HITOTSUBASHI UNIVERSITY

Hi-Stat Discussion Paper

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Model Selection Criteria in Multivariate Models with Multiple Structural Changes

Eiji Kurozumi
Purevdorj Tuvaandorj

June 2010

Model Selection Criteria in Multivariate Models with Multiple Structural Changes[1]

Eiji Kurozumi[2]   Purevdorj Tuvaandorj

Department of Economics   Department of Economics

Hitotsubashi University   McGill University

May, 2010

Abstract

This paper considers the issue of selecting the number of regressors and the number of structural breaks in multivariate regression models in the possible presence of multiple structural changes. We develop a modified Akaike's information criterion (AIC), a modified Mallows' C_p criterion and a modified Bayesian information criterion (BIC). The penalty terms in these criteria are shown to be different from the usual terms. We prove that the modified BIC consistently selects the regressors and the number of breaks whereas the modified AIC and the modified C_p criterion tend to overly choose them with positive probability. The finite sample performance of these criteria is investigated through Monte Carlo simulations and it turns out that our modification is successful in comparison to the classical model selection criteria and the sequential testing procedure with the robust method.

JEL classification: C13; C32

Keywords: structural breaks, AIC; Mallows' Cp; BIC; information criteria

1. Introduction

This paper considers the selection of regressors and estimation of the number of structural changes in multivariate regression models in the possible presence of multiple structural changes. Many methods for the selection of regressors have been proposed in the econometric and statistical literature, and it is often the case in practical analyses that the regressors are selected using either testing procedures or model selection criteria. The former methods select the regressors by testing the significance of the coefficients of the regressors and deleting the insignificant coefficients from the models, while the model selection criteria choose the regressors that minimize the given risk functions. The representative model selection criteria in econometric analysis are the Akaike information criterion (AIC) by Akaike (1973), the C_p criterion by Mallows (1973) and the Bayesian information criterion (BIC) by Schwarz (1978) among others. See Burnham and Anderson (2002) and Konishi and Kitagawa (2008) for a general treatment of the model selection criteria.

In addition to the selection of the regressors, we need to consider the possibility of structural changes when we investigate data covering a relatively long sample period. In such a case we usually test for structural changes. Various tests for structural changes have been proposed in the literature, and the most commonly used tests in recent practical analyses are the sup-type test of Andrews (1993) and the exponential and average-type tests of Andrews, Lee and Ploberger (1996) among others. These tests assume the null hypothesis of no changes against the alternative of (multiple) change(s), whereas Bai and Perron (1998) and Bai (1999) proposed tests for the null of ℓ breaks against the alternative of $\ell + 1$ breaks for univariate models. These tests are extended to multivariate models by Qu and Perron (2007), who give a comprehensive treatment on the issue of the estimation, inference, and computation in a system of equations with multiple structural changes. Their treatment is general enough in that less restrictive assumptions are placed on the error term and that models such as vector autoregressions (VAR), seemingly unrelated regressions and panel data models are included in their setup as special cases. See Perron (2006) for a review of the testing and estimation of structural changes.

Once the evidence of structural breaks is found, the next step is to estimate the number

of breaks. Bai (1997b, 1999), Bai and Perron (1998) and Qu and Perron (2007) proposed to implement tests for structural changes sequentially and proved that the estimated number of structural changes is consistent by letting the significance level go to zero. Alternatively, in the statistical literature, the model selection criteria have been proposed to select the number of breaks. For independent normal random variables with mean shifts, Yao (1988) and Zhang and Siegmund (2007) derived the modified BIC and Ninomiya (2005) proposed to modify the AIC, while Liu, Wu and Zidek (1997) considered the modified BIC in regression models with i.i.d. regressors. According to these works, the penalty terms of these new criteria are different from those of the corresponding classical criteria because of the irregularity in the change points. Although these results are of interest from a statistical point of view, they cannot be directly applied to economic data because while economic time series variables are typically serially correlated, in the above papers, assumptions such as i.i.d. observations and regressors are made. Exceptions are Ninomiya (2006) and Hansen (2009). The former considered the modified AIC in finite order autoregressive models, but derived under the assumption of known variance. Hansen (2009) established the modified Mallows' C_p criterion but with only a single break being allowed.

In this paper we develop the model selection criteria in multivariate models allowing lagged dependent variables as regressors in the possible presence of multiple structural changes in both the coefficients and the variance matrices. Our criteria have an advantage over the existing ones in that (i) multivariate models are considered, (ii) serial correlation is taken into account in models by allowing serially correlated regressors, including lagged dependent variables, (iii) structural changes in the variance matrices are allowed. We theatrically derive the AIC, C_p criterion and BIC in models with structural changes and show that the penalty terms should be modified compared with those of the corresponding classical ones. We confirm by Monte Carlo simulations that this modification of the penalty terms is very important to correctly select the regressors and the number of structural changes in finite samples.

The rest of this paper is organized as follows. We explain the model and assumptions in Section 2. Section 3 establishes the modified AIC, the modified C_p criterion and the modified BIC with multiple structural breaks and discusses the consistency of these criteria. In Section 4, we investigate the finite sample performance of our model selection criteria via simulations.

Concluding remarks are provided in Section 5.

2. Model and Assumptions

Let us consider the following n-dimensional regression model with m structural changes ($m+1$ regimes):

$$y_t = \Phi_j x_{jt} + \varepsilon_t \quad (j = 1, \cdots, m+1 \quad \text{and} \quad t = T_{j-1} + 1, \cdots, T_j) \tag{1}$$

where y_t and x_{jt} are $n \times 1$ and $p_{x_j} \times 1$ vectors of observations, respectively, ε_t is an error term and Φ_j is an $n \times p_{x_j}$ unknown coefficient matrix in the jth regime. Typically, the regressor x_{jt} includes a constant but trending regressors are not allowed in our model. We use the term $p_{\phi_j} = np_{x_j}$ to denote the number of unknown coefficients in each regime, so that the total number of coefficients is given by $p_\phi^{all} = \sum_{j=1}^{m+1} p_{\phi_j} = \sum_{j=1}^{m+1} np_{x_j}$. Similarly, by allowing structural changes in the variance matrix of ε_t, the number of unknown variance components in each regime is $p_\sigma = n(n+1)/2$ and that in all regimes is given by $p_\sigma^{all} = (m+1)p_\sigma = (m+1)n(n+1)/2$. We set $T_0 = 0$ and $T_{m+1} = T$, so that the total number of observations is T. In model (1) there are m structural changes ($m+1$ regimes) with change points given by $T_1, \cdots, T_m$. We allow the lagged dependent variables as regressors and in that case, the initial observations of y_t for $t \leq 0$ are assumed to be given. Thus, model (1) includes a VAR model as a special case. Note that the different regressors and the different orders of the lagged dependent variables are allowed depending on the regimes. The main purpose of this paper is to derive the model selection criteria to choose the regressors among the $\bar{p}_x$ candidates for regressors and to estimate the number of structural changes m. In what follows, while we will continue using "choose the number p_x among the $\bar{p}_x$ regressors," we, however, imply "choose the regressors $x_{1t}, x_{2t}, \cdots, x_{m+1t}$ for all the regimes among the $\bar{p}_x$ candidates for regressors."

Model (1) can be rewritten as $y_t = (x'_{jt} \otimes I_n)\phi_j + \varepsilon_t$ for the jth regime where $\phi_j = \text{vec}(\Phi_j)$ is a $p_{\phi_j} \times 1$ vector. We denote the true value of a parameter with superscript 0. For example, ϕ_j^0 and T_j^0 denote the true value of ϕ_j in the jth regime and the true jth break point, respectively. Hence, the data generating process is given by

$$y_t = (x'_{jt} \otimes I_n)\phi_j^0 + \varepsilon_t.$$

3

The following assumptions are supposed mainly for the derivation of the modified AIC and the modified C_p criterion.

Assumption A1 *(a) There exists a positive integer $l_0 > 0$ such that for all $l > l_0$, the minimum eigenvalues of $(1/l)\sum_{t=T^0_{j-1}+1}^{T^0_{j-1}+l} x_{jt}x'_{jt}$ and $(1/l)\sum_{t=T^0_j-l}^{T^0_j} x_{jt}x'_{jt}$ are bounded away from zero $(j = 1, \cdots, m^0 + 1)$. (b) $\sum_{t=k}^{l} x_{jt}x'_{jt}$ is invertible for $l - k > k_0$ for some $0 < k_0 < \infty$. (c) $\sup_{j,t} E\|x_{jt}\|^{4+\delta} < \infty$ for some $\delta > 0$.*

Assumption A2 *When the lagged dependent variables are allowed as regressors, all the characteristic roots associated with the lag polynomials are inside the unit circle.*

Assumption A3 *(a) $\varepsilon_t = (\Sigma^0_j)^{1/2}\eta_t$ for $T^0_{j-1} + 1 \leq t \leq T^0_j$ $(j = 1, \cdots, m^0 + 1)$, where Σ^0_j is a symmetric and positive definite unknown matrix and $\{\eta_t\}$ is a martingale difference sequence with respect to $\mathcal{F}_t = \sigma\{\eta_t, \eta_{t-1}, \cdots, z_{t+1}, z_t, \cdots, \}$ with $E[\eta_t\eta'_t|\mathcal{F}_{t-1}] = I_n$ for all t. (b) $\sup_t E\|\eta_t\|^{4+\delta} < \infty$ for some $\delta > 0$. (c) $E[\eta_{it}\eta_{jt}\eta_{kt}] = 0$ $(i,j,k = 1, \cdots, n)$. (d) $(1/\Delta T^0_j)\mathrm{tr}\left\{\left[\sum_{t=T^0_{j-1}+1}^{T^0_j}(\eta_t\eta'_t - I_n)\right]^2\right\} \xrightarrow{p} \kappa_{4j}$, where κ_4 is some positive number and $\Delta T^0_j = T^0_j - T^0_{j-1}$ $(j = 1, \cdots, m^0 + 1)$.*

Assumption A4 *$\phi^0_{j+1} - \phi^0_j = v_T\delta_j$ and $\Sigma^0_{j+1} - \Sigma^0_j = v_T\Psi_j$, where $(\delta_j, \Psi_j) \neq 0$ $(j = 1, \cdots, m^0)$, $\Sigma^0_j \to \Sigma^0$ as $T \to \infty$ for all j and v_T is a sequence of positive numbers such that $v_T \to 0$ and $\sqrt{T}v_T/(\log T)^2 \to \infty$.*

Assumption A5 *$0 = \lambda_0 < \lambda^0_1 < ... < \lambda^0_{m^0} < \lambda_{m^0+1} = 1$, where $T^0_j = [T\lambda^0_j]$ $(j = 0, \cdots, m^0 + 1)$.*

Assumption A6 *The following weak law of large numbers and the functional central limit theorems hold $(j = 1, \cdots, m^0)$:*

$$\frac{1}{\Delta T^0_j}\sum_{t=T^0_{j-1}+1}^{T^0_j}\left(x_{jt}x'_{jt}\otimes(\Sigma^0_j)^{-1}\right) \xrightarrow{p} Q_{1j}, \quad \frac{1}{\Delta T^0_{j+1}}\sum_{t=T^0_j+1}^{T^0_{j+1}}\left(x_{j+1t}x'_{j+1t}\otimes(\Sigma^0_{j+1})^{-1}\right) \xrightarrow{p} Q_{2j},$$

$$v_T\sum_{t=T^0_j-[vv_T^{-2}]}^{T^0_j}(\eta_t\eta'_t - I_n) \Rightarrow \xi_{1j}(v), \quad v_T\sum_{t=T^0_j+1}^{T^0_j+[vv_T^{-2}]}(\eta_t\eta'_t - I_n) \Rightarrow \xi_{2j}(v),$$

4

$$v_T \sum_{t=T_j^0-[vv_T^{-2}]}^{T_j^0} \left(x_{jt} \otimes (\Sigma_j^0)^{-1/2} \right) \eta_t \Rightarrow Q_{1j}^{1/2} \zeta_{1j}(v), \quad v_T \sum_{t=T_j^0+1}^{T_j^0+[vv_T^{-2}]} \left(x_{j+1t} \otimes (\Sigma_{j+1}^0)^{-1/2} \right) \eta_t \Rightarrow Q_{2j}^{1/2} \zeta_{2j}(v)$$

for $v \geq 0$, where Q_{1j} and Q_{2j} are positive definite matrices, $\xrightarrow{p}$ and $\Rightarrow$ signify convergence in probability and weak convergence of the associated probability measures, respectively, each entry of $\xi_{1j}(v)$ and $\xi_{2j}(v)$ is a (nonstandard) Brownian motion process on $[0,\infty)$ with $Var(\text{vec}(\xi_{1j}(1)) = \Omega_{1j}$ and $Var(\text{vec}(\xi_{2j}(1)) = \Omega_{2j}$, while each element of $\zeta_{1j}(v)$ and $\zeta_{2j}(v)$ is a standard Brownian motion process on $[0,\infty)$, and $\xi_{1j}(v)$, $\xi_{2j}(v)$, $\zeta_{1j}(v)$ and $\zeta_{2j}(v)$ are independent of each other.

The above assumptions satisfy or are similar to the conditions provided by the existing literature. For detailed explanations, see Bai (1997a) and Bai and Perron (1998) for the univariate case and Bai (2000) and Qu and Perron (2007) for the multivariate case. Note that we do not allow serial correlation in the error term to derive the model selection criteria. This is more restrictive as compared to the assumptions in Qu and Perron (2007). However, since the lagged dependent variables are allowed as regressors and some elements of x_{jt} can be the lagged values of the other elements, Assumption A3 may not be too restrictive for practical purposes. We should also note that the regressor x_{jt} is not necessarily homogeneous in all the regimes. In other words, the regime-wise heteroskedastic regressors are allowed in our model. It is known that the assumption of heteroskedasticity in x_{jt} and the shrinking shifts in Assumption A4 result in the asymmetric limiting distributions of the break point estimators. This asymmetry will make the modified AIC and the modified C_p criterion take relatively complicated forms.

We estimate model (1) by the quasi-maximum likelihood (QML) method, conditional on the initial values if the lagged dependent variables are allowed as regressors. Let $\phi = [\phi_1', \phi_2', \cdots, \phi_{m+1}']'$, $\sigma = [\text{vec}(\Sigma_1)', \text{vec}(\Sigma_2)', \cdots, \text{vec}(\Sigma_{m+1})']'$, $\theta = [\phi', \sigma']'$ and $\mathcal{T} = [T_1, T_2, \cdots, T_m]'$. Then, given the number of breaks and the regressors, the log-likelihood function, denoted by $\ell_{m,p_x}(\mathcal{T}, \theta | y, x)$, becomes

$$\ell_{m,p_x}(\mathcal{T}, \theta | y, x) = -\frac{nT}{2} \log 2\pi - \sum_{j=1}^{m+1} \frac{\Delta T_j}{2} \log |\Sigma_j|$$

5

$$-\frac{1}{2}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}+1}^{T_j}\{y_t-(x_{jt}'\otimes I_n)\phi_j\}'\Sigma_j^{-1}\{y_t-(x_{jt}'\otimes I_n)\phi_j\}, \tag{2}$$

where $\Delta T_j = T_j - T_{j-1}$ and the subscripts m and p_x signify that the log-likelihood function depends on the number of structural changes and the selected regressors, respectively. The maximum likelihood estimators (MLE) of θ and $\mathcal{T}$ for given m and p_x are obtained by maximizing (2) over $\{(\lambda_1,\cdots,\lambda_m); |\lambda_{j+1}-\lambda_j| \geq \epsilon$ for $j=1,\cdots,m\}$ for small $\epsilon > 0$ and are denoted by $\hat{\theta}$ and $\hat{\mathcal{T}}$, respectively.

Under Assumptions A1-A6, Qu and Perron (2007) showed that the MLEs of ϕ_j and Σ_j have the standard asymptotic distributions as in the case of known break points, whereas the limiting distributions of the break date estimators are given by, for $j=1,\cdots,m^0$,

$$v_T^2(\hat{T}_j - T_j^0) \xrightarrow{d} \operatorname*{argmax}_v B_j^I(v) \tag{3}$$

where the superscript I denotes the dependency on the indicator function given by $I(v \leq 0)$,

$$B_j^I(v) = \begin{cases} B_{1j}(v) = \sqrt{\omega_{1j}}W_{1j}(|v|) - \frac{|v|}{2}\gamma_{1j} & : & v \leq 0 \\ B_{2j}(v) = \sqrt{\omega_{2j}}W_{2j}(v) - \frac{v}{2}\gamma_{2j} & : & v > 0, \end{cases} \tag{4}$$

$$\omega_{1j} = \frac{1}{4}\operatorname{vec}(A_{1j})'\Omega_{1j}\operatorname{vec}(A_{1j})+\delta_j'Q_{1j}\delta_j, \quad \gamma_{1j} = \frac{1}{2}\operatorname{tr}(A_{1j}^2)+\delta_j'Q_{1j}\delta_j, \quad A_{1j} = (\Sigma_j^0)^{-1/2}\Psi_j(\Sigma_j^0)^{-1/2},$$

$$\omega_{2j} = \frac{1}{4}\operatorname{vec}(A_{2j})'\Omega_{2j}\operatorname{vec}(A_{2j})+\delta_j'Q_{2j}\delta_j, \quad \gamma_{2j} = \frac{1}{2}\operatorname{tr}(A_{2j}^2)+\delta_j'Q_{2j}\delta_j, \quad A_{2j} = (\Sigma_{j+1}^0)^{-1/2}\Psi_j(\Sigma_{j+1}^0)^{-1/2},$$

and $W_{1j}(v)$ and $W_{2j}(v)$ are independent standard Brownian motions on $[0,\infty)$.

Before moving to the derivation of the model selection criteria, we give the following lemma.

