing each segment (since the segments are independent from each other). The ultimate goal is to obtain the optimal segmentation: $\tilde{p}(N,l) = \{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_l\}$, which minimizes the target function, i.e.:

$$e\,[\tilde{p}(N,l)] = \min_{p \in p(N,l)} e\,[p(N,l)] \tag{2.1.7}$$

let $\hat{\delta}$ denote the resulting estimates based on the given l partition $(i_1, i_2, \ldots, i_l)$. Substituting these estimates in the objective function and denoting the resulting sum of squared residuals as $SSE(i_1, i_2, \ldots, i_l)$, the estimated break points $\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_l\}$ can be alternatively denoted as

$$\{\tilde{i}_1, \tilde{i}_2, \ldots, \tilde{i}_l\} = \arg\min_{i_1, i_2, \ldots, i_l} SSE(i_1, i_2, \ldots, i_l)$$

Because the target function is the additive function of diameters, it is obvious that $e[\,p(N,l)]$ satisfies the separability condition in a multi-stage decision-making problem in dynamic programming. Thus, by using the technique of backward recursive optimization, the following relationship can be derived:

$$\begin{aligned} e[\tilde{p}(n,s)] &= \min_{s \le j_s < n} \{e[\tilde{p}(j_s - 1, s-1)] + d(j_s, n)\} \\ &= e[\tilde{p}(\tilde{j}_s - 1, s-1)] + d\ (\tilde{j}_s, n) \qquad (1 < s < n \le N) \end{aligned} \tag{2.1.8}$$

In (2.1.8), $\tilde{p}(j_s - 1, s-1)$ represents the optimal *s-1* segmentation of first $j_s - 1$ observations, while $\tilde{j}_s$ is the corresponding value of j_s that minimizes the value of (2.1.8), denoted as *g(n ,s)*. Especially when *s=2*, (2.1.8) becomes:

$$\begin{aligned} e[\tilde{p}(n,2)] &= \min_{2 \le j_2 < n} \{d(1, j_2 - 1) + d(j_2, n)\} \\ &= d(1, \tilde{j}_2 - 1) + d(\tilde{j}_2, n) \qquad (2 < n \le N) \end{aligned} \tag{2.1.9}$$

Equations (2.1.8) and (2.1.9) are the main devices to obtain optimal segmentation. Using (2.1.8) recursively, we can obtain all $e[\tilde{p}(n,l')], l' = 3,4,...,N-1, (l' < n < N)$ and the corresponding $g(n,l')$; For any given *l, (l=1,2,...,[N/m])* (*N* is the sample size while *m* is the number of independent variables), check out $\tilde{i}_l = g(N,l)$, which is the subscript of the first data point of the *l*th segment. Then, from

$$\tilde{i}_s = g(\tilde{i}_{s+1} - 1, s), (s = 1,2,...,l) \tag{2.1.10}$$

we obtain all the subscripts of the first data points of the rest segments, $\tilde{i}_{l-1}, \tilde{i}_{l-2}, ..., \tilde{i}_2$. Thus the optimal segmentation is derived as:

$$\tilde{p}(N,l) = \{\tilde{i}_1, \tilde{i}_2, ..., \tilde{i}_l\}. \tag{2.1.11}$$

The critical step of the RS method is the recursive equation (2.1.9). That is why it is called "*recursive segmentation method*".

2.2 Full Information Method of Estimation

Structural change analysis using limited information method allows each equation among the whole system to react differently to external shocks, or the structural breaks. That means each equation might change its structural at different timing. However, intuition would surely suggest that full information, or systems methods of estimation are asymptotically better than limited information methods which estimate the system one equation at a time, since the latter neglect information contained in other equations while the former brings efficiency gains. Now we examine the structural instability using the technique of joint estimation of the entire system of equations.

The stated discussion about structural changes detection and, in particular, the RS method is also applicable to this circumstance. We may formulate the full system as

$$Y = Z\delta + U$$

where $E(U) = 0$ and $E[UU'] = \Sigma \otimes I$. In line with the principle of system methods, the technique of three-stage least square is used for joint estimation of the entire system of equations. Thus the 3SLS estimator is

$$\hat{\delta}_{3SLS} = [\hat{Z}'(\Sigma^{-1} \otimes I)\hat{Z}]^{-1}\hat{Z}'(\Sigma^{-1} \otimes I)Y \tag{2.2.1}$$

where $\hat{Z}$ is the IV estimator for 2SLS.