Lemma 1 *Let $B_j^I(v)$ be defined as in (4). Then,*

$$E\left[\max_v B_j^I(v)\right] = \frac{r_{1j}^2 + r_{1j}r_{2j} + r_{2j}^2}{r_{1j} + r_{2j}}, \tag{5}$$

$$E\left[a_j^I\left|\operatorname*{argmax}_v B_j^I(v)\right|\right] = 2\left\{\left(\frac{a_2}{\gamma_{2j}}\right)\frac{r_{2j}^2(2r_{1j}+r_{2j})}{(r_{1j}+r_{2j})^2} + \left(\frac{a_1}{\gamma_{1j}}\right)\frac{r_{1j}^2(r_{1j}+2r_{2j})}{(r_{1j}+r_{2j})^2}\right\}, \tag{6}$$

$$E\left[a_j^I\operatorname*{argmax}_v B_j(v)\right] = 2\left\{\left(\frac{a_2}{\gamma_{2j}}\right)\frac{r_{2j}^2(2r_{1j}+r_{2j})}{(r_{1j}+r_{2j})^2} - \left(\frac{a_1}{\gamma_{1j}}\right)\frac{r_{1j}^2(r_{1j}+2r_{2j})}{(r_{1j}+r_{2j})^2}\right\}, \tag{7}$$

where $a_j^I = a_1$ if $v \leq 0$ and $a_j^I = a_2$ if $v > 0$, $r_{1j} = \omega_{1j}/\gamma_{1j}$ and $r_{2j} = \omega_{2j}/\gamma_{2j}$.

6

Although the above three expectations are generally expressed using complicated forms, they can be written in a simpler manner in special cases. For example, when $a_1 = \gamma_{1j}$ and $a_2 = \gamma_{2j}$, (6) and (7) become

$$E\left[a_j^I \left|\operatorname*{argmax}_v B_j^I(v)\right|\right] = 2\frac{r_{1j}^2 + r_{1j}r_{2j} + r_{2j}^2}{r_{1j} + r_{2j}},$$

$$E\left[a_j^I \operatorname*{argmax}_v B_j^I(v)\right] = 2\frac{(r_{2j} - r_{1j})(r_{1j}^2 + 3r_{1j}r_{2j} + r_{2j}^2)}{(r_{1j} + r_{2j})^2}.$$

In addition, if x_{jt} is homoskedastic across the regimes, $B_j^I(v)$ has a symmetric distribution with $\omega_{1j} = \omega_{2j} = \omega_j$ and $\gamma_{1j} = \gamma_{2j} = \gamma_j$ so that the expectations reduce to

$$E\left[\max_v B_j^I(v)\right] = \frac{3}{2}r_j, \quad E\left[a_j^I \left|\operatorname*{argmax}_v B_j^I(v)\right|\right] = 3r_j, \quad E\left[a_j^I \operatorname*{argmax}_v B_j^I(v)\right] = 0,$$

where $r_j = r_{1j} = r_{2j}$. Moreover, if $r_{1j} = r_{2j} = 1$ as in the case of the Gaussian error, then we have

$$E\left[\max_v B_j^I(v)\right] = \frac{3}{2}, \quad E\left[a_j^I \left|\operatorname*{argmax}_v B_j^I(v)\right|\right] = 3, \quad E\left[a_j^I \operatorname*{argmax}_v B_j^I(v)\right] = 0,$$

which are the same as obtained in Ninomiya (2005). Lemma 1 will be used to derive the modified AIC and the modified C_p criterion in the next section.

3. Derivation of Model Selection Criteria

In this section we derive the three model selection criteria, AIC, C_p criterion and BIC, taking structural changes into account. More precisely, let $\bar{p}_x$ and $\bar{m}$ be the largest number of regressors and the largest number of structural changes, respectively, which we have to prespecify. We propose to choose p_x and m among the $\bar{p}_x$ candidates for regressors and $0 \leq m \leq \bar{m}$, respectively, based on the derived model selection criteria, where though we conventionally state "choose p_x," we imply that we select an optimal set of regressors $x_{1t}, x_{2t}, \cdots, x_{m+1t}$ among the $\bar{p}_x$ candidates. Note that all of the following three criteria are designed to choose the model that minimizes them.

3.1. Akaike information criterion

Akaike information criterion (AIC) is defined as the unbiased estimator of (-2) times the expected log-likelihood given by $E_y[E_{y^*}[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y^*, x^*)]]$, which is in turn equivalent to 2 times the Kullback-Leibler (KL) information, where y^* and x^* have the same distribution as y and x but are independent of y and x, and E_y and E_{y^*} are expectation operators with respect to y and y^*, respectively. See, for example, Burnham and Anderson (2002) and Konishi and Kitagawa (2008). Note that $\hat{\mathcal{T}}$ and $\hat{\theta}$ are based not on y^* but on y. Since we choose the model that minimizes the AIC, we can interpret the chosen model as optimal in the sense that it minimizes the KL information.

Akaike (1973) proposed to estimate the expected log-likelihood by the empirical log-likelihood but he also showed that the empirical log-likelihood is the biased estimator of the expected log-likelihood in finite samples. As in Akaike (1973) we consider the following criterion that depends on the number of structural changes m and the selected regressors, which we conventionally denote as p_x:

$$AIC(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + 2b_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}), \tag{8}$$

where $b_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}) = E_y[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) - E_{y^*}[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y^*, x^*)]]$ corresponds to the bias. Since the first term on the right hand side of (8) is the maximized log-likelihood obtained by the QML estimation, we only need to evaluate the bias term explicitly.

In order to calculate the bias, we decompose $b_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta})$ into four parts as follows:

$$
\begin{aligned}
b_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}) &= E_y\left[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) - E_{y^*}\left[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y^*, x^*)\right]\right] \\
&= E_y\left[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) - \ell_{m,p_x}(\mathcal{T}^0, \theta^0|y, x)\right] \\
&\quad + E_y\left[\ell_{m,p_x}(\mathcal{T}^0, \theta^0|y, x) - E_{y^*}\left[\ell_{m,p_x}(\mathcal{T}^0, \theta^0|y^*, x^*)\right]\right] \\
&\quad + E_y\left[E_{y^*}\left[\ell_{m,p_x}(\mathcal{T}^0, \theta^0|y^*, x^*)\right] - E_{y^*}\left[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta}|y^*, x^*)\right]\right] \\
&\quad + E_y\left[E_{y^*}\left[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta}|y^*, x^*)\right] - E_{y^*}\left[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y^*, x^*)\right]\right] \\
\\
&= b_{m,p_x,1} + b_{m,p_x,2} + b_{m,p_x,3} + b_{m,p_x,4}, \quad \text{say.} \tag{9}
\end{aligned}
$$

As in the literature we evaluate $b_{m,p_x,1}$ to $b_{m,p_x,4}$ up to the $O(1)$ terms for the true values of m and p_x and obtain the modified AIC.

Proposition 1 *Under Assumptions A1-A6 with $m = m^0$ and $p_x = p_x^0$, the bias terms $b_{m,p_x,1}$, $b_{m,p_x,2}$, $b_{m,p_x,3}$ and $b_{m,p_x,4}$ are, up to the $O(1)$ terms, given by*

$$b_{m,p_x,1} = \frac{p_\phi^{all}}{2} + \sum_{j=1}^{m+1} \frac{\kappa_{4j}}{4} + \sum_{j=1}^{m} \frac{r_{1j}^2 + r_{1j}r_{2j} + r_{2j}^2}{r_{1j} + r_{2j}},$$

$$b_{m,p_x,2} = 0, \qquad b_{m,p_x,3} = \sum_{j=1}^{m+1} \frac{\kappa_{4j}}{4} + \frac{p_\phi^{all}}{2}, \quad b_{m,p_x,4} = \sum_{j=1}^{m} \frac{r_{1j}^2 + r_{1j}r_{2j} + r_{2j}^2}{r_{1j} + r_{2j}}.$$

Proposition 1 suggests that the modified AIC should be defined as

$$MAIC(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat\theta | y, x) + 2p_\phi^{all} + \sum_{j=1}^{m+1} \hat\kappa_{4j} + 4\sum_{j=1}^{m} \frac{\hat{r}_{1j}^2 + \hat{r}_{1j}\hat{r}_{2j} + \hat{r}_{2j}^2}{\hat{r}_{1j} + \hat{r}_{2j}}, \qquad (10)$$

where $\hat{\ }$ denotes the consistent estimator of the corresponding parameter. The estimators of the parameters are given by $\hat\kappa_{4j} = \mathrm{tr}(\hat\Omega_j)$,

$$\hat{r}_{1j} = \frac{\frac{1}{4}\mathrm{vec}(\hat{A}_{1j})'\hat\Omega_j \mathrm{vec}(\hat{A}_{1j}) + \hat\delta_j'\hat{Q}_{1j}\hat\delta_j}{\frac{1}{2}\mathrm{tr}(\hat{A}_{1j}^2) + \hat\delta_j'\hat{Q}_{1j}\hat\delta_j}, \quad \text{and} \quad \hat{r}_{2j} = \frac{\frac{1}{4}\mathrm{vec}(\hat{A}_{2j})'\hat\Omega_{j+1} \mathrm{vec}(\hat{A}_{2j}) + \hat\delta_j'\hat{Q}_{2j}\hat\delta_j}{\frac{1}{2}\mathrm{tr}(\hat{A}_{2j}^2) + \hat\delta_j'\hat{Q}_{2j}\hat\delta_j},$$

$$\text{where} \quad \hat\Omega_j = \frac{1}{\Delta\hat{T}_j} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \mathrm{vec}(\hat\eta_t\hat\eta_t' - I_n)\mathrm{vec}(\hat\eta_t\hat\eta_t' - I_n)', \quad \hat\delta_j = \hat\phi_{j+1} - \hat\phi_j,$$

$$\hat{A}_{1j} = \hat\Sigma_j^{-1/2}\left(\hat\Sigma_{j+1} - \hat\Sigma_j\right)\hat\Sigma_j^{-1/2}, \quad \hat{A}_{2j} = \hat\Sigma_{j+1}^{-1/2}\left(\hat\Sigma_{j+1} - \hat\Sigma_j\right)\hat\Sigma_{j+1}^{-1/2}, \quad \hat\Sigma_j = \frac{1}{\Delta\hat{T}_j}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat\varepsilon_t\hat\varepsilon_t'$$

$$\hat{Q}_{1j} = \frac{1}{\Delta\hat{T}_j}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} (x_{jt}x_{jt}' \otimes \hat\Sigma_j^{-1}) \quad \text{and} \quad \hat{Q}_{2j} = \frac{1}{\Delta\hat{T}_{j+1}}\sum_{t=\hat{T}_j+1}^{\hat{T}_{j+1}} (x_{jt}x_{jt}' \otimes \hat\Sigma_{j+1}^{-1})$$

with $\hat\varepsilon_t = y_t - (x_{jt}' \otimes I_n)\hat\phi_j$ for $\hat{T}_{j-1} + 1 \le t \le \hat{T}_j$ ($j = 1, \cdots, m$).

Although (10) is expressed in a complicated form, the modified AIC can be simplified in several interesting cases. For example, when there are no weakly exogenous regressors and model (1) is a pure VAR model, the limiting distributions of the break point estimators are symmetric with $\omega_{1j} = \omega_{2j} = \omega_j$ and $\gamma_{1j} = \gamma_{2j} = \gamma_j$ as shown by Bai (2000). In this case, $r_{1j} = r_{2j} = r_j$ so that

$$MAIC(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat\theta | y, x) + 2p_\phi^{all} + \sum_{j=1}^{m+1} \hat\kappa_{4j} + 6\sum_{j=1}^{m} \hat{r}_j.$$

9

When η_t is Gaussian, it can be shown that $\kappa_{4j} = n(n + 1) = 2p_\sigma$ for all j. In addition, we have $\omega_{1j} = \gamma_{1j}$ and $\omega_{2j} = \gamma_{2j}$ because $\text{vec}(A_{1j})'\Omega_{1j}\text{vec}(A_{1j})/2 = \text{tr}(A_{1j}^2)$ and $\text{vec}(A_{2j})'\Omega_{2j}\text{vec}(A_{2j})/2 = \text{tr}(A_{2j}^2)$ as shown in Remark 5 of Qu and Perron (2007). As a result, $r_{1j} = r_{2j} = 1$ so that the modified AIC reduces to

$$MAIC(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + 2(p_\phi^{all} + p_\sigma^{all}) + 6m. \tag{11}$$

Even if η_t is not Gaussian but if there are no changes in the variance matrices, the modified AIC takes a simple form. In this case, $B_j^I(s)$ becomes as given by (4) with $\omega_{1j} = \gamma_{1j} = \delta_j' Q_{1j}\delta_{1j}$ and $\omega_{2j} = \gamma_{2j} = \delta_j' Q_{2j}\delta_j$. Again, we have $r_{1j} = r_{2j} = 1$, so that the modified AIC reduces to

$$MAIC(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + 2p_\phi^{all} + 6m \tag{12}$$

where we omit κ_4 because this term is common for all the models in this case.

In any case we can see that the penalty term of the modified AIC is different from the classical one, which is 2 times the number of unknown coefficients given by $2p_\phi^{all}$. In order to see the intuitive meaning of the penalty term of the modified AIC, we first consider the case where m is prefixed and one additional regressor is included in each regime. In this case, the third and fourth terms on the right hand side of (10) are basically the same except for the changes in the estimators of κ_{4j}, r_{1j} and r_{2j} while the second term is increased by $2(m+1)n$ because the difference between the number of unknown coefficients in the two cases is $2\sum_{j=1}^{m+1} n(p_{x_j} + 1) - 2\sum_{j=1}^{m+1} np_{x_j} = 2(m+1)n$. Since $(m+1)n$ is the number of additional unknown coefficients, the penalty on the additional regressors is the same as in the classical case.

Next, let us consider the case where the number of structural changes is increased from $m - 1$ to m. For ease of exposition, consider the case where the additional break is found in the last regime with $p_{x_{m+1}}$ regressors. From (10) the penalty term is increased by

$$2\left(p_{x_{m+1}} + \frac{\hat{\kappa}_{4,m+1}}{2} + 2\frac{\hat{r}_{1m}^2 + \hat{r}_{1m}\hat{r}_{2m} + \hat{r}_{2m}^2}{\hat{r}_{1m} + \hat{r}_{2m}}\right). \tag{13}$$

The first term in the parentheses of (13) is interpreted as the penalty on the additional regressors, while the second term corresponds to the penalty on the increased number of

unknown variance components. That is, when the number of structural changes increases, we need to estimate the variance matrix in the additional new regime given by Σ_{m+1} and $\hat{\kappa}_{4,m+1}/2$ can be interpreted as the penalty on Σ_{m+1}. In fact, when η_t is Gaussian, $\kappa_{4,m+1}/2$ becomes equal to $n(n+1)/2$, which is the same as the number of unknown components, p_σ, in Σ_{m+1}.

On the other hand, the third term in the parentheses of (13) can be interpreted as the penalty on searching for an additional break because this term *does not* appear when the mth break point, T_m^0, is known but *does* appear when we estimate it, as is understood from the proof of Proposition 1. When we estimate the additional unknown break, we look for the break point that maximizes the log-likelihood; further, the maximizing point is not necessarily the true break date. In general, the maximization is possibly attained at a point different from the true break date in finite samples. In other words, the uncertainty of the break point always leads to a larger log-likelihood (or a smaller model selection criterion) and the third term of (13) can be interpreted as the penalty on this uncertainty. In fact, we can see from (3) and Lemma 1 that the third term in the parentheses of (13) is an approximation of $E[\gamma_j^I v_T^2 |\hat{T}_m - T_m^0|]$, which can be seen as a measure of the uncertainty of the mth break point, where $\gamma_j^I = \gamma_{1j}$ for $v \leq 0$ and $\gamma_j^I = \gamma_{2j}$ for $v > 0$. Thus, the more uncertain or more volatile the break point estimator, the heavier the penalty imposed on the modified criterion. As is seen in (11) and (12), this penalty becomes equal to 6 ($= 2 \times 3$) in special cases such as when the error is Gaussian and when there are no breaks in the variance matrix, because $E[\gamma_j^I v_T^2 |\hat{T}_m - T_m^0|] = 3$ in these cases. We also note that the third term of (13) is an increasing function of $r_{1j} = \omega_{1j}/\gamma_{1j}$ and $r_{2j} = \omega_{2j}/\gamma_{2j}$, both of which become larger when the break point estimator is more volatile (for larger ω_{1j} and ω_{2j}). Since this penalty on the uncertainty is positive, the modified AIC will choose a smaller number of structural changes than the classical AIC. This property will be confirmed by the simulations in a later section.

3.2. Mallows' C_p criterion

Mallows (1973) focused on the prediction of the conditional mean of a univariate model and proposed as a measure of adequacy for prediction the scaled sum of squared forecast errors. In this subsection we extend the Mallows' C_p criterion to multivariate models with multiple

structural changes by introducing a multivariate version of the scaled sum of square forecast errors. The model minimizing the modified C_p criterion is optimal from the viewpoint of the minimization of the risk function based on the forecast errors.

Let $\mu_t^0 = (x'_{jt} \otimes I_n)\phi_j^0$ be the conditional mean of y_t for $T_{j-1}^0 + 1 \leq t \leq T_j^0$ and $\hat{\mu}_t = (x'_{jt} \otimes I_n)\hat{\phi}_j$ be its estimator for $\hat{T}_{j-1} + 1 \leq t \leq \hat{T}_j$ ($j = 1, \cdots, m+1$). As suggested by Mallows (1973) we adopt as a measure of adequacy of prediction the trace of the scaled residual variance matrix given by

$$J_{m,p_x} = \sum_{j=1}^{m+1} \text{tr}\left[\Sigma_j^{0-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \left(\hat{\mu}_t - \mu_t^0\right)\left(\hat{\mu}_t - \mu_t^0\right)'\right].$$

Let $\hat{\varepsilon}_t = y_t - \hat{\mu}_t$ for $\hat{T}_{j-1} + 1 \leq t \leq \hat{T}_j$ as defined before and $\tilde{\varepsilon}_t = y_t - \tilde{\mu}_t$ with $\tilde{\mu}_t = (x'_{jt} \otimes I_n)\hat{\phi}_j$ for $T_{j-1}^0 + 1 \leq t \leq T_j^0$. That is, $\hat{\varepsilon}_t$ is the residual from the ML estimation while $\tilde{\varepsilon}_t$ is the residual when we forecast the conditional mean by $(x_{jt} \otimes I_n)\hat{\phi}_j$ in the true regimes. Since $y_t = \mu_t^0 + \varepsilon_t = \hat{\mu}_t + \hat{\varepsilon}_t = \tilde{\mu}_t + \tilde{\varepsilon}_t$, J_{m,p_x} becomes

$$
\begin{aligned}
J_{m,p_x} &= \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} (\varepsilon_t - \hat{\varepsilon}_t)'\Sigma_j^{0-1}(\varepsilon_t - \hat{\varepsilon}_t) \\
&= \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \varepsilon_t'\Sigma_j^{0-1}\varepsilon_t - 2\sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} (y_t - \hat{\mu}_t)'\Sigma_j^{0-1}\varepsilon_t + \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \hat{\varepsilon}_t'\Sigma_j^{0-1}\hat{\varepsilon}_t \\
&= -\sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \varepsilon_t'\Sigma_j^{0-1}\varepsilon_t + 2\sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} (\hat{\mu}_t - \mu_t^0)'\Sigma_j^{0-1}\varepsilon_t \\
&\quad -\sum_{j=1}^{m+1}\left(\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t'\Sigma_j^{0-1}\hat{\varepsilon}_t - \sum_{t=T_{j-1}^0+1}^{T_j^0} \hat{\varepsilon}_t'\Sigma_j^{0-1}\hat{\varepsilon}_t\right) + \sum_{j=1}^{m+1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t'\Sigma_j^{0-1}\hat{\varepsilon}_t \\
&= J_{m,p_x,1} + J_{m,p_x,2} + J_{m,p_x,3} + \sum_{j=1}^{m+1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t'\Sigma_j^{0-1}\hat{\varepsilon}_t, \quad \text{say.} \quad (14)
\end{aligned}
$$

Following Mallows' (1973) original work, the modified C_p criterion is defined as

$$MC_p(m, p_x) = \sum_{j=1}^{m+1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t'\hat{\Sigma}_{t,\bar{m},\bar{p}_x}^{-1}\hat{\varepsilon}_t + E[J_{m,p_x,1} + J_{m,p_x,2} + J_{m,p_x,3}]$$

12

where $\hat{\Sigma}_{t,\bar{m},\bar{p}_x}$ is the estimator of the variance matrix based on the most general model using $m = \bar{m}$ and $p_x = \bar{p}_x$. For a univariate case ($n = 1$) with no structural changes, Mallows (1973) showed that $E[J_{m,p_x,1} + J_{m,p_x,2} + J_{m,p_x,3}] = 2p_x - T$ and hence the modified C_p criterion takes a well known form by ignoring $-T$. For our model (1) with structural changes, it is easy to see that $E[J_{m,p_x,1}] = -nT$, which is common for all the models and can be ignored. On the other hand, it can be shown that the dominant term in $J_{m,p_x,3}$ is of order v_T^{-1} while $J_{m,p_x,2}$ is $O_p(1)$. As can be seen in the proof of Proposition 2 we have

$$
\begin{aligned}
v_T J_{m,p_x,3} &= -\sum_{j=1}^{m} 1(\hat{T}_j < T_j^0)\mathrm{tr}(A_{1j})v_T^2(\hat{T}_j - T_j^0) + 1(\hat{T}_j > T_j^0)\mathrm{tr}(A_{2j})v_T^2(\hat{T}_j - T_j^0) + o_p(1) \\
&\xrightarrow{d} -\sum_{j=1}^{m} \mathrm{tr}(A_j^I)\operatorname*{argmax}_{v} B_j^I(v) = \bar{J}_{m,p_x,3}, \quad \text{say,}
\end{aligned}
\tag{15}
$$

where $A_j^I = A_{1j}$ when $v \leq 0$ and $A_j^I = A_{2j}$ when $v > 0$. Then, the natural candidate for the modified C_p criterion would be $MC_p(m, p_x) = \sum_{j=1}^{m+1}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t'\hat{\Sigma}_{t,\bar{m},\bar{p}_x}^{-1}\hat{\varepsilon}_t + E[\bar{J}_{m,p_x,3}]$. However, this criterion may not be informative for the choice of models. For example, when x_{jt} is homoskedastic in all the regimes, $\operatorname{argmax}_v B_j^I(v)$ has a symmetric distribution, which implies $E[\bar{J}_{m,p_x,3}] = 0$. In this case, the modified criterion consists of only $\sum_{j=1}^{m+1}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t'\hat{\Sigma}_{t,\bar{m},\bar{p}_x}^{-1}\hat{\varepsilon}_t$, which is always minimized when we choose the most general model.

In order to avoid the above problem we need to evaluate the second dominant term in $J_{m,p_x,3}$, which would be of the same order as is $J_{m,p_x,2}$. The problem here is that the second dominant term in $J_{m,p_x,3}$ will be obtained by the higher order expansion of $v_T^2(\hat{T}_j - T_j^0)$, which is tedious to derive in practice. We might construct a new criterion by ignoring the whole term of $J_{m,p_x,3}$ but such a criterion may not be optimal from the viewpoint of a measure of adequacy for prediction.

Because of the above reason, we do not construct the modified C_p criterion under Assumptions A1-A6. Instead, we consider the same criterion under more restrictive assumptions; we impose restrictions on the breaks in the variance matrices such that $\Sigma_{j+1}^0 - \Sigma_j^0 = v_T^2\Psi_j$. In other words, we allow only those breaks that are smaller than the ones supposed in Assumption A4. Note that by changing the assumption, the limiting distributions of the break point estimators are not affected by the breaks in the variance matrices but depend only on the

breaks in the coefficients, whereas a measure of prediction given by J_{m,p_x} still depends on the breaks in the variance matrices through $J_{m,p_x,3}$.

Proposition 2 *Under Assumptions A1-A6 with $\Sigma_{j+1}^0 - \Sigma_j^0 = v_T^2 \Psi_j$ and with $m = m^0$ and $p_x = p_x^0$, the expectations of the first three terms of J_{m,p_x} are given by, up to the $O(1)$ terms,*

$$E[J_{m,p_x,1}] = -nT, \quad E[J_{m,p_x,2}] = 2p_\phi^{all} + 6m, \quad E[J_{m,p_x,3}] = \frac{3}{2}\sum_{j=1}^{m}\left(\frac{\text{tr}(A_{1j})}{\gamma_{1j}} - \frac{\text{tr}(A_{2j})}{\gamma_{2j}}\right),$$

where $\gamma_{1j} = \delta_j' Q_{1j}\delta_j$ and $\gamma_{2j} = \delta_j' Q_{2j}\delta_j$ in this case.

Proposition 2 suggests that the modified C_p criterion should be given by, ignoring the $O(1)$ terms,

$$MC_p(m,p_x) = \sum_{j=1}^{m+1}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t' \hat{\Sigma}_{t,\bar{m},\bar{p}_x}^{-1}\hat{\varepsilon}_t + 2p_\phi^{all} + 6m + \frac{3}{2}\sum_{j=1}^{m}\left(\frac{\text{tr}(\hat{A}_{1j})}{\hat{\gamma}_{1j}} - \frac{\text{tr}(\hat{A}_{2j})}{\hat{\gamma}_{2j}}\right). \quad (16)$$

Note that the last penalty term might be negative depending on the asymmetric property of the limiting distributions of the break point estimators. In the symmetric case the last penalty term disappears so that the modified C_p criterion reduces to

$$MC_p(m,p_x) = \sum_{j=1}^{m+1}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t' \hat{\Sigma}_{t,\bar{m},\bar{p}_x}^{-1}\hat{\varepsilon}_t + 2p_\phi^{all} + 6m. \quad (17)$$

For example, the two sided Brownian motions $B_j^I(v)$ $(j = 1, \cdots, m^0)$ become symmetric when there are no structural changes in the variance matrices, when (1) is a pure VAR model and when x_{jt} is homogeneous across the regimes.

From (16) we can see that the penalty term of the modified C_p criterion is different from the classical criterion. As in the case of the modified AIC, the second term on the right hand side of (16) is the penalty on the additional regressors while the third term can be interpreted as the penalty on the uncertainty associated with estimating the break points. The last term of (16) is related with the breaks in the variance matrices and the asymmetry of the break point estimators. From the proof of Proposition 2 we can see that the last penalty term is an approximation of $-\sum_{j=1}^{m}\text{tr}(A_j)E[v_T^2(\hat{T}_j - T_j^0)]$ where we used the fact that $A_{1j} = A_{2j} = (\Sigma^0)^{-1/2}\Psi(\Sigma^0)^{-1/2} = A_j$, say, asymptotically. This implies

14

that the additional positive penalty is imposed when $\Psi_j > 0$ and $E[v_T^2(\hat{T}_j - T_j^0)] < 0$ (or $\Psi_j < 0$ and $E[v_T^2(\hat{T}_j - T_j^0)] > 0$) and vice versa. To interpret this property, let us assume that Σ_j^0 for $j = 1, \cdots, m+1$ are known. In this case, the contribution to J_{m,p_x} from the jth regime should be given by $\text{tr}[\Sigma_j^{0-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} (\hat{\mu}_t - \mu_t^0)(\hat{\mu}_t - \mu_t^0)']$ but we need to replace T_j^0 with $\hat{T}_j$ for $j = 1, \cdots, m$ in practice. That is, the actual contribution is given by $\text{tr}[\Sigma_j^{0-1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} (\hat{\mu}_t - \mu_t^0)(\hat{\mu}_t - \mu_t^0)']$. In this case, if $\Psi_j > 0$, or equivalently, Σ_j^{0-1} is larger than Σ_{j+1}^{0-1} in terms of the matrix, and if $\hat{T}_j$ tends to be smaller than T_j^0, then the contribution to J_{m,p_x} from the jth regime decreases more than expected. In order to adjust the smaller contribution from the jth regime, we need an additional positive penalty. On the other hand, if, again, Σ_j^{0-1} is relatively large but if the distribution of $\hat{T}_j$ is skewed to the right, the contribution to J_{m,p_x} from the jth regime is larger than expected and hence we need to reduce this contribution by a negative penalty. Thus, the last term of (16) can be interpreted as the penalty on the overweight or underweight on the forecast errors caused by the asymmetry of the break point estimators.

We should keep in mind that the modified C_p criterion given by (16) is optimal from the viewpoint of minimizing the risk function based on the prediction errors only in the case where the magnitude of structural changes in the variance matrices is negligibly small as compared to the magnitude of shifts in the coefficients.

3.3. Bayesian information criterion

Schwarz (1978) considered the problem of model selection in the Bayesian framework, in which a model is selected based on the posterior probability. In this subsection, we derive the modified BIC for models with structural changes using the Laplace approximation technique as explained in Konishi and Kitagawa (2008). The model selected by the modified BIC is optimal from the viewpoint of the maximization of the posterior probability.

Let $M(m, p_x, \mathcal{T})$ be a model for given m, p_x and $\mathcal{T}$, $P(M(m, p_x, \mathcal{T}))$ be the prior probability of a given model, $f_M(y|\theta, x)$ be the probability density function (pdf) of y conditional on θ and x for a given model $M = M(m, p_x, \mathcal{T})$ and $\pi_M(\theta)$ be the prior pdf for θ. Then, the

posterior probability of model $M(m, p_x, \mathcal{T})$ is given by

$$P\left(M(m, p_x, \mathcal{T})|y, x\right) = \frac{g_M(y|x)P\left(M(m, p_x, \mathcal{T})\right)}{\sum g_M(y|x)P\left(M(m, p_x, \mathcal{T})\right)}, \tag{18}$$

where the summation is taken by over models $M(m, p_x, \mathcal{T})$ and $g_M(y|x)$ is the marginal distribution of y conditional on x defined as $g_M(y|x) = \int f_M\left(y|\theta, x\right)\pi_M(\theta)d\theta$. We adopt the model that maximizes the posterior probability (18), but the maximization of (18) is equivalent to the maximization of the numerator on the right hand side of (18) because the denominator is common for all the models. Thus, we consider the maximization of $g_M(y|x)P(M(m, p_x, \mathcal{T}))$, or equivalently, the minimization of

$$-2\log\left\{g_M(y|x)P\left(M(m, p_x, \mathcal{T})\right)\right\} = -2\log g_M(y|x) - 2\log P\left(M(m, p_x, \mathcal{T})\right). \tag{19}$$

In order to evaluate (19), we make the following assumptions.

Assumption A7 *(a) The conditional pdf of y, $f_M(y|\theta, x)$, is Gaussian. (b) The priors for $\phi_1, \cdots, \phi_{m+1}$ and $\Sigma_1^{-1}, \cdots, \Sigma_{m+1}^{-1}$ are noninformative. (c) The prior probability of $M(m, p_x, \mathcal{T})$ is given by $P(M(m, p_x, \mathcal{T})) = \alpha_{m,T}/T^m$ where $0 < \underline{\alpha} < \alpha_{m,T} < \bar{\alpha} < \infty$.*

Note that we do not have to assume shrinking shifts in the Bayesian framework. Instead, we make assumptions on the priors. Assumption A7 (a) states that we base our analysis on a Gaussian distribution. We can interpret Assumption A7 (b) such that we do not have a priori information on the distributional property of the parameters. Note that the noninformative priors for the reciprocal of the variance matrices are sometimes considered in the Bayesian framework. Assumption A7 (c) is motivated from the following three examples:

- (Example 1) Let us first consider the continuous time framework with $0 < t < T$. For a given m, let $S_1, S_2, \cdots, S_m$ be m candidates for the break dates that are independently uniformly distributed on $(0, T)$. Then, the pdf of S_j is $1/T$ for $j = 1, \cdots, m$. In this case the break points $T_1, T_2, \cdots, T_m$ correspond to the order statistics $S_{(1)}, S_{(2)}, \cdots, S_{(m)}$ where $S_{(j)}$ is the jth smallest value among $S_1, S_2, \cdots, S_m$. As a result, the joint pdf of $T_1, T_2, \cdots, T_m$ is given by $m!/T^m$. This motivates us to assume that for discrete time $t = 1, 2, \cdots, T$ the joint probability function of $T_1, T_2, \cdots, T_m$ is proportional to $1/T^m$

16

and given by $m!/\prod_{j=1}^{m}(1 - j/T)/T^m$. With the assumption that the prior probability of m is given by $1/(\bar{m}+1)$ (the uniform prior for $m = 0, 1, \cdots, \bar{m}$), the prior probability of a given model becomes $\alpha_{m,T}/T^m$ where $\alpha_{m,T} = \{m!/(\bar{m} + 1)\}/\prod_{j=1}^{m}(1 - j/T)$.

- (Example 2) Let us again consider the continuous time framework as in Example 1 and suppose that the prior for m is a Poisson process with mean $T\beta$ where the prior for β is noninformative ($\beta > 0$). In this case, the pdf of $T_1, T_2, \cdots, T_m$ conditional on m is given by $m!/T^m$ (Theorem 2.3.1 in Ross, 1996) and hence the prior probability of a given model becomes $\int_0^{\infty} e^{-T\beta}(T\beta)^m/m!(m!/T^m)d\beta = m!/T^{m+1}$. This motivates us to assume the prior given in Assumption A7 (c) in the discrete time framework.

- (Example 3) Suppose that the probability of the occurrence of structural change at each time is given by a Binomial distribution with parameter ρ, for which the prior is uniform on $(0, 1)$. In this case, the prior probability of a given model becomes $\int_0^1 \rho^m (1 - \rho)^m d\rho = m!/\prod_{j=1}^{m}\{1 - (j - 1)/T\}/(T + 1)/T^m$, which is of the same form as in Assumption A7 (c).

Let the MLE of $\theta = [\phi', \sigma']'$ for a given set of break points $\mathcal{T}$ be $\check{\theta} = [\check{\phi}', \check{\sigma}']'$ where $\check{\phi} = [\check{\phi}_1, \check{\phi}_2, \cdots, \check{\phi}_{m+1}]'$ and $\check{\sigma} = [\text{vec}(\check{\Sigma}_1)', \text{vec}(\check{\Sigma}_2)', \cdots, \text{vec}(\check{\Sigma}_{m+1})']'$. Note that $\hat{\theta}$ is the global MLE with $\mathcal{T} = \hat{\mathcal{T}}$ whereas $\check{\theta}$ is obtained for an arbitrary given $\mathcal{T}$. Thus, $\check{\theta}$ is different from $\hat{\theta}$ in general and they are the same only in the case where $\mathcal{T} = \hat{\mathcal{T}}$.

While the second term on the right hand side of (19) is given by $2m \log T - 2 \log \alpha_{m,T}$, we need to evaluate the first term to obtain the modified BIC.