Again, the model is assumed to have $l-1$ structural changes in the whole sample period, i.e., l subsamples. Following the definition of diameter and target function stated in section 2.1, we have

$$d\ (i_s, i_{s+1}-1) = \sum_{h=1}^{M} d_h(i_s, i_{s+1}-1) \tag{2.2.2}$$

where $d_h(i_s, i_{s+1}-1)$ is the diameter of the h th equation, for the individual segment staring from i_s to $i_{s+1}-1$. $d\ (i_s, i_{s+1}-1)$ is the summation of all the diameters throughout the system. Given the structural changes in the form of $P(N,l) = \{i_1, i_2, ..., i_l\}$, we have

$$e\ [p(N,l)] = \sum_{s=1}^{l} d\ (i_s, i_{s+1}-1) = \sum_{s=1}^{l}\sum_{h=1}^{M} d_h(i_s, i_{s+1}-1) \tag{2.2.3}$$

Those corresponding diameters can be calculated from 3SLS estimators. Similarly, we have the optimum of target function as $e\ [\tilde{p}(N,l)] = \min_{p \in p(N,l)} e\ [p(N,l)]$. Again, the estimated break points will be $\{\tilde{i}_1, \tilde{i}_2, ..., \tilde{i}_l\} = \arg\min_{i_1, i_2, ..., i_l} SSE(i_1, i_2, ..., i_l)$.

It is obvious that the technique of backward recursive optimization and dynamic programming procedure are applicable and we can apply RS method again to detect the structural changes without grid search calculation. The use of full information or system method in model estimation makes use of the cross-equation correlations of the disturbances. In so doing, the structural change analysis is conducted with regard to the whole simultaneous equation system instead of with each equation at one time. The structural changes occurred are therefore assumed to affect the whole system, with each equation included, at the same timing.

3 Determining the Number of Breaks

By using recursive classification, we can obtain different recursive segmentations simultaneously, given different number of segments l. If and only if $l=l^0$ (l^0 is the true number of structure changes), the optimal l segmentation can optimally fit in the given data and reflect the structure changing characteristics of the model. In practice we may not have such information of the exact number of changes or the number of segments. Another standard problem is that an improvement in the objective function is always possible by allowing more breaks. Therefore, in determining optimal l^0, we have to take both the goodness-of-fit and the efficiency of the model into consideration, so that maximum amount of accurate information can be obtained. Information criterion which derives from maximizing the posterior likelihood in a model selection paradigm and enjoys widespread use in model identification provides a natural baseline.

In the multiple change point model, it has been found that sequential estimation is consistent to estimate the model without treating all break points simultaneously. The basic logic is to estimate the break point using the whole sample data and then to divide the sample into two sub-samples at the estimated break point which allows the greatest reduction in the sum of squared residuals. Estimate an additional break whenever the sub-sample fails the parameter constancy test. This step is repeated until all the sub-domains do not reject the null hypothesis of no structural changes. Although it yields consistent estimates of the break points, the estimates are not guaranteed to be identical to those obtained by global minimization.

In this section, we use model selection method, and in particular, information criterion to determine the number of breaks. Information theory and, in particular the Kullback-Leibler (Kullback and Leibler 1951) “distance” or “information” forms the deep theoretical basis for data-based model selection. Akaike (1973) found a simple relationship between expected Kullback-Leibler information and Fisher's maximized log-likelihood function. This relationship leads to a simple, effective, and very general methodology for selecting a parsimonious model for the analysis of empirical data.

We know that the general form of Information Criterion (IC) is:

$$IC_s = -2\ln[L(M_s)] + P(m_s) \tag{3.1}$$

where M_s represents the sth model, m_s is the corresponding number of independent variables. It is expected intuitively that a more complicated model will provide a better approximation to reality. But, on the contrary, in most practical situations a less complicated model is likely to be preferred if we wish to pursue the accuracy of estimation. $L(M_s)$ is the value of the maximum likelihood function of the model, which takes into account of the goodness-of-fit of the model. $P(m_s)$ is the penalty function, which is an increasing function with respect to m_s and penalizes the index when the model's efficiency decreases. Thus, the RS model should be the one with smallest IC value. By using computer simulation, the investigation of the penalty function with different values of N, m and the variance σ^2 suggests that the AIC function by Akaike, BIC of Schwarz and CAI of Sugiura are all appropriate.