Proposition 3 *Under Assumption A7, the logarithm of the marginal pdf of y given x is expressed as*

$$\log g_M(y|x) = \ell_{m,p_x}(\mathcal{T}, \check{\theta}|y, x) - \sum_{j=1}^{m+1} \frac{p_{\phi_j} + p_{\sigma}}{2} \log(T_j - T_{j-1}) + O_p(1). \tag{20}$$

Proposition 3 suggests that, since the second term on the right hand side of (19) is given

by $2m \log T + O(1)$ from Assumption A7 (c), we should minimize

$$
\begin{aligned}
&-2 \log \{g_M(y|x) P(M(m, p_x, \mathcal{T}))\} \\
&= -2\ell_{m,p_x}(\mathcal{T}, \check{\theta}|y, x) + \sum_{j=1}^{m+1} \left(p_{\phi_j} + p_\sigma\right) \log(T_j - T_{j-1}) + 2m \log T + O_p(1).
\end{aligned}
$$

It is not difficult to see that the first term on the right hand side of the above equation domi-
nates the other terms and hence we only have to consider the minimization of $-\ell_{m,p_x}(\mathcal{T}, \check{\theta}|y, x)$
as long as T is sufficiently large. Since $\min_{\mathcal{T}} -\ell_{m,p_x}(\mathcal{T}, \check{\theta}|y, x) = -\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x)$, we replace
$\mathcal{T}$ and $\check{\theta}$ with the MLEs $\hat{\mathcal{T}}$ and $\hat{\theta}$ and propose the following modified BIC:

$$
MBIC1(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + \sum_{j=1}^{m+1} \left(p_{\phi_j} + p_\sigma\right) \log(\hat{T}_j - \hat{T}_{j-1}) + 2m \log T. \tag{21}
$$

Since $\log(\hat{T}_j - \hat{T}_{j-1}) = \log T + \log(\hat{\lambda}_j - \hat{\lambda}_{j-1}) = \log T + O_p(1)$, we may also simplify the
modified BIC (21) as

$$
MBIC2(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + \left(p_\phi^{all} + p_\sigma^{all} + 2m\right) \log T. \tag{22}
$$

It is not difficult to interpret the penalty terms of the two modified BICs. The second
term on the right hand side of (21) and $(p_\phi^{all} + p_\sigma^{all}) \log T$ of (22) are the penalty on the
additional unknown coefficients and variance components while $2m \log T$ can be interpreted
as the penalty on the uncertainty of the break points.

Since the penalty of the classical BIC on the unknown coefficients is given by $p_\phi^{all} \log T$,
we can see from (21) that $MBIC1$ will tend to choose more regressors than the classical BIC
for a given $m \geq 1$. We also note that the modified BIC (22) takes a form similar to Yao's
(1988) BIC, which is given by

$$
MBIC_y(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + (p_\phi^{all} + p_\sigma^{all} + m) \log T. \tag{23}
$$

On comparing the penalty term, we can see that our modified BIC will tend to choose less
number of structural changes than Yao's BIC.[3]

---

[3]Yao (1988) obtained the modified BIC by treating $T_1, T_2, \cdots, T_m$ as unknown parameters; counting the
total number of unknown parameters in the model, which equals $p_\phi^{all} + p_\sigma^{all} + m$; and inserting this number
into the formula of the classical BIC. Since we treat the change points in a different way, our modified BIC
does not coincide with Yao's BIC.

3.4. Consistency

In this subsection we investigate whether or not the model selection criteria derived in this paper can choose the appropriate regressors and the true number of structural changes. It is well known that the BIC can consistently choose the true lag length for time series models with no structural changes while the AIC and the C_p criterion tend to choose longer lags than the true one. See, for example, Shibata (1976), Hannan (1980) and Hannan and Deistler (1988). Thus, we expect that the modified BIC can consistently choose the regressors while the modified AIC and the modified C_p criterion may choose a larger set of regressors than the true ones. In addition, since our two modified BICs take a form similar to Yao's (1988) BIC, which is proved to consistently estimate the number of breaks for a local level Gaussian model, we expect that our modified BIC may have the same property.

In the following, we conventionally use the statement "$p_x > p_x^0$," which means that the true regressors are included in each regime and there are extra regressors at least in one regime. On the other hand, "$p_x < p_x^0$" implies that some true regressors are excluded at least from one regime irrespective of whether or not the extra regressors are included in some regimes. In the case of "$p_x = p_x^0$" each regime includes only the true regressors. Then, $\{p_x < p_x^0\} \cup \{p_x > p_x^0\} \cup \{p_x = p_x^0\}$ covers all the possible choices of the regressors.

To investigate the consistency of the estimated p_x and m, let $MIC(m, p_x)$ be a general expression of the model selection criterion defined as follows:

$$MIC(m, p_x) = -2\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) + \sum_{j=1}^{m+1}(p_{\phi_j} + p_\sigma)g_1(T) + mg_2(T), \qquad (24)$$

where $g_1(T)$ and $g_2(T)$ are sequences of positive non-decreasing numbers. Suppose that $m^0 \leq \bar{m}$ and that $\bar{p}_x$ includes all the true regressors. Let $\hat{m}$ and $\hat{p}_x$ be chosen such that the MIC is minimized over $0 \leq m \leq \bar{m}$ and among the $\bar{p}_x$ regressors.

Proposition 4 *Assume that Assumptions A1-A6 hold. (i) If $g_1(T) \to \infty$ while $g_i(T)/(Tv_T^2) \to 0$ for $i = 1$ and 2, then $P(\hat{m} = m_0$ and $\hat{p}_x = p_x^0) \to 1$ as $T \to \infty$.*
(ii) If $g_2(T) \to \infty$ and $g_2(T)/(Tv_T^2) \to 0$ while $g_1(T) = O(1)$, then $P(\hat{m} = m_0$ and $\hat{p}_x \geq p_x^0) \to 1$ as $T \to \infty$.
(iii) If $g_i(T) = O(1)$ for $i = 1$ and 2, then $P(\hat{m} \geq m_0$ and $\hat{p}_x \geq p_x^0) \to 1$ as $T \to \infty$.

Proposition 4(i) implies that the divergence of $g_1(T)$ guarantees the consistency of both $\hat{m}$ and $\hat{p}$ irrespective whether or not $g_2(T)$ goes to infinity. Intuitively, this is because the consistency of $\hat{m}$ ($\hat{p}$) requires the divergence of the penalty term when the extra breaks (extra regressors) are included. Since the coefficient associated with $g_1(T)$ increases when either the extra regressors or the extra breaks are included, we have Proposition 4(i). In the case of (ii) the penalty term does not diverge when the extra regressors are included and hence there is a positive probability of $\hat{p}_x > p_x^0$. (iii) can be interpreted similarly.

From Proposition 4(i) we can see that $\hat{m}$ and $\hat{p}_x$ based on the modified BICs are consistent. Of interest is that both the classical BIC and Yao's (1988) BIC also have the consistent property. On the other hand, the estimators based on the modified AIC are not consistent but they tend to be greater than m^0 and p_x^0 with positive probability from Proposition 4(iii). Similarly, we deduce that the modified C_p criterion does not deliver consistent estimators of m and p_x. Therefore, we can say that the modified BICs have more plausible property than the modified AIC and the modified C_p criterion, at least asymptotically. However, as we will see in the next section, this is not always the case in finite samples.

4. Finite Sample Property

In this section we investigate the finite sample property of the model selection criteria developed in the previous section. We consider univariate AR(1), AR(2) and MA(1) processes generated by $\varepsilon_t \sim i.i.d.N(0,1)$ possibly with structural changes as the data generating process (DGP) and examine the performance of the modified criteria by estimating the lag length p_y as well as the number of breaks m.

In the case of no breaks, the DGP is given by

$$\text{DGP0} \quad : \quad y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad : \quad 1 \leq t \leq T,$$

where the sets of the parameters are summarized in the first panel of Table 1. DGP0AR1a corresponds to the AR(1) case with weak positive serial correlation while y_t is moderately serially correlated for DGP0A1b. DGP0AR2a-b are the AR(2) cases with real valued characteristic roots, whereas DGP0AR2c-d have complex roots. We choose these parameters so

that $\phi_1 + \phi_2 = 0.3$ (the case of weak serial correlation) or 0.7 (the case of moderate serial correlation). Similarly, we consider two DGPs for the MA(1) case (DGP0MA1a-b).

A process with one time break is generated by

$$
\text{DGP1} \quad : \quad \begin{cases} y_t = c_1 + \phi_{11}y_{t-1} + \phi_{21}y_{t-2} + \varepsilon_t - \theta_{11}\varepsilon_{t-1} & : & 1 \le t \le [T/2] \\ y_t = c_2 + \phi_{12}y_{t-1} + \phi_{22}y_{t-2} + \varepsilon_t - \theta_{12}\varepsilon_{t-1} & : & [T/2] + 1 \le t < T, \end{cases}
$$

where the sets of the parameters are given in the second panel of Table 1. For the AR(1) (DGP1AR1a-b) and AR(2) (DGP1AR2a-d) cases, the sum of the AR coefficients changes from 0.7 to 0.3 or 0.3 to 0.7. The MA(1) case (DGP1MA1a-b) has similar structural changes.

The DGP with two time breaks is given by

$$
\text{DGP2} \quad : \quad \begin{cases} y_t = c_1 + \phi_{11}y_{t-1} + \phi_{21}y_{t-2} + \varepsilon_t - \theta_{11}\varepsilon_{t-1} & : & 1 \le t \le [T/3] \\ y_t = c_2 + \phi_{12}y_{t-1} + \phi_{22}y_{t-2} + \varepsilon_t - \theta_{12}\varepsilon_{t-1} & : & [T/3] + 1 \le t < [2T/3] \\ y_t = c_3 + \phi_{13}y_{t-1} + \phi_{23}y_{t-2} + \varepsilon_t - \theta_{13}\varepsilon_{t-1} & : & [2T/3] + 1 \le t < T, \end{cases}
$$

where the sets of the parameters are summarized in the last panel of Table 1. For DGP2AR1a, AR2a, AR2c and MA1a, serial correlation is weakened by the first break but the process returns to the first regime after the second break (the first and third regimes have the same parameters). DGP2AR1b, AR2b, AR2d and MA1b correspond to the case where the level of the process goes down and the process becomes less persistent gradually because of structural changes.

We set $T = 120$ or 300 while the trimming parameter ϵ is set as 0.05 or 0.15. All computations are carried out by using the GAUSS matrix language with 5,000 replications.[4]

We estimate the AR(p_y) model including a constant with structural changes and select the lag length p_y and the number of breaks m based on the modified AIC in (12), the modified C_p criterion in (17) and the two modified BICs in (21) and (22), with a restriction such that the lag lengths are the same in all the regimes. We set $\bar{p}_y = 4$ and $\bar{m} = 5$ so that we choose a model from among $0 \le p_y \le 4$ and $0 \le m \le 5$.

To see the effect of our modification, we also estimate p_y and m by the classical model selection criteria and Yao's (1988) modified BIC. In addition, we compare the finite sample performance of the model selection criteria with that of the sequential testing procedure.

---

[4] We need to efficiently calculate the maximized likelihood for the case of multiple structural changes to save computational time. The method is explained in Bai and Perron (1998, 2003) and we made use of the program provided by them.

Since we do not know the true lag length, we first estimate the number of breaks by the testing procedure that is robust to heteroskedasticity and autocorrelation, and then estimate the lag length using the estimated number of breaks. More precisely, following the suggestion by Bai and Perron (2006), we first test for the null of no breaks using the $UDmax$ test at the 5% significance level allowing different second moments of the regressors as well as the heterogeneity of the variances. Since p_y is unknown, we regress y_t on a constant in each regime and construct the test statistic using the autocorrelation and heteroskedasticity consistent (HAC) estimate of the variance of the error terms with the prewhitening method. If the null hypothesis is rejected, we continuously use the $\sup F(\ell+1|\ell)$ test constructed in the same way as the $UDmax$ test until it cannot reject the hypothesis. Once the number of breaks is estimated, the lag length is estimated using the Wald test by the general to specific rule. This robust testing procedure is denoted by "$Sq(rb)$".

We also consider the hybrid of the modified criteria with the testing procedure; we first estimate p_y and m by the modified criteria and using the estimated lag length, estimate the number of breaks with the testing procedure.[5] Note that we do not use the HAC estimates of the variance to construct the test statistics in this case because the lag length is estimated. The hybrid method is denote by "$Sq(MIC)$." For example, "$Sq(MAIC)$" signifies that the lag length is selected by the modified AIC while the number of breaks is estimated by the testing procedure.

Table 2a reports the frequencies of selecting the true model for the case of no breaks. The entries for the AR(1) and AR(2) cases are the frequencies of $\hat{p}_y = 1$ and $\hat{m} = 0$, respectively. We focus only on the estimation of the number of breaks for the MA(1) case and the entries in this case correspond to the frequencies of $\hat{m} = 0$ irrespective of any values of $\hat{p}_y$, because any finite order lags are incorrect. From the panels DGP0AR1a-b, we can see that the classical AIC and C_p criterion rarely choose the true model for the AR(1) case while the classical BIC has a better finite sample property when $T = 300$, although its performance is not necessarily satisfactory when $T = 120$. This poor finite sample performance of the classical criteria is dramatically improved by our modification; in particular, the modified BICs have

---

[5]We also conducted simulations for the hybrid of the classical model selection criteria, such as AIC and BIC, with the testing procedure, but the performance is poor and we do not report the results.

a high probability of selecting the true model, as expected from Proposition 4. The problem of the classical criteria is that they tend to choose large number of structural changes. For example, in the case of DGP0AR1a with $T = 120$ and $\epsilon = 0.05$ the probabilities for AIC, BIC and C_p when $\hat{m} = 5$ and $\hat{p} = 0$ are 0.456, 0.403 and 0.475, respectively. This tendency to over-estimate is well corrected by including the penalty term on the additional breaks. Comparing the modified criteria with the robust testing procedure and the hybrid methods, we find that our modified criteria work better when $T = 120$. Further, while the hybrid methods perform better than the modified AIC and the modified C_p criterion in some cases, they are not as good as the modified BICs in this case. We also note that all the methods perform better for larger T and larger ϵ, but the modified criteria are not as sensitive to the value of ϵ as the testing procedure and the hybrid method. In particular, when $T = 120$ and $\epsilon = 0.05$, the sequential testing procedure and the hybrid method do not work well.

For the AR(2) case with DGP0AR2a, it is difficult for all the methods to choose the true model. This is because the coefficient associated with y_{t-2} is so small that shorter lags tend to be selected. Because the penalty of the modified AIC and the modified C_p criterion is not as heavy as that of the modified BICs, the former two methods choose the true model with a higher probability. For the other AR(2) case (DGP0AR2b-d) and the MA(1) case (DGP0MA1a-b), the overall performance is similar to the AR(1) case.

Table 2b reports the result for the case of one time break. As in the case of no breaks, the classical criteria do not perform well because they tend to choose larger breaks; this tendency to over-estimate is fixed by our modification. In this case, while the hybrid method with the modified BICs tends to choose the true model more frequently than the modified BICs for the AR(1) and AR(2) cases, the relation is reversed for the MA(1) case. However, the performance of both the methods is not satisfactory for DGP1AR2a-b.

The result for the case of two time breaks is summarized in Table 2c. As a whole, it seems difficult to choose the true model for the AR(1) and AR(2) cases, especially, when $T = 120$. The modified AIC works best among the modified criteria but its performance is dominated by the hybrid method in many cases. On the other hand, the modified BICs work better than the hybrid method for the MA(1) case, although both perform quite well in this case.

23

To summarize our simulation results, we can say that the modified BICs perform relatively well when $m^0 \leq 1$ and $\epsilon = 0.05$, while the hybrid method with the modified BICs, that is, the sequential testing procedure with the lag order selected by the modified BICs, may be recommended if one is confident that the distance between the two consecutive break fractions is not so close, such as $\epsilon = 0.15$, or if it is believed that the model has more than one break with a high probability.

5. Conclusion

This paper developed the model selection criteria to select the regressors and the number of structural changes in multivariate regression models, including a VAR model as a special case. We derived the modified AIC, the modified C_p criterion and the modified BICs. The penalty terms of these criteria are determined not in ad hoc ways but based on the risk functions given for the criteria. We showed that the modified BICs can consistently estimate the number of structural changes and the regressors while the modified AIC and the modified C_p criterion tend to choose a larger model with a positive probability. The consistency of the modified BICs is a plausible theoretical property and by reflecting this nice nature, the modified BICs perform well in finite samples. Because it is important to consistently estimate the number of breaks and given the simulation results, the modified BICs and the hybrid method are recommended to be implemented for practical analyses.

Appendix

Since all the model selection criteria are derived when $m = m^0$ and $p = p^0$, we omit the superscript 0 for notational convenience.

Proof of Lemma 1: In this proof we omit a subscript j for notational convenience. For example, $B_{1j}(v)$ is abbreviated as $B_1(v)$. As explained in Appendix B of Bai (1997a), $\max_{v \leq 0} B_1(v)$ and $\max_{v > 0} B_2(v)$ are distributed as exponential distributions with parameters γ_1/ω_1 and γ_2/ω_2, respectively, and hence

$$
\begin{aligned}
P\left(\max_v B^I(v) \leq b\right) &= P\left(\max\left\{\max_{v \leq 0} B_1(v),\ \max_{v > 0} B_2(v)\right\} \leq b\right) \\
&= P\left(\max_{v \leq 0} B_1(v) \leq b\right) P\left(\max_{v > 0} B_2(v) \leq b\right)
\end{aligned}
$$

$$= \left(1 - e^{-(\gamma_1/\omega_1)b}\right)\left(1 - e^{-(\gamma_2/\omega_2)b}\right),$$

where the second equality holds because $B_1(v)$ and $B_2(v)$ are independent. Then, the probability density function of $\max_v B^I(v)$ is given by

$$f(b) = (\gamma_1/\omega_1)e^{-(\gamma_1/\omega_1)b} + (\gamma_2/\omega_2)e^{-(\gamma_2/\omega_2)b} - \left\{(\gamma_1/\omega_1) + (\gamma_2/\omega_2)\right\}e^{-\{(\gamma_1/\omega_1)+(\gamma_2/\omega_2)\}b}.$$

Carrying out the integration $\int_{b>0} bf(b)db$ and letting $r_1 = \omega_1/\gamma_1$ and $r_2 = \omega_2/\gamma_2$, we obtain (5).

Next, let $\hat{v} = \operatorname{argmax}_v B^I(v)$. By change of variable with $s = (\gamma_1^2/\omega_1)v$ as in Qu and Perron (2007) we can see that

$$\hat{v} = \frac{\omega_1}{\gamma_1^2}\operatorname*{argmax}_s \tilde{B}^I(s), \quad \text{where} \quad \tilde{B}^I(s) = \begin{cases} W_1(|s|) - \frac{|s|}{2} & : \quad s \le 0 \\ \sqrt{r_\omega}W_2(s) - \frac{s}{2}r_\gamma & : \quad s > 0, \end{cases}$$

where $r_\omega = \omega_2/\omega_1$ and $r_\gamma = \gamma_2/\gamma_1$. Then, it is sufficient to calculate $E[\operatorname{argmax}_s \tilde{B}^I(s)1(s \le 0)]$ and $E[\operatorname{argmax}_s \tilde{B}^I(s)1(s > 0)]$ in order to obtain (6) and (7).

Following Appendix B of Bai (1997a) it can be shown that the probability density function (pdf) of $\hat{s} = \operatorname{argmax}_s \tilde{B}^I(s)$ is given by

$$g(s) = \begin{cases} -\frac{1}{2}\Phi\left(-\frac{1}{2}\sqrt{|s|}\right) + \left(\frac{r_\gamma}{r_\omega} + \frac{1}{2}\right)e^{(1/2)\{(r_\gamma/r_\omega)+(r_\gamma/r_\omega)^2\}|s|}\Phi\left(-\left(\frac{r_\gamma}{r_\omega} + \frac{1}{2}\right)\sqrt{|s|}\right) & : \quad s \le 0 \\ -\frac{(r_\gamma/\sqrt{r_\omega})^2}{2}\Phi\left(-\frac{(r_\gamma/\sqrt{r_\omega})}{2}\sqrt{s}\right) \\ \quad + \left(r_\gamma + \frac{(r_\gamma/\sqrt{r_\omega})^2}{2}\right)e^{(1/2)(r_\gamma+r_\omega)s}\Phi\left(-\left(\sqrt{r_\omega} + \frac{(r_\gamma/\sqrt{r_\omega})}{2}\right)\sqrt{s}\right) & : \quad s > 0 \end{cases},$$

which is obtained based on the result on an additive process by Bhattacharya and Brockwell (1976),[6] where $\Phi(\cdot)$ denotes a cumulative distribution function of a standard normal random variable. By carrying out the integration we obtain

$$\int_{-\infty}^0 sg(s)ds = -\frac{2r_\gamma(r_\gamma + 2r_\omega)}{(r_\gamma + r_\omega)^2} = -\frac{2r_1(r_1 + 2r_2)}{(r_1 + r_2)^2},$$

$$\int_0^\infty sg(s)ds = \frac{2r_\omega^2(r_\omega + 2r_\gamma)}{r_\gamma^2(r_\gamma + r_\omega)^2} = \frac{2}{r_\gamma}\frac{r_2^2(2r_1 + r_2)}{r_1(r_1 + r_2)^2},$$

where the second equalities of the above two integrals are obtained by using $r_\omega = (r_2/r_1)r_\gamma$. Noting that

$$E\left[a_j^I\left|\operatorname*{argmax}_s \tilde{B}^I(s)\right|\right] = a_2\int_0^\infty sg(s)ds - a_1\int_{-\infty}^0 sg(s)ds,$$

---

[6]This result is obtained by replacing ϕ and ξ in Bai (1997a) with r_ω and r_γ, respectively. Note that there are typos in equation (B.1) and the definition of $g(x)$ in Bai (1997a). The first term on the right hand side of $g(x)$ for $x < 0$ must have a negative sign.

$$E\left[a_j^I \operatorname*{argmax}_s \tilde{B}^I(s)\right] = a_2 \int_0^\infty sg(s)ds + a_1 \int_{-\infty}^0 sg(s)ds,$$

(6) and (7) are established.∎

Proof of Proposition 1: In this proof we restrict our analysis on the set given by $\{T_j | T_j = T_j^0 + cv_T^{-2}, \ -M \le c \le M\}$ for some large M $(j = 1, \cdots, m)$ because $\hat{T}_j - T_j = O_p(v_T^{-2})$ as shown by Qu and Perron (2007). We first evaluate $b_{m,p_x,1}$. Since $E_y[\ell_{m,p_x}(\mathcal{T}^0, \theta^0 | y, x)] = -(nT/2)\log(2\pi) - (1/2)\sum_{j=1}^{m+1} \Delta T_j^0 \log |\Sigma_j^0| - (nT/2)$, we can see that

$$\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta} | y, x) - E\left[\ell_{m,p_x}(\mathcal{T}^0, \theta^0 | y, x)\right] = R_{11} + R_{12}, \tag{25}$$

$$\text{where} \quad R_{11} = -\sum_{j=1}^{m+1} \left(\frac{\Delta \hat{T}_j}{2} \log |\hat{\Sigma}_j| - \frac{\Delta T_j^0}{2} \log |\Sigma_j^0|\right),$$

$$R_{12} = -\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{2} \left(\log |\hat{\Sigma}_j| - \log |\Sigma_j^0|\right) - \sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j - \Delta T_j^0}{2} \log |\Sigma_j^0|.$$

By expanding $\log |\hat{\Sigma}_j|$ around $\log |\Sigma_j^0|$, R_{11} is expressed as

$$
\begin{aligned}
R_{11} &= -\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{2} \left[\operatorname{tr}\left\{\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \Sigma_j^0\right)\right\} - \frac{1}{2}\operatorname{tr}\left\{\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \Sigma_j^0\right)\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \Sigma_j^0\right)\right\}\right] + o_p(1) \\
&= -\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{2} \left[\operatorname{tr}\left\{\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \tilde{\Sigma}_j\right)\right\} + \operatorname{tr}\left\{\Sigma_j^{0-1}\left(\tilde{\Sigma}_j - \Sigma_j^0\right)\right\} \right.\\
&\qquad\qquad\qquad \left. -\frac{1}{2}\operatorname{tr}\left\{\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \Sigma_j^0\right)\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \Sigma_j^0\right)\right\}\right] + o_p(1), \tag{26}
\end{aligned}
$$

where $\tilde{\Sigma}_j = \sum_{t=T_{j-1}^0+1}^{T_j^0} \tilde{\varepsilon}_t \tilde{\varepsilon}_t' / \Delta \hat{T}_j$ with $\tilde{\varepsilon}_t = y_t - (x_t' \otimes I_n)\hat{\phi}_j$ for $T_{j-1}^0 + 1 \le t \le T_j^0$.

For $\hat{T}_j < T_j^0$, the first term in the square brackets on the right hand side of (26) becomes

$$
\begin{aligned}
&-\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{2} \operatorname{tr}\left\{\Sigma_j^{0-1}\left(\hat{\Sigma}_j - \tilde{\Sigma}_j\right)\right\} \\
&= -\frac{1}{2}\sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \left\{\varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_{j+1} - \phi_j^0)\right\}' \Sigma_{j+1}^{0-1} \left\{\varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_{j+1} - \phi_j^0)\right\} \\
&\qquad\qquad - \left\{\varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_j - \phi_j^0)\right\}' \Sigma_j^{0-1} \left\{\varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_j - \phi_j^0)\right\} \\
&= -\frac{1}{2}\sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \operatorname{tr}\left(\Sigma_{j+1}^{0-1}\varepsilon_t\varepsilon_t'\right) - \operatorname{tr}\left(\Sigma_j^{0-1}\varepsilon_t\varepsilon_t'\right)
\end{aligned}
$$

26

$$-2(\hat{\phi}_{j+1} - \phi_j^0)' \sum_{t=\hat{T}_j+1}^{T_j^0} \left(x_t \otimes \Sigma_{j+1}^{0-1}\right) \varepsilon_t + (\hat{\phi}_{j+1} - \phi_j^0)' \sum_{t=\hat{T}_j+1}^{T_j^0} \left(x_t x_t' \otimes \Sigma_{j+1}^{0-1}\right) (\hat{\phi}_{j+1} - \phi_j^0) + o_p(1).$$

Since $\hat{\phi}_{j+1} - \phi_j^0 = (\phi_{j+1}^0 - \phi_j^0) + (\hat{\phi}_{j+1} - \phi_{j+1}^0) = v_T \delta_j + O_p(1/\sqrt{T})$ and

$$
\begin{aligned}
\Sigma_j^{0-1} &= \Sigma_{j+1}^{0-1} + \Sigma_{j+1}^{0-1}(\Sigma_{j+1}^0 - \Sigma_j^0)\Sigma_{j+1}^{0-1} + \Sigma_{j+1}^{0-1}(\Sigma_{j+1}^0 - \Sigma_j^0)\Sigma_j^{0-1}(\Sigma_{j+1}^0 - \Sigma_j^0)\Sigma_{j+1}^{0-1} \\
&= \Sigma_{j+1}^{0-1} + v_T \Sigma_{j+1}^{0-1/2} A_{2j} \Sigma_{j+1}^{0-1/2} + v_T^2 \Sigma_{j+1}^{0-1/2} A_{2j}^2 \Sigma_{j+1}^{0-1/2} + O(v_T^3),
\end{aligned}
$$

we can see that, for $\hat{T}_j < T_j^0$,

$$-\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{2} \text{tr} \left\{ \Sigma_j^{0-1} \left(\hat{\Sigma}_j - \tilde{\Sigma}_j \right) \right\} = \frac{1}{2} \sum_{j=1}^{m} \text{tr} \left(v_T A_{1j} \sum_{t=\hat{T}_j+1}^{T_j^0} \eta_t \eta_t' \right) - v_T^2 (T_j^0 - \hat{T}_j) \text{tr} \left(A_{1j}^2 \right)$$

$$+ 2 v_T \delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0} \left(x_t \otimes \Sigma_j^{0-1} \right) \varepsilon_t - v_T^2 \delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0} \left(x_t x_t' \otimes \Sigma_j^{0-1} \right) \delta_j + o_p(1). \quad (27)$$

Similarly, the second and third terms in the square brackets on the right hand side of (26) are expressed as

$$-\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{2} \text{tr} \left\{ \Sigma_j^{0-1} (\tilde{\Sigma}_j - \Sigma_j^0) \right\}$$

$$= -\frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \left\{ \varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_j - \phi_j^0) \right\}' \Sigma_j^{0-1} \left\{ \varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_j - \phi_j^0) \right\} + \frac{nT}{2}$$

$$= \frac{1}{2} \sum_{j=1}^{m+1} (\hat{\phi}_j - \phi_j^0) \sum_{t=T_{j-1}^0+1}^{T_j^0} \left(x_t x_t' \otimes \Sigma_j^{0-1} \right) (\hat{\phi}_j - \phi_j^0) - \frac{1}{2} \sum_{j=1}^{m+1} \text{tr} \left\{ \sum_{t=T_{j-1}^0+1}^{T_j^0} \left(\eta_t \eta_t' - I_n \right) \right\} + o_p(1).$$

$$(28)$$

$$\sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j}{4} \text{tr} \left\{ \Sigma_j^{0-1} \left(\hat{\Sigma}_j - \Sigma_j^0 \right) \Sigma_j^{0-1} \left(\hat{\Sigma}_j - \Sigma_j^0 \right) \right\} = \sum_{j=1}^{m+1} \frac{1}{4\Delta T_j^0} \text{tr} \left\{ \left[\sum_{t=T_{j-1}^0+1}^{T_j^0} \left(\eta_t \eta_t' - I_n \right) \right]^2 \right\} + o_p(1).$$

$$(29)$$

On the other hand, R_{12} becomes, for $\hat{T}_j < T_j^0$,

$$R_{12} = \frac{1}{2} \sum_{j=1}^{m} (T_j^0 - \hat{T}_j) \left(\log |\Sigma_j^0| - \log |\Sigma_{j+1}^0| \right)$$

$$= \frac{1}{2} \sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \mathrm{tr}\left(-v_T A_{1j}\right) + \frac{v_T^2 (T_j^0 - \hat{T}_j)}{2} \mathrm{tr}\left(A_{1j}^2\right) + o_p(1), \tag{30}$$

where the last equality is obtained by expanding $\log |\Sigma_{j+1}^0|$ around $\log |\Sigma_j^0|$.

Then, by combining (27)–(30), we have, for $\hat{T}_j < T_j^0$,

$$R_{11} + R_{12} + \frac{1}{2} \sum_{j=1}^{m+1} \mathrm{tr}\left\{ \sum_{t=T_{j-1}^0+1}^{T_j^0} \left(\eta_t \eta_t' - I_n\right) \right\}$$

$$= \frac{1}{2} \sum_{j=1}^{m+1} (\hat{\phi}_j - \phi_j^0) \sum_{t=T_{j-1}^0+1}^{T_j^0} \left(x_t x_t' \otimes \Sigma_j^{0-1}\right) (\hat{\phi}_j - \phi_j^0) + \sum_{j=1}^{m+1} \frac{1}{4\Delta T_j^0} \mathrm{tr}\left\{ \left[\sum_{t=T_{j-1}^0+1}^{T_j^0} \left(\eta_t \eta_t' - I_n\right) \right]^2 \right\}$$

$$+ \sum_{j=1}^{m} \mathrm{tr}\left\{ \frac{v_T}{2} A_{1j} \sum_{t=\hat{T}_j+1}^{T_j^0} \left(\eta_t \eta_t' - I_n\right) \right\} - \frac{v_T^2}{4} (T_j^0 - \hat{T}_j) \mathrm{tr}\left(A_{1j}^2\right)$$

$$+ v_T \delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0} \left(x_t \otimes \Sigma_j^{0-1}\right) \varepsilon_t - \frac{v_T^2}{2} \delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0} \left(x_t x_t' \otimes \Sigma_j^{0-1}\right) \delta_j + o_p(1)$$

$$\xrightarrow{d} \frac{1}{2} \sum_{j=1}^{m+1} \chi_{p_{\phi_j}}^2 + \sum_{j=1}^{m+1} \frac{\kappa_{4j}}{4} + \sum_{j=1}^{m} \max_v B_j^I(v), \tag{31}$$

where $\chi_{p_{\phi_j}}^2$ $(j = 1, \cdots, m+1)$ are independent chi-square distributions with p_{ϕ_j} degrees of freedom. Since the same convergence holds for $\hat{T}_j > T_j^0$ and the expectation of the left hand side of (31) equals $b_{m,p_x,1}$, we can see using Lemma 1 that, up to the $O(1)$ terms,

$$b_{m,p_x,1} = \frac{p_\phi^{all}}{2} + \sum_{j=1}^{m+1} \frac{\kappa_{4j}}{4} + \sum_{j=1}^{m} \frac{r_{1j}^2 + r_{1j} r_{2j} + r_{2j}^2}{r_{1j} + r_{2j}}.$$

We next evaluate $b_{m,p_x,3}$ because $b_{m,p_x,2} = 0$ is obvious. Since

$$E_{y^*}[\ell_{m,p_x}(T^0, \hat{\theta}|y^*, x^*)] = -\frac{nT}{2} \log(2\pi) - \sum_{j=1}^{m+1} \frac{\Delta T_j^0}{2} \log |\hat{\Sigma}_j| - \frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \mathrm{tr}\left(\hat{\Sigma}_j^{-1} \Sigma_j^0\right)$$

$$- \frac{1}{2} \sum_{j=1}^{m+1} \Delta T_j^0 (\hat{\phi}_j - \phi_j^0)' \left(E_{y^*}[x_t^* x_t^{*\prime}] \otimes \hat{\Sigma}_j^{-1}\right) (\hat{\phi}_j - \phi_j^0),$$

28

we can see that

$$E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \theta^0 | y^*, x^*)] - E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta} | y^*, x^*)]$$

$$= \sum_{j=1}^{m+1} \frac{\Delta T_j^0}{2} \left(\log |\hat{\Sigma}_j| - \log |\Sigma_j^0| \right) - \frac{nT}{2} + \frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \operatorname{tr} \left(\hat{\Sigma}_j^{-1} \Sigma_j^0 \right)$$

$$+ \frac{1}{2} \sum_{j=1}^{m+1} \Delta T_j^0 (\hat{\phi}_j - \phi_j^0)' \left(E_{y^*}[x_t^* x_t^{*\prime}] \otimes \hat{\Sigma}_j^{-1} \right) (\hat{\phi}_j - \phi_j^0)$$

$$= \sum_{j=1}^{m+1} \frac{\Delta T_j^0}{4} \operatorname{tr} \left\{ \Sigma_j^{0-1} (\hat{\Sigma}_j - \Sigma_j^0) \Sigma_j^{0-1} (\hat{\Sigma}_j - \Sigma_j^0) \right\}$$

$$+ \frac{1}{2} \sum_{j=1}^{m+1} \Delta T_j^0 (\hat{\phi}_j - \phi_j^0)' \left(E_{y^*}[x_t^* x_t^{*\prime}] \otimes \hat{\Sigma}_j^{-1} \right) (\hat{\phi}_j - \phi_j^0) + o_p(1)$$

$$\xrightarrow{d} \sum_{j=1}^{m+1} \frac{\kappa_{4j}}{4} + \frac{1}{2} \sum_{j=1}^{m+1} \chi_{p_{\phi_j}}^2, \tag{32}$$

where the second equality is obtained by expanding $\log |\hat{\Sigma}_j|$ around $\log |\Sigma_j^0|$ and by using the relation

$$\operatorname{tr} \left(\hat{\Sigma}_j^{-1} \Sigma_j^0 \right) = n - \operatorname{tr} \left\{ \Sigma_j^{0-1} (\hat{\Sigma}_j - \Sigma_j^0) \right\} + \operatorname{tr} \left\{ \Sigma_j^{0-1} (\hat{\Sigma}_j - \Sigma_j^0) \hat{\Sigma}_j^{-1} (\hat{\Sigma}_j - \Sigma_j^0) \right\},$$

which holds because

$$\hat{\Sigma}_j^{-1} = \Sigma_j^{0-1} - \Sigma_j^{0-1} (\hat{\Sigma}_j - \Sigma_j^0) \Sigma_j^{0-1} + \Sigma_j^{0-1} (\hat{\Sigma}_j - \Sigma_j^0) \hat{\Sigma}_j^{-1} (\hat{\Sigma}_j - \Sigma_j^0) \Sigma_j^{0-1}.$$