Based on the results obtained in section 2.2, for different number l, we have found the optimal segmentations and the corresponding estimation of the whole sample. Now the determination of l^0 will be obtained according to the IC criteria, i.e., the one which allows the greatest reduction in the IC value: $l^0 = \arg\min IC(l^s, i_1, i_2, ..., i_{l^s})$.

4. Conclusion

In this paper we present a comprehensive treatment of issues related to the estimation of linear models with multiple structural changes, to detection of the presence of multiple structural changes and to the determination of the number of changes present. It will further allow us to state facts about the number of segments present in the horizon covered, the magnitude of the mean and variance in each subsample, the nature of the dynamics in the error component, and the timing of the changes in regime.

In particular, the RS method we present is able to correctly detect and estimate the existence and the timing of unknown changing points. This method provides a systematic and operational approach that can accurately detect structural changing points without any prior information or knowledge of the pattern and timing of possible structural shifts. The method is based on the principle of dynamic programming and allows global minimizers to be obtained using a number of sums of squared residuals rather than an exhaustive grid search.

The single equation estimation of structural changes enables us to detect the structural instability in individual equation, which is not necessary that of the entire simultaneous system. Since the changing points estimated among different equations may differ to some extent, this provides a new angle to explain the spillover effect of some policy implementation. Meanwhile, allowance is made to adapt different estimating criterion for different equation and the flexibility, in this respect, is gained. Another obvious practical consideration of estimating equation-wide instead of system-wide is the computational simplicity of the single equation methods. But the current state of available software has all but eliminates this advantage.

Although the systems methods are asymptotically better, they are risky at propagating any specification error in one structural model throughout the system. But obtaining a unique structural changing points estimator help in consistency interpretation and enjoy the simplicity in coping with one system with the same number of segments, rather than with various numbers and different timing of structural changes for individual equation.

Reference

- Akaike, H., 1973, “Information Theory and an Extension of the Maximum Likelihood Principle,” *Second International Symposium on Information Theory*.
- Andrews, D. W. K., and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615-640.
- Bai, Jushan, 1997, “Estimating Multiple Breaks One at a Time.” *Econometric Theory*, June 13.3, pp. 315-52;
- Chong, Terence Tai-leung, 1995, “Partial Parameter Consistency in a Misspecified Structural Change Model,” *Economics Letters*, October, 49:4, pp351-57
- Chow, G. C., 1960, “Tests of equality between sets of coefficients in two linear regressions,” *Econometrica*, 28, pp. 591-605.
- Erlat, 1983, “A Note on Testing Structural Change in a Single Equation Belonging to a Simultaneous System”, *Economics Letters* **13** (1983), pp. 185–189
- Fisher. W. D. 1958, “On grouping for maximum homogeneity”. *Journal of the American Statistical Association.*, 53:789--798
- Giles, 1981, “Testing for parameter stability in structural econometric relationships”, *Economics Letters* **7** (1981), pp. 323–326
- Guthery Scott B., 1974, “Partition Regression (in Theory and Methods)”, *Journal of the American Statistical Association*, Vol. 69, No. 348. pp. 945-947.
- Harvey and Phillips, 1980, “Testing for Serial Correlation in Simultaneous Equation Models”, *Econometrica*, Vol. 48, No. 3. pp. 747-760.
- Huang Weihong, Liu Bao and Zhang Shiying, 1985, “An Optimal Classification Model-building Method for Structure-changed Models”, *Journal of Systems Engineering*, Vol. 1, pp. 11-23.
- Kullback, S. and Leibler, R. A., 1951, “On Information and Sufficiency.” Annals of Mathematical Statistics 22:76-86.
- Lo and Newey, 1985, “A Large Sample Chow Test for the Linear Simultaneous Equation”, *Economics Letters* **18** (1985), pp. 351–353
- Schwarz, G., 1978, “Estimating the dimension of a model,” *Annals of Statistics*, 6, 461-464.
- Sugiura, N., 1978, “Further analysis of the data by Akaike's information criterion and the finite corrections,” *Communications in Statistics*, Theory and Methods, A7:13-26.