From (32), we have, up to the $O(1)$ terms,

$$b_{m,p_x,3} = E_y \left[E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \theta^0 | y^*, x^*)] - E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta} | y^*, x^*)] \right] = \sum_{j=1}^{m+1} \frac{\kappa_{4j}}{4} + \frac{p_\phi^{all}}{2}.$$

For $b_{m,p_x,4}$, we write

$$E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta} | y^*, x^*)] - E_{y^*}[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta} | y^*, x^*)] = R_{41} + R_{42} + R_{43},$$

$$\text{where} \qquad R_{41} = \sum_{j=1}^{m+1} \frac{\Delta \hat{T}_j - \Delta T_j^0}{2} \log |\hat{\Sigma}_j|,$$

$$R_{42} = \frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \left\{ y_t^* - (x_t^{*\prime} \otimes I_n)\hat{\phi}_j \right\}' \hat{\Sigma}_j^{-1} \left\{ y_t^* - (x_t^{*\prime} \otimes I_n)\hat{\phi}_j \right\},$$

$$R_{43} = -\frac{1}{2} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \left\{ y_t^* - (x_t^{*\prime} \otimes I_n)\hat{\phi}_j \right\}' \hat{\Sigma}_j^{-1} \left\{ y_t^* - (x_t^{*\prime} \otimes I_n)\hat{\phi}_j \right\}.$$

Similarly to the evaluation of R_{12}, by the Taylor expansion of $\log|\hat{\Sigma}_{j+1}|$, R_{41} can be expressed as, for $\hat{T}_j < T_j^0$,

$$
\begin{aligned}
R_{41} &= \sum_{j=1}^{m} \frac{\hat{T}_j - T_j^0}{2} \left(\log|\hat{\Sigma}_j| - \log|\hat{\Sigma}_{j+1}| \right) \\
&= \frac{1}{2} \sum_{j=1}^{m} (T_j^0 - \hat{T}_j) \left[\mathrm{tr}\left\{ \hat{\Sigma}_j^{-1} \left(\hat{\Sigma}_{j+1} - \hat{\Sigma}_j \right) \right\} \right. \\
&\qquad\qquad \left. - \frac{1}{2}\mathrm{tr}\left\{ \hat{\Sigma}_j^{-1} \left(\hat{\Sigma}_{j+1} - \hat{\Sigma}_j \right) \hat{\Sigma}_j^{-1} \left(\hat{\Sigma}_{j+1} - \hat{\Sigma}_j \right) \right\} \right] + o_p(1). \qquad (33)
\end{aligned}
$$

On the other hand, $R_{42} + R_{43}$ becomes, for $\hat{T}_j < T_j^0$,

$R_{42} + R_{43}$
$$
\begin{aligned}
&= E_{y^*} \left[\frac{1}{2} \sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \left\{ \varepsilon_t^* - (x_t^{*\prime} \otimes I_n)(\hat{\phi}_{j+1} - \phi_j^0) \right\}' \hat{\Sigma}_{j+1}^{-1} \left\{ \varepsilon_t^* - (x_t^{*\prime} \otimes I_n)(\hat{\phi}_{j+1} - \phi_j^0) \right\} \right. \\
&\qquad\qquad \left. - \left\{ \varepsilon_t^* - (x_t^{*\prime} \otimes I_n)(\hat{\phi}_j - \phi_j^0) \right\}' \hat{\Sigma}_j^{-1} \left\{ \varepsilon_t^* - (x_t^{*\prime} \otimes I_n)(\hat{\phi}_j - \phi_j^0) \right\} \right] \\
&= \frac{1}{2} \sum_{j=1}^{m} (T_j^0 - \hat{T}_j) v_T^2 \delta_j' \left(E[x_t^* x_t^{*\prime}] \otimes \hat{\Sigma}_{j+1}^{-1} \right) \delta_j - (T_j^0 - \hat{T}_j)\mathrm{tr}\left\{ \hat{\Sigma}_j^{-1} \left(\hat{\Sigma}_{j+1} - \hat{\Sigma}_j \right) \hat{\Sigma}_j^{-1}\Sigma_j^0 \right\} \\
&\qquad + (T_j^0 - \hat{T}_j)\mathrm{tr}\left\{ \hat{\Sigma}_j^{-1} \left(\hat{\Sigma}_{j+1} - \hat{\Sigma}_j \right) \hat{\Sigma}_{j+1}^{-1} \left(\hat{\Sigma}_{j+1} - \hat{\Sigma}_j \right) \hat{\Sigma}_j \Sigma_j^0 \right\} + o_p(1). \qquad (34)
\end{aligned}
$$

Thus, by combining (33) and (34), we have, for $\hat{T}_j < T_j^0$,

$$
\begin{aligned}
&E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \theta^0|y^*)] - E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta}|y^*)] \\
&= \sum_{j=1}^{m} v_T^2 (T_j^0 - \hat{T}_j) \left\{ \frac{1}{2}\delta_j' \left(E[x_t^* x_t^{*\prime}] \otimes \hat{\Sigma}_j^{-1} \right) \delta_j + \frac{1}{4}\mathrm{tr}\left(A_{1j}^2 \right) \right\} + o_p(1) \\
&\xrightarrow{d} \sum_{j=1}^{m} \frac{\gamma_j^I}{2} \left| \mathrm{argmax}_v B_j^I(v) \right|,
\end{aligned}
$$

where $\gamma_j^I = \gamma_{1j}$ when $v \leq 0$ and $\gamma_j^I = \gamma_{2j}$ when $v > 0$. Since the same convergence holds for $\hat{T}_j > T_j^0$, we have, by Lemma 1,

$$
b_{m,p_x,4} = E_y \left[E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \theta^0|y^*)] - E_{y^*}[\ell_{m,p_x}(\mathcal{T}^0, \hat{\theta}|y^*)] \right] = \sum_{j=1}^{m} \frac{r_{1j}^2 + r_{1j}r_{2j} + r_{2j}^2}{r_{1j} + r_{2j}}
$$
$$(35)$$

up to the $O(1)$ terms.∎

Proof of Proposition 2: $E[J_{m,p_x,1}] = -nT$ is obvious. For $J_{m,p_x,2}$ we expand it as, for $\hat{T}_j < T_j^0$,

$$
\begin{aligned}
J_{m,p_x,2} &= 2\sum_{j=1}^{m+1}\sum_{t=T_{j-1}^0+1}^{T_j^0} \left\{(\hat{\mu}_t - \tilde{\mu}_t) + (\tilde{\mu}_t - \mu_t^0)\right\}' \Sigma_j^{0-1}\varepsilon_t \\
&= 2\sum_{j=1}^{m}\sum_{t=\hat{T}_j+1}^{T_j^0} \left\{(x_t' \otimes I_n)(\hat{\phi}_{j+1} - \hat{\phi}_j)\right\}' \Sigma_j^{0-1}\varepsilon_t \\
&\quad + 2\sum_{j=1}^{m+1}\sum_{t=T_{j-1}^0+1}^{T_j^0} \left\{(x_t' \otimes I_n)(\hat{\phi}_j - \phi_j^0)\right\}' \Sigma_j^{0-1}\varepsilon_t \\
&= 2\sum_{j=1}^{m}v_T\delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0}\left(x_t \otimes \Sigma_j^{0-1}\right)\varepsilon_t - v_T\delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0}\left(x_t x_t' \otimes \Sigma_j^{0-1}\right)\delta_j \\
&\quad + v_T\delta_j' \sum_{t=\hat{T}_j+1}^{T_j^0}\left(x_t x_t' \otimes \Sigma_j^{0-1}\right)\delta_j \\
&\quad + 2\sum_{j=1}^{m+1}(\hat{\phi}_j - \phi_j^0)' \sum_{t=T_{j-1}^0+1}^{T_j^0}\left(x_t x_t' \otimes \Sigma_j^{0-1}\right)(\hat{\phi}_j - \phi_j^0) + o_p(1) \\
&\xrightarrow{d} \sum_{j=1}^{m}\left\{2\max_v B_j^I(v) + \gamma_j^I \left|\operatorname*{argmax}_v B_j^I(v)\right|\right\} + 2\sum_{j=1}^{m+1}\chi_{p_{\phi_j}}^2,
\end{aligned}
$$

where, under the assumption of $\Sigma_{j+1} - \Sigma_j = v_T^2\Psi_j$, $\gamma_j^I = \gamma_{1j} = \delta_j'Q_{1j}\delta_j$ when $v \leq 0$ and $\gamma_j^I = \gamma_{2j} = \delta_j'Q_{2j}\delta_j$ when $v > 0$, $B_j^I(v)$ is defined as (4) with $\omega_{1j} = \gamma_{1j} = \delta_j'Q_{1j}\delta_j$ and $\omega_{2j} = \gamma_{2j} = \delta_j'Q_{2j}\delta_j$ because only the changes in the coefficients affect the limiting distributions of the break points in this case. Note that the same convergence holds for $\hat{T}_j > T_j^0$. From Lemma 1 we can see that

$$
E\left[2\max_v B_j^I(v)\right] = E\left[\gamma_j^I \left|\operatorname*{argmax}_v B_j^I(v)\right|\right] = 3
$$

in this case. Hence, we obtain $E[J_{m,p_x,2}] = 6m + 2p_\phi^{all}$ up to the $O(1)$ terms.

Similarly, using $\Sigma_{j+1}^0 - \Sigma_j^0 = v_T^2 \Psi_j$, $J_{m,p_x,3}$ is expressed as, for $\hat{T}_j < T_j^0$,

$$
\begin{aligned}
J_{m,p_x,3} &= -\sum_{j=1}^{m+1} \left(\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} \hat{\varepsilon}_t' \Sigma_j^{0-1} \hat{\varepsilon}_t - \sum_{t=T_{j-1}^0+1}^{T_j^0} \hat{\varepsilon}_t' \Sigma_j^{0-1} \hat{\varepsilon}_t \right) \\
&= -\sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \hat{\varepsilon}_t' (\Sigma_{j+1}^{0-1} - \Sigma_j^{0-1}) \hat{\varepsilon}_t \\
&= \sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \left\{ \varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_{j+1} - \phi_j^0) \right\}' \\
&\qquad \left\{ v_T^2 \Sigma_j^{0-1/2} A_{1j} \Sigma_j^{0-1/2} - v_T^4 \Sigma_j^{0-1/2} A_{1j}^2 \Sigma_j^{0-1/2} \right\} \left\{ \varepsilon_t - (x_t' \otimes I_n)(\hat{\phi}_{j+1} - \phi_j^0) \right\} \\
&= \sum_{j=1}^{m} \operatorname{tr} \left(v_T^2 A_{1j} \sum_{t=\hat{T}_j+1}^{T_j^0} \eta_t \eta_t' \right) + o_p(1) \\
&\Rightarrow -\sum_{j=1}^{m} \operatorname{tr}(A_j^I) \operatorname*{argmax}_{v} B_j^I(v), \quad\quad\quad\quad (36)
\end{aligned}
$$

and the same convergence holds for $\hat{T}_j > T_j^0$. Then, from Lemma 1 we can see, up to the $O(1)$ term, that

$$
E[J_{m,p_x,3}] = E\left[-\sum_{j=1}^{m} \operatorname{tr}(A_j^I) \operatorname*{argmax}_{v} B_j^I(v) \right] = \sum_{j=1}^{m} \frac{3}{2} \left(\frac{\operatorname{tr}(A_{1j})}{\gamma_{1j}} - \frac{\operatorname{tr}(A_{2j})}{\gamma_{2j}} \right).
$$

The limiting distribution in (15) is obtained similarly to (36) under Assumption A4. ∎

Proof of Proposition 3: We first note that the logarithm of $g_M(y|x)$ can be expressed as

$$
\log g_M(y|x) = \log \int \exp \left\{ \ell_{m,p_x}(\mathcal{T}, \theta | y, x) \right\} \pi_M(\theta) d\theta \quad\quad\quad\quad (37)
$$

where $d\theta = d\phi d\Sigma^{-1}$ with $d\phi = d\phi_1 d\phi_2 \cdots d\phi_{m+1}$ and $d\Sigma^{-1} = d\Sigma_1^{-1} d\Sigma_2^{-1} \cdots d\Sigma_{m+1}^{-1}$ from Assumption A7 (b). We expand the log likelihood as

$$
\begin{aligned}
\ell_{m,p_x}(\mathcal{T}, \theta | y, x) &= \ell_{m,p_x}(\mathcal{T}, \check{\theta} | y, x) - \left[\ell_{m,p_x}(\mathcal{T}, \check{\phi}, \sigma | y, x) - \ell_{m,p_x}(\mathcal{T}, \phi, \sigma | y, x) \right] \\
&\qquad - \left[\ell_{m,p_x}(\mathcal{T}, \check{\theta} | y, x) - \ell_{m,p_x}(\mathcal{T}, \check{\phi}, \sigma | y, x) \right] \\
&= \ell_{m,p_x}(\mathcal{T}, \check{\theta} | y, x) + L_1(\phi, \sigma) + L_2(\sigma), \quad \text{say.} \quad\quad (38)
\end{aligned}
$$

Since ϕ appears only in L_1 while both L_1 and L_2 depend on σ, we first evaluate the integral of $\exp(L_1)$ with respect to ϕ and next obtain the integral of $\int \exp(L_1)d\phi \exp(L_2)$ with respect to Σ^{-1}.

From the direct calculation we can see that

$$L_1(\phi, \sigma) = -\frac{1}{2} \sum_{j=1}^{m} (\phi_j - \check{\phi}_j)' \sum_{t=T_{j-1}+1}^{T_j} \left(x_{jt} x_{jt}' \otimes \Sigma_j^{-1} \right) (\phi_j - \check{\phi}_j) = \sum_{j=1}^{m+1} \left(-\frac{1}{2} q_j \right), \quad \text{say,}$$

where we used the fact that $\sum_{t=T_{j-1}+1}^{T_j} (x_{jt} \otimes I_n) y_t = \sum_{t=T_{j-1}+1}^{T_j} \left(x_{jt} x_{jt}' \otimes I_n \right) \check{\phi}_j$ from the first order condition on the maximization. Using this expression we can see that

$$
\begin{aligned}
\int e^{L_1(\phi, \sigma)} d\phi &= \prod_{j=1}^{m+1} \int e^{-q_j/2} d\phi_j \\
&= \prod_{j=1}^{m+1} \int (2\pi)^{-p_{x_j}/2} \left| \sum_{t=T_{j-1}+1}^{T_j} \left(x_{jt} x_{jt}' \otimes \Sigma_j^{-1} \right) \right|^{1/2} e^{-q_j/2} d\phi_j \\
&\quad \times (2\pi)^{p_{x_j}/2} \left| \sum_{t=T_{j-1}+1}^{T_j} \left(x_{jt} x_{jt}' \otimes \Sigma_j^{-1} \right) \right|^{-1/2} \\
&= \prod_{j=1}^{m+1} (2\pi)^{p_{x_j}/2} (\Delta T_j)^{-p_{\phi_j}/2} \left| \check{\Sigma}_{j,x} \right|^{-n/2} |\Sigma_j|^{-p_{x_j}/2},
\end{aligned}
\tag{39}
$$

where $\check{\Sigma}_{j,x} = (\Delta T_j)^{-1} \sum_{t=T_{j-1}+1}^{T_j} x_{jt} x_{jt}'$. The last equality holds because the integrand on the right hand side of the second equality is the pdf of a p_{x_j} dimensional normal distribution.

For $L_2(\sigma)$ we can see that

$$L_2(\sigma) = \frac{nT}{2} + \sum_{j=1}^{m+1} \left\{ \frac{\Delta T_j}{2} \log |\check{\Sigma}_j| / |\Sigma_j| - \frac{1}{2} \text{tr} \left(S_j \Sigma_j^{-1} \right) \right\}
\tag{40}$$

where $S_j = \sum_{t=T_{j-1}+1}^{T_j} \check{\varepsilon}_t \check{\varepsilon}_t'$. Using (39) and (40) we have

$$
\begin{aligned}
\int\int e^{L_1(\phi, \sigma)+L_2(\sigma)} d\phi d\Sigma^{-1} &= e^{nT/2} \prod_{j=1}^{m+1} (2\pi)^{p_{x_j}/2} \left| \check{\Sigma}_{j,x} \right|^{-n/2} (\Delta T_j)^{-p_{\phi_j}/2} \left| \check{\Sigma}_j \right|^{\Delta T_j/2} \\
&\quad \times \int \left| \Sigma_j^{-1} \right|^{(\Delta T_j - p_{x_j})/2} e^{-(1/2)\text{tr}(S_j \Sigma_j^{-1})} d\Sigma_j^{-1}.
\end{aligned}
\tag{41}
$$

33

Note that

$$
\int \left|\Sigma_j^{-1}\right|^{(\Delta T_j - p_{x_j})/2} e^{-(1/2)\mathrm{tr}(S_j \Sigma_j^{-1})} d\Sigma_j^{-1}
$$

$$
= \int \frac{\left|\Sigma_j^{-1}\right|^{(\Delta T_j^* - n - 1)/2} e^{-(1/2)\mathrm{tr}(S_j \Sigma_j^{-1})}}{2^{n\Delta T_j^*/2} \pi^{n(n-1)/4} \left|S_j^{-1}\right|^{\Delta T_j^*/2} \prod_{i=1}^{n} \Gamma\left[\frac{1}{2}(\Delta T_j^* + 1 - i)\right]} d\Sigma_j^{-1}
$$

$$
\times 2^{n\Delta T_j^*/2} \pi^{n(n-1)/4} \left|S_j^{-1}\right|^{\Delta T_j^*/2} \prod_{i=1}^{n} \Gamma\left[\frac{1}{2}(\Delta T_j^* + 1 - i)\right]
$$

$$
= 2^{n\Delta T_j^*/2} \pi^{n(n-1)/4} \Delta T_j^{-n\Delta T_j^*/2} \left|\check{\Sigma}_j\right|^{-\Delta T_j^*/2} \prod_{i=1}^{n} \Gamma\left[\frac{1}{2}(\Delta T_j^* + 1 - i)\right],
$$

where $\Delta T_j^* = \Delta T_j - p_{x_j} + n + 1$ and the last equality holds because the integrand on the right hand side of the first equality is the pdf of the Wishart distribution. Then, the logarithm of (41) becomes

$$
\log \int \int e^{L_1(\phi,\sigma) + L_2(\sigma)} d\phi d\Sigma^{-1} = \frac{nT}{2} + \sum_{j=1}^{m+1} \left\{ \frac{n\Delta T_j^*}{2} \log 2 - \frac{n(\Delta T_j^* + p_{x_j})}{2} \log \Delta T_j \right.
$$

$$
\left. + \sum_{i=1}^{n} \log \Gamma\left[\frac{1}{2}(\Delta T_j^* + 1 - i)\right] \right\} + O_p(1).
$$

Using the Stirling's formula, the sum of the logarithms of the Gamma functions becomes

$$
\sum_{i=1}^{n} \log \Gamma\left[\frac{1}{2}(\Delta T_j^* + 1 - i)\right]
$$

$$
= \frac{n}{2} \log(2\pi) + \sum_{i=1}^{n} \left(\frac{\Delta T_j^* - i}{2} \log \frac{\Delta T_j^* + 1 - i}{2} - \frac{\Delta T_j^* + 1 - i}{2} + \frac{\vartheta_i}{6(\Delta T_j^* + 1 - i)} \right)
$$

$$
= \left(\frac{n\Delta T_j^*}{2} - \frac{n(n+1)}{4} \right) (\log \Delta T_j - \log 2) - \frac{nT}{2} + O(1),
$$

where $0 < \vartheta_i < 1$. The second equality holds because $\sum_{i=1}^{n}(\Delta T_j^* - i)/2 = n\Delta T_j^*/2 - n(n+1)/4$, $\sum_{i=1}^{n}(\Delta T_j^* + 1 - i) = nT + O(1)$ and

$$
\log(\Delta T_j^* + 1 - i) = \log \Delta T_j + \log\left(1 + \frac{-p_{x_j} + n + 2 - i}{\Delta T_j}\right) = \log \Delta T_j + O\left(\frac{1}{\Delta T_j}\right).
$$

Thus, we have

$$
\log \int \int e^{L_1(\phi,\sigma) + L_2(\sigma)} d\phi d\Sigma^{-1} = \sum_{j=1}^{m+1} \left\{ -\frac{np_{x_j}}{2} - \frac{n(n+1)}{4} \right\} \log \Delta T_j + O_p(1)
$$

$$
= -\sum_{j=1}^{m+1} \frac{p_{\phi_j} + p_\sigma}{2} \log \Delta T_j + O_p(1). \tag{42}
$$

From (37), (38) and (42) we finally have

$$\log g_M(y|x) = \ell_{m,p_x}(\mathcal{T}, \check{\theta}|y, x) - \sum_{j=1}^{m+1} \frac{p_{\phi_j} + p_\sigma}{2} \log \Delta T_j + O_p(1). \blacksquare$$

Proof of Proposition 4: (i) We will show that $MIC(m, p_x) - MIC(m^0, p_x^0) \to \infty$ if $m \neq m^0$ or $p_x \neq p_x^0$. Let $\hat{\mathcal{T}}^0$ and $\hat{\theta}^0$ be the MLEs when $m = m^0$ and $p_x = p_x^0$. Then,

$$\begin{aligned} MIC(m, p_x) - MIC(m^0, p_x^0) &= -2\left[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) - \ell_{m^0,p_x^0}(\mathcal{T}^0, \theta^0|y, x)\right] \quad (43) \\ &\quad + 2\left[\ell_{m^0,p_x^0}(\hat{\mathcal{T}}^0, \hat{\theta}^0|y, x) - \ell_{m^0,p_x^0}(\mathcal{T}^0, \theta^0|y, x)\right] + o_p(Tv_T^2). \end{aligned}$$

When $m < m^0$ or $p_x < p_x^0$, there exists at least one regime in which the estimators of the coefficients are inconsistent. In this case, the first term on the right hand side of (43) is greater than cTv_T^2 for some $c > 0$ with a large probability as shown in the proof of Theorem 6 of Bai (2000), while the second term is shown to be $O_p(1)$ in the same way as Lemma 13 of Bai (2000). As a result, the left hand side of (43) diverges to infinity as $T \to \infty$.

When $m > m^0$ and $p_x > p_x^0$, we rewrite (43) as

$$\begin{aligned} MIC(m, p_x) - MIC(m^0, p_x^0) &= -2\left[\ell_{m,p_x}(\hat{\mathcal{T}}, \hat{\theta}|y, x) - \ell_{m^0,p_x}(\hat{\mathcal{T}}^*, \hat{\theta}^*|y, x)\right] \quad (44) \\ &\quad -2\left[\ell_{m^0,p_x}(\hat{\mathcal{T}}^*, \hat{\theta}^*|y, x) - \ell_{m^0,p_x^0}(\hat{\mathcal{T}}^0, \hat{\theta}^0|y, x)\right] \\ &\quad + c_1 g_1(T) + (m - m_0)g_2(T), \end{aligned}$$

where $\mathcal{T}^*$ and θ^* are the MLEs of $\mathcal{T}$ and θ when $m = m^0$ and $p_x > p_x^0$, and c_1 is the difference of the number of the unknown parameters, which is positive. Since $p_x > p_x^0$, the model with $m = m_0$ and p_x may be seen as the true model with the zero coefficients associated with the additional regressors. Then, the first term on the right hand side of (44) can be seen as the likelihood ratio test statistic for m, which is $O_p(1)$. On the other hand, the second term is the likelihood ratio test statistic for the extra regressors, which is $O_p(1)$. As a result, the left hand side of (44) goes to infinity when $m > m_0$ and $p_x > p_x^0$ because $c_1 g_1(T) \to \infty$.

In exactly the same manner we can see that the left hand side of (44) diverges to infinity when $m = m^0$ and $p_x > p_x^0$ and when $m > m^0$ and $p_x = p_x^0$ because c_1 is positive in both the cases. Thus, we have (i).

Similarly, in the case of (ii), the left hand side of (44) goes to infinity when $m > m^0$ and $p_x \geq p_x^0$ because $(m - m^0)g_2(T) \to \infty$, while it does not diverge when $m = m^0$ and $p_x \geq p_x^0$ because $c_1 g_1(T) = O(1)$. Hence, we have (ii).

In the case of (iii) we have the same equality as (43) with the $o_p(Tv_T^2)$ term replaced by the $O_p(1)$ term and thus the left hand side of (43) goes to infinity when $m < m^0$ or $p_x < p_x^0$. On the other hand, we can see that (44) does not go to infinity when $m \geq m^0$ and $p_x \geq p_x^0$ because $g_1(T)$ and $g_2(T)$ are $O(1)$. As a result, we have $P(m \geq m^0 \text{ and } p_x \geq p_x^0) \to 1$. ∎

References

[1] Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle, in B. N. Petrov and F. Csáki, ed., *2nd International Symposium on Information Theory, Akadémiai Kiadó*. Budapest, pp. 267–281.

[2] Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61, 821-856.

[3] Andrews, D. W. K., I. Lee and W. Ploberger (1996). Optimal Changepoint Tests for Normal Linear Regression. *Journal of Econometrics* 70, 9-38.

[4] Bai, J. (1997a). Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics* 79, 551-563.

[5] Bai, J. (1997b). Estimating Multiple Breaks One at a Time. *Econometric Theory* 13, 315-352.

[6] Bai, J. (1999). Likelihood Ratio Tests for Multiple Structural Changes. *Journal of Econometrics* 91, 299-323.

[7] Bai, J. (2000). Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices. *Annals of Economics and Finance* 1, 303-339.

[8] Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66, 47-78.

[9] Bai, J. and P. Perron (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics* 18, 1-22.

[10] Bai, J. and P. Perron (2006). Estimating and Testing Linear Models with Multiple Structural Changes, in D. Corbae, S. N. Durlauf and B. E. Hansen ed., *Econometric Theory and Practice.* Cambridge University Press, Cambridge.

[11] Bhattacharya, P. K. and P. J. Brockwell (1976). The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory. *Z. Wahrschein. Verw. Gebiete* 37, 51-75.

[12] Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference, 2nd ed.* Springer, New York.

[13] Hannan, E. J. (1980). The Estimation of the Order of an ARMA Process. *Annals of Statistics* 8, 1071-1081.

[14] Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems.* Wiley, New York.

[15] Hansen, B. E. (2009). Averaging Estimators for Regressions with a Possible Structural Break. *Econometric Theory* 25, 1498-1514.

[16] Konishi, S. and G. Kitagawa (2008). *Information Criteria and Statistical Modeling.* Springer-Verlag, New-York.

[17] Mallows, C. L. (1973). Some Comments on C_p. *Technometrics* 15, 661-675.

[18] Liu, J., S. Wu and J. V. Zidek (1997). On Segmented Multivariate Regressions. *Statistica Sinica* 7, 497-525.

[19] Ninomiya, Y. (2005). Information Criterion for Gaussian Change-Point Model. *Statistics and Probability Letters* 72, 237-247.

[20] Ninomiya, Y. (2006). AIC for Change-Point Models and its Application. Research Memorandum 1002, *Institute of Statistical Mathematics.*

[21] Perron, P. (2006). Dealing with Structural Breaks, in T. C. Mills and K. Patterson, ed., *Palgrave Handbook of Econometrics, Vol.1: Econometric Theory.* Palgrave Macmillan, New York.

[22] Qu, Z. and P. Perron (2007). Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica* 75, 459-502.

[23] Ross, S. M. (1996). *Stochastic Processes, 2nd ed.* Wiley, New York.

[24] Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6, 461-464.

[25] Shao, J. (1997). An Asymptotic Theory for Linear Model Selection. *Statistica Sinica* 7, 221-226.

[26] Shibata, R. (1976). Selection of the Order of an Autoregressive Model by Akaike's Information Criterion. *Biometrika* 63, 117-126.

[27] Yao, Y. C. (1988). Estimating the Number of Change-Points via Schwarz' Criterion. *Statistics and Probability Letters* 6, 181-189.

[28] Zhang, N. R. and D. Siegmund (2007). Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data. *Biometrics* 63, 22-32.

Table 1. The Parameter Setting for Simulations

			ϕ_1		ϕ_2		θ_1	
No Break								
DGP0AR1a			0.3		0.0		0.0	
DGP0AR1b			0.7		0.0		0.0	
DGP0AR2a			0.2		0.1		0.0	
DGP0AR2b			0.4		0.3		0.0	
DGP0AR2c			0.7		-0.4		0.0	
DGP0AR2d			1.3		-0.6		0.0	
DGP0MA1a			0.0		0.0		0.3	
DGP0MA1b			0.0		0.0		0.7	

One Time Break

	c_1	c_2	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	θ_{11}	θ_{12}
DGP1AR1a	1.0	0.0	0.7	0.3	0.0	0.0	0.0	0.0
DGP1AR1b	0.0	1.0	0.3	0.7	0.0	0.0	0.0	0.0
DGP1AR2a	1.0	0.0	0.4	0.2	0.3	0.1	0.0	0.0
DGP1AR2b	0.0	1.0	0.2	0.4	0.1	0.3	0.0	0.0
DGP1AR2c	1.0	0.0	1.3	0.7	-0.6	-0.4	0.0	0.0
DGP1AR2d	0.0	1.0	0.7	1.3	-0.4	-0.6	0.0	0.0
DGP1MA1a	1.0	0.0	0.0	0.0	0.0	0.0	0.7	0.3
DGP1MA1b	0.0	1.0	0.0	0.0	0.0	0.0	0.3	0.7

Two Time Breaks

	c_1	c_2	c_3	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{21}	ϕ_{22}	ϕ_{23}	θ_{11}	θ_{12}	θ_{13}
DGP2AR1a	1.0	0.0	1.0	0.7	0.3	0.7	0.0	0.0	0.0	0.0	0.0	0.0
DGP2AR1b	1.0	0.0	-1.0	0.7	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
DGP2AR2a	1.0	0.0	1.0	0.4	0.2	0.4	0.3	0.1	0.3	0.0	0.0	0.0
DGP2AR2b	1.0	0.0	-1.0	0.4	0.3	0.2	0.3	0.2	0.1	0.0	0.0	0.0
DGP2AR2c	1.0	0.0	1.0	1.3	0.7	1.3	-0.6	-0.4	-0.6	0.0	0.0	0.0
DGP2AR2d	1.0	0.0	-1.0	1.3	1.0	0.7	-0.6	-0.5	-0.4	0.0	0.0	0.0
DGP2MA1a	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.3	0.7
DGP2MA1b	1.0	0.0	-1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.3	0.1

Table 2a. The frequencies of Selecting the True Model (No break)

T	120	300	120	120	T	120	300	120	120
ϵ	0.05	0.05	0.15	0.15	ϵ	0.05	0.05	0.15	0.15
	\multicolumn{4}{c}{DGP0AR1a}		\multicolumn{4}{c}{DGP0AR1b}			
AIC	0.000	0.000	0.047	0.066	AIC	0.000	0.000	0.024	0.040
BIC	0.156	0.794	0.529	0.908	BIC	0.540	0.827	0.756	0.907
C_p	0.000	0.000	0.052	0.070	C_p	0.000	0.000	0.027	0.042
$MAIC$	0.517	0.559	0.608	0.652	$MAIC$	0.438	0.473	0.583	0.613
$MBIC_y$	0.724	0.964	0.768	0.968	$MBIC_y$	0.916	0.970	0.933	0.976
$MBIC_1$	0.780	0.970	0.795	0.973	$MBIC_1$	0.923	0.973	0.942	0.979
$MBIC_2$	0.801	0.973	0.804	0.974	$MBIC_2$	0.954	0.979	0.955	0.980
MC_p	0.538	0.683	0.662	0.730	MC_p	0.512	0.660	0.669	0.717
$Sq(MAIC)$	0.336	0.574	0.609	0.680	$Sq(MAIC)$	0.223	0.478	0.580	0.644
$Sq(MBIC_y)$	0.427	0.810	0.722	0.912	$Sq(MBIC_y)$	0.346	0.696	0.788	0.891
$Sq(MBIC_1)$	0.445	0.811	0.735	0.912	$Sq(MBIC_1)$	0.345	0.696	0.787	0.891
$Sq(MBIC_2)$	0.448	0.811	0.737	0.912	$Sq(MBIC_2)$	0.346	0.696	0.788	0.891
$Sq(MC_p)$	0.313	0.580	0.614	0.693	$Sq(MC_p)$	0.197	0.488	0.570	0.663
$Sq(rb)$	0.228	0.556	0.631	0.774	$Sq(rb)$	0.083	0.308	0.543	0.719
	\multicolumn{4}{c}{DGP0AR2a}		\multicolumn{4}{c}{DGP0AR2b}			
AIC	0.000	0.000	0.008	0.037	AIC	0.000	0.000	0.024	0.078
BIC	0.001	0.073	0.020	0.151	BIC	0.062	0.893	0.460	0.945
C_p	0.000	0.000	0.008	0.039	C_p	0.000	0.000	0.029	0.082
$MAIC$	0.108	0.286	0.159	0.362	$MAIC$	0.347	0.479	0.539	0.639
$MBIC_y$	0.039	0.183	0.053	0.188	$MBIC_y$	0.584	0.966	0.700	0.969
$MBIC_1$	0.054	0.191	0.061	0.193	$MBIC_1$	0.688	0.966	0.730	0.970
$MBIC_2$	0.062	0.193	0.064	0.194	$MBIC_2$	0.747	0.970	0.754	0.970
MC_p	0.197	0.412	0.213	0.432	MC_p	0.504	0.696	0.662	0.753
$Sq(MAIC)$	0.043	0.275	0.160	0.377	$Sq(MAIC)$	0.127	0.462	0.524	0.684
$Sq(MBIC_y)$	0.011	0.143	0.049	0.178	$Sq(MBIC_y)$	0.162	0.618	0.587	0.876
$Sq(MBIC_1)$	0.015	0.147	0.054	0.181	$Sq(MBIC_1)$	0.181	0.619	0.599	0.876
$Sq(MBIC_2)$	0.017	0.147	0.056	0.181	$Sq(MBIC_2)$	0.185	0.619	0.603	0.876
$Sq(MC_p)$	0.063	0.334	0.189	0.408	$Sq(MC_p)$	0.147	0.474	0.557	0.694
$Sq(rb)$	0.011	0.119	0.069	0.243	$Sq(rb)$	0.009	0.041	0.186	0.346

Table 2a. (Continued)

T	120	300	120	120	T	120	300	120	120
ϵ	0.05	0.05	0.15	0.15	ϵ	0.05	0.05	0.15	0.15
	DGP0AR2c					DGP0AR2d			
AIC	0.003	0.005	0.136	0.136	AIC	0.000	0.001	0.098	0.117
BIC	0.888	0.960	0.911	0.967	BIC	0.877	0.957	0.915	0.966
C_p	0.002	0.004	0.159	0.143	C_p	0.000	0.001	0.110	0.124
$MAIC$	0.594	0.595	0.683	0.680	$MAIC$	0.523	0.572	0.650	0.673
$MBIC_y$	0.945	0.978	0.947	0.980	$MBIC_y$	0.955	0.977	0.959	0.978
$MBIC_1$	0.936	0.978	0.946	0.980	$MBIC_1$	0.940	0.976	0.957	0.979
$MBIC_2$	0.953	0.981	0.954	0.981	$MBIC_2$	0.965	0.980	0.966	0.980
MC_p	0.586	0.713	0.732	0.756	MC_p	0.550	0.703	0.715	0.750
$Sq(MAIC)$	0.227	0.595	0.683	0.711	$Sq(MAIC)$	0.124	0.528	0.645	0.703
$Sq(MBIC_y)$	0.309	0.780	0.856	0.917	$Sq(MBIC_y)$	0.175	0.701	0.829	0.910
$Sq(MBIC_1)$	0.309	0.780	0.856	0.917	$Sq(MBIC_1)$	0.175	0.701	0.828	0.910
$Sq(MBIC_2)$	0.309	0.780	0.856	0.917	$Sq(MBIC_2)$	0.175	0.701	0.829	0.910
$Sq(MC_p)$	0.196	0.588	0.672	0.719	$Sq(MC_p)$	0.109	0.524	0.634	0.708
$Sq(rb)$	0.627	0.870	0.882	0.898	$Sq(rb)$	0.451	0.861	0.886	0.898
	DGP0MA1a					DGP0MA2b			
AIC	0.009	0.010	0.179	0.176	AIC	0.071	0.194	0.427	0.478
BIC	0.902	0.967	0.942	0.980	BIC	0.983	0.997	0.988	0.998
C_p	0.009	0.010	0.210	0.188	C_p	0.074	0.184	0.502	0.503
$MAIC$	0.822	0.817	0.892	0.894	$MAIC$	0.886	0.900	0.936	0.934
$MBIC_y$	0.994	0.998	0.995	0.999	$MBIC_y$	0.998	1.000	0.999	1.000
$MBIC_1$	0.992	0.999	0.997	1.000	$MBIC_1$	0.992	0.999	0.998	1.000
$MBIC_2$	1.000	1.000	1.000	1.000	$MBIC_2$	1.000	1.000	1.000	1.000
MC_p	0.919	0.970	0.982	0.987	MC_p	0.955	0.984	0.991	0.993
$Sq(MAIC)$	0.500	0.820	0.902	0.926	$Sq(MAIC)$	0.173	0.834	0.932	0.961
$Sq(MBIC_y)$	0.741	0.920	0.964	0.969	$Sq(MBIC_y)$	0.389	0.871	0.957	0.971
$Sq(MBIC_1)$	0.741	0.920	0.964	0.969	$Sq(MBIC_1)$	0.388	0.871	0.956	0.971
$Sq(MBIC_2)$	0.742	0.920	0.965	0.969	$Sq(MBIC_2)$	0.389	0.871	0.957	0.971
$Sq(MC_p)$	0.429	0.845	0.925	0.952	$Sq(MC_p)$	0.125	0.835	0.938	0.965
$Sq(rb)$	0.664	0.885	0.940	0.970	$Sq(rb)$	0.958	0.997	0.999	1.000

Table 2b. The frequencies of Selecting the True Model (One Time break)

T	120	300	120	120	T	120	300	120	120
ϵ	0.05	0.05	0.15	0.15	ϵ	0.05	0.05	0.15	0.15
	DGP1AR1a					DGP1AR1b			
AIC	0.000	0.000	0.082	0.099	AIC	0.000	0.000	0.062	0.092
BIC	0.263	0.804	0.672	0.924	BIC	0.195	0.782	0.612	0.911
C_p	0.000	0.000	0.097	0.105	C_p	0.000	0.000	0.080	0.099
$MAIC$	0.362	0.475	0.514	0.656	$MAIC$	0.303	0.443	0.457	0.639
$MBIC_y$	0.509	0.935	0.550	0.946	$MBIC_y$	0.425	0.924	0.479	0.936
$MBIC_1$	0.439	0.892	0.450	0.903	$MBIC_1$	0.359	0.869	0.377	0.879
$MBIC_2$	0.306	0.840	0.305	0.841	$MBIC_2$	0.242	0.802	0.243	0.803
MC_p	0.282	0.645	0.333	0.713	MC_p	0.247	0.626	0.284	0.697
$Sq(MAIC)$	0.375	0.586	0.504	0.698	$Sq(MAIC)$	0.367	0.580	0.460	0.687
$Sq(MBIC_y)$	0.556	0.825	0.678	0.923	$Sq(MBIC_y)$	0.534	0.826	0.633	0.915
$Sq(MBIC_1)$	0.571	0.819	0.673	0.918	$Sq(MBIC_1)$	0.561	0.819	0.626	0.908
$Sq(MBIC_2)$	0.577	0.819	0.673	0.918	$Sq(MBIC_2)$	0.566	0.819	0.623	0.906
$Sq(MC_p)$	0.280	0.568	0.435	0.685	$Sq(MC_p)$	0.270	0.559	0.394	0.673
$Sq(rb)$	0.437	0.568	0.409	0.616	$Sq(rb)$	0.437	0.567	0.394	0.599
	DGP1AR2a					DGP1AR2b			
AIC	0.000	0.000	0.032	0.131	AIC	0.000	0.000	0.022	0.116
BIC	0.001	0.294	0.034	0.500	BIC	0.000	0.240	0.021	0.438
C_p	0.000	0.000	0.040	0.143	C_p	0.000	0.000	0.028	0.125
$MAIC$	0.102	0.416	0.214	0.634	$MAIC$	0.064	0.358	0.161	0.586
$MBIC_y$	0.012	0.395	0.031	0.422	$MBIC_y$	0.005	0.325	0.012	0.350
$MBIC_1$	0.035	0.413	0.046	0.428	$MBIC_1$	0.016	0.334	0.020	0.343
$MBIC_2$	0.012	0.255	0.015	0.258	$MBIC_2$	0.004	0.185	0.005	0.187
MC_p	0.162	0.598	0.148	0.653	MC_p	0.114	0.557	0.099	0.600
$Sq(MAIC)$	0.128	0.498	0.201	0.665	$Sq(MAIC)$	0.112	0.457	0.155	0.621
$Sq(MBIC_y)$	0.094	0.480	0.105	0.578	$Sq(MBIC_y)$	0.090	0.451	0.080	0.544
$Sq(MBIC_1)$	0.170	0.555	0.163	0.649	$Sq(MBIC_1)$	0.172	0.527	0.133	0.612
$Sq(MBIC_2)$	0.227	0.546	0.216	0.635	$Sq(MBIC_2)$	0.235	0.528	0.187	0.608
$Sq(MC_p)$	0.191	0.537	0.229	0.657	$Sq(MC_p)$	0.196	0.503	0.184	0.609
$Sq(rb)$	0.097	0.375	0.151	0.542	$Sq(rb)$	0.094	0.370	0.133	0.544

Table 2b. (Cotinued)

T	120	300	120	120	T	120	300	120	120
ϵ	0.05	0.05	0.15	0.15	ϵ	0.05	0.05	0.15	0.15
	DGP1AR2c					DGP1AR2d			
AIC	0.000	0.003	0.142	0.171	AIC	0.000	0.002	0.142	0.179
BIC	0.786	0.956	0.857	0.979	BIC	0.750	0.955	0.833	0.978
C_p	0.001	0.004	0.178	0.184	C_p	0.000	0.002	0.179	0.192
$MAIC$	0.457	0.550	0.636	0.696	$MAIC$	0.424	0.536	0.618	0.685
$MBIC_y$	0.741	0.991	0.748	0.993	$MBIC_y$	0.691	0.990	0.696	0.992
$MBIC_1$	0.676	0.983	0.711	0.987	$MBIC_1$	0.625	0.980	0.659	0.987
$MBIC_2$	0.520	0.984	0.520	0.984	$MBIC_2$	0.463	0.983	0.462	0.983
MC_p	0.437	0.710	0.597	0.776	MC_p	0.420	0.710	0.575	0.773
$Sq(MAIC)$	0.419	0.647	0.656	0.743	$Sq(MAIC)$	0.407	0.649	0.645	0.739
$Sq(MBIC_y)$	0.595	0.841	0.841	0.934	$Sq(MBIC_y)$	0.584	0.845	0.820	0.933
$Sq(MBIC_1)$	0.585	0.838	0.832	0.931	$Sq(MBIC_1)$	0.578	0.842	0.813	0.930
$Sq(MBIC_2)$	0.588	0.840	0.831	0.934	$Sq(MBIC_2)$	0.576	0.845	0.810	0.933
$Sq(MC_p)$	0.313	0.625	0.610	0.740	$Sq(MC_p)$	0.296	0.626	0.597	0.735
$Sq(rb)$	0.381	0.112	0.045	0.071	$Sq(rb)$	0.392	0.104	0.039	0.062
	DGP1MA1a					DGP1MA1b			
AIC	0.020	0.063	0.341	0.405	AIC	0.018	0.062	0.351	0.398
BIC	0.939	0.983	0.972	0.992	BIC	0.939	0.980	0.969	0.991
C_p	0.030	0.073	0.422	0.437	C_p	0.032	0.070	0.421	0.431
$MAIC$	0.782	0.805	0.904	0.907	$MAIC$	0.789	0.805	0.899	0.903
$MBIC_y$	0.995	0.999	0.997	0.999	$MBIC_y$	0.995	0.998	0.997	0.999
$MBIC_1$	0.980	0.998	0.995	0.999	$MBIC_1$	0.982	0.996	0.995	0.999
$MBIC_2$	1.000	1.000	1.000	1.000	$MBIC_2$	0.999	1.000	0.999	1.000
MC_p	0.933	0.970	0.990	0.991	MC_p	0.937	0.971	0.990	0.992
$Sq(MAIC)$	0.684	0.854	0.932	0.947	$Sq(MAIC)$	0.678	0.862	0.932	0.944
$Sq(MBIC_y)$	0.868	0.922	0.968	0.966	$Sq(MBIC_y)$	0.866	0.920	0.967	0.966
$Sq(MBIC_1)$	0.842	0.915	0.964	0.965	$Sq(MBIC_1)$	0.841	0.915	0.962	0.964
$Sq(MBIC_2)$	0.869	0.922	0.969	0.966	$Sq(MBIC_2)$	0.866	0.920	0.967	0.966
$Sq(MC_p)$	0.622	0.863	0.936	0.954	$Sq(MC_p)$	0.619	0.873	0.940	0.955
$Sq(rb)$	0.906	0.962	0.975	0.984	$Sq(rb)$	0.903	0.959	0.975	0.987

Table 2c. The frequencies of Selecting the True Model (Two Time breaks)

T	120	300	120	120	T	120	300	120	120
ϵ	0.05	0.05	0.15	0.15	ϵ	0.05	0.05	0.15	0.15
	DGP2AR1a					DGP2AR1b			
AIC	0.000	0.000	0.173	0.210	AIC	0.000	0.000	0.211	0.238
BIC	0.132	0.613	0.332	0.781	BIC	0.056	0.719	0.349	0.892
C_p	0.000	0.000	0.201	0.219	C_p	0.000	0.000	0.240	0.254
$MAIC$	0.154	0.376	0.216	0.633	$MAIC$	0.214	0.479	0.315	0.703
$MBIC_y$	0.056	0.326	0.062	0.330	$MBIC_y$	0.080	0.689	0.126	0.696
$MBIC_1$	0.036	0.176	0.030	0.176	$MBIC_1$	0.103	0.581	0.102	0.580
$MBIC_2$	0.005	0.055	0.005	0.054	$MBIC_2$	0.027	0.385	0.029	0.381
MC_p	0.047	0.324	0.027	0.319	MC_p	0.091	0.539	0.072	0.576
$Sq(MAIC)$	0.132	0.367	0.152	0.562	$Sq(MAIC)$	0.148	0.558	0.239	0.701
$Sq(MBIC_y)$	0.179	0.483	0.195	0.693	$Sq(MBIC_y)$	0.152	0.736	0.264	0.882
$Sq(MBIC_1)$	0.190	0.470	0.193	0.677	$Sq(MBIC_1)$	0.212	0.732	0.289	0.875
$Sq(MBIC_2)$	0.187	0.459	0.189	0.658	$Sq(MBIC_2)$	0.211	0.732	0.275	0.876
$Sq(MC_p)$	0.091	0.299	0.100	0.452	$Sq(MC_p)$	0.111	0.507	0.189	0.654
$Sq(rb)$	0.140	0.226	0.141	0.286	$Sq(rb)$	0.157	0.387	0.229	0.508
	DGP2AR2a					DGP2AR2b			
AIC	0.000	0.000	0.059	0.224	AIC	0.000	0.000	0.047	0.219
BIC	0.000	0.043	0.002	0.078	BIC	0.000	0.016	0.002	0.034
C_p	0.000	0.000	0.074	0.247	C_p	0.000	0.001	0.056	0.237
$MAIC$	0.032	0.254	0.064	0.457	$MAIC$	0.042	0.216	0.052	0.339
$MBIC_y$	0.000	0.005	0.001	0.007	$MBIC_y$	0.000	0.004	0.001	0.005
$MBIC_1$	0.002	0.008	0.001	0.008	$MBIC_1$	0.005	0.011	0.002	0.009
$MBIC_2$	0.000	0.001	0.000	0.001	$MBIC_2$	0.000	0.001	0.000	0.001
MC_p	0.026	0.159	0.011	0.130	MC_p	0.029	0.132	0.008	0.101
$Sq(MAIC)$	0.047	0.166	0.041	0.305	$Sq(MAIC)$	0.039	0.155	0.036	0.238
$Sq(MBIC_y)$	0.052	0.202	0.036	0.325	$Sq(MBIC_y)$	0.022	0.110	0.015	0.152
$Sq(MBIC_1)$	0.084	0.233	0.052	0.361	$Sq(MBIC_1)$	0.047	0.161	0.030	0.211
$Sq(MBIC_2)$	0.105	0.245	0.075	0.375	$Sq(MBIC_2)$	0.075	0.168	0.048	0.211
$Sq(MC_p)$	0.075	0.156	0.054	0.243	$Sq(MC_p)$	0.058	0.174	0.041	0.235
$Sq(rb)$	0.036	0.225	0.066	0.452	$Sq(rb)$	0.032	0.216	0.052	0.408

T	120	300	120	120	T	120	300	120	120
ϵ	0.05	0.05	0.15	0.15	ϵ	0.05	0.05	0.15	0.15
		DGP2AR2c					DGP2AR2d		
AIC	0.000	0.003	0.246	0.300	AIC	0.000	0.003	0.253	0.302
BIC	0.311	0.872	0.357	0.911	BIC	0.161	0.469	0.131	0.473
C_p	0.000	0.004	0.307	0.321	C_p	0.001	0.005	0.306	0.327
$MAIC$	0.243	0.510	0.409	0.749	$MAIC$	0.170	0.432	0.192	0.603
$MBIC_y$	0.085	0.624	0.083	0.627	$MBIC_y$	0.034	0.166	0.027	0.161
$MBIC_1$	0.087	0.564	0.089	0.571	$MBIC_1$	0.062	0.162	0.034	0.155
$MBIC_2$	0.011	0.247	0.010	0.247	$MBIC_2$	0.005	0.036	0.003	0.035
MC_p	0.156	0.667	0.135	0.722	MC_p	0.060	0.347	0.029	0.325
$Sq(MAIC)$	0.162	0.587	0.284	0.761	$Sq(MAIC)$	0.146	0.388	0.121	0.508
$Sq(MBIC_y)$	0.235	0.727	0.362	0.908	$Sq(MBIC_y)$	0.198	0.490	0.153	0.621
$Sq(MBIC_1)$	0.233	0.724	0.359	0.905	$Sq(MBIC_1)$	0.195	0.486	0.150	0.616
$Sq(MBIC_2)$	0.232	0.716	0.354	0.896	$Sq(MBIC_2)$	0.190	0.487	0.145	0.618
$Sq(MC_p)$	0.123	0.546	0.238	0.727	$Sq(MC_p)$	0.101	0.356	0.100	0.472
$Sq(rb)$	0.028	0.001	0.002	0.001	$Sq(rb)$	0.024	0.021	0.003	0.031
		DGP2MA1a					DGP2MA1b		
AIC	0.018	0.067	0.520	0.581	AIC	0.007	0.028	0.445	0.497
BIC	0.950	0.983	0.989	0.996	BIC	0.855	0.957	0.963	0.990
C_p	0.051	0.090	0.607	0.626	C_p	0.024	0.040	0.527	0.532
$MAIC$	0.754	0.800	0.936	0.951	$MAIC$	0.717	0.750	0.906	0.937
$MBIC_y$	0.942	0.998	0.946	1.000	$MBIC_y$	0.885	0.997	0.892	0.998
$MBIC_1$	0.873	0.995	0.900	0.999	$MBIC_1$	0.795	0.994	0.815	0.998
$MBIC_2$	0.598	1.000	0.598	1.000	$MBIC_2$	0.586	0.995	0.586	0.995
MC_p	0.939	0.975	0.955	0.995	MC_p	0.799	0.966	0.767	0.992
$Sq(MAIC)$	0.623	0.853	0.965	0.977	$Sq(MAIC)$	0.609	0.836	0.905	0.972
$Sq(MBIC_y)$	0.870	0.931	0.978	0.986	$Sq(MBIC_y)$	0.813	0.928	0.937	0.984
$Sq(MBIC_1)$	0.815	0.917	0.972	0.983	$Sq(MBIC_1)$	0.774	0.912	0.927	0.982
$Sq(MBIC_2)$	0.853	0.932	0.970	0.986	$Sq(MBIC_2)$	0.793	0.928	0.921	0.984
$Sq(MC_p)$	0.532	0.856	0.956	0.980	$Sq(MC_p)$	0.491	0.846	0.849	0.978
$Sq(rb)$	0.920	0.967	0.991	0.995	$Sq(rb)$	0.813	0.911	0.961	0.980