Linear model selection in the presence of outliers and break points

Wolfgang Polasek and Lei Ren

Institute of Statistics and Econometrics
University of Basel
Holbeinstrasse 12, 4051 Basel, Switzerland
Email: Wolfgang@iso.iso.unibas.ch
November 10, 1999

Abstract

The presence of outliers and break points are important questions in any applications of econometric time series analysis. This paper shows how Bayes tests for different type and complexity can be constructed using the concept of marginal likelihood. The results can be viewed as an extension of Polasek and Ren (1997). If the location of the break point is not known, the lag length, and possible heteroskedastic errors are present, one can calculate the Bayes factors in a normal-gamma regression model quite easily. The search for outliers and break points can be extended to the multivariate normal-Wishart regression model. The approach is demonstrated with Swiss macroeconomic time series GNP and consumption.

`Keywords`: Bayes tests, marginal likelihood, outliers, break points, heteroskedasticity, variable selection, autoregressive processes.

1 Introduction

The selection of models for practical applications is despite many discussions and many new developments an unsolved problem. Automatic model selection from a Bayesian perspective is very time consuming as well since the calculations of Bayes factors always requires is the concept of informative criteria as e.g. AIC and BIC (Aitkin 1973). Recently, the advance of the automatic model selection by intrinsic Bayes factors (Berger and Perrichi 1996) and fractional Bayes factors (O'Hagen 1995) have become a model in Bayesian statistics. Therefore we will make an attempt in this paper to show how the marginal likelihood differ by assuming a full conjugate prior and a a fractional approach.
Time series models react very sensitive on the presence of outliers and break points. Therefore the detection of "change" in a time series is a topic with a long history from a classical point of view (see e.g. Hackl 1989) and in recent time also from a Bayesian perspective. Modeling outliers have become easy using the Gibbs sampler (Verdinelli and Wasserman (1990)), but also structural breaks can be analysed from different perspective as in Polasek and Ren (1997). This paper is concerned with the question if the two problems, outliers and structural breaks, can be analysed in a combined way using the concept of marginal likelihood.
For simplicity reasons we put the time series data for the autoregressive processes into linear regression framework. Assuming a conjugate normal-Wishart distribution we can use the closed integration formula of Polasek and Ren (1997) to calculate the marginal likelihoods. In their discussion, Barbieri and Conigliani (in Polasek 1997) showed how the marginal likelihood (and the fractional Bayes factors) can be calculated if the exact likelihood function for AR process is used.
Therefore we analyse in section 2 the problem of univariate outliers and break points and in section 3 the problem of multivariate outliers and break points. In section 4 we demonstrate the approach by economic examples. Section 5 concludes with the discussions.

2 Marginal likelihood for outliers

2.1 Univariate outlier models

In this section we consider the univariate regression model with outliers. Let y be a $T \times 1$ vector of the dependent variable and X a $T \times k$ matrix of full rank of independent variables. In case of an AR(p) process the first column of X is a vector of ones and the other columns are p lags such that $k = p + 1$. We consider T outlier models, indexed by $j = 1, \dots, T$, i.e.

$$y = X_j \beta_j + D_j \theta + \varepsilon, \quad j = 1, \dots, T, \tag{1}$$

where D_j is a dummy variable, i.e. the j-th unity vector of length T.
The model can be written in compact form as

$$y = \tilde{X}_j \tilde{\beta}_j + \varepsilon \tag{2}$$

with $\tilde{X}_j = (X : D_j)$ and $\tilde{\beta}_j = (\beta_j : \theta)$. If no confusion is possible, we will drop the index j. The OLS estimate of model (2) is given by

$$\hat{\tilde{\beta}}_j = (\tilde{X}_j' \tilde{X}_j)^{-1} \tilde{X}_j' y. \tag{3}$$

To carry out a Bayes test we need an informative prior distribution. We suggest the following simple procedure which is based on the idea of tightness restriction as in Litterman (1986) for the outlier model

$$\tilde{\beta} | \sigma^2 \sim N \left[\tilde{b}_* = \begin{pmatrix} b_* \\ 0 \end{pmatrix}, \tilde{H}_* = \begin{pmatrix} H_* & 0 \\ 0 & s_*^2 \end{pmatrix} \right], \tag{4}$$

where b_* can be set to zero or a unity vector if we believe in a random walk prior distribution. For $\tilde{H}_*$ we assume the tightness structure

$$\tilde{H}_* = diag(h^{-1}, 1, \dots, p, s_*^2),$$

where h is a small number to ensure a non-informative prior for the intercept. For the variance of the outlier we assume simply that it is about the same size as the data dispersion: $s_*^2 = var(y)$. Thus, we assume for the prior precision $\sigma^2 \sim \Gamma(s_*^{-2}/2, n_*/2)$ a gamma distribution, where $n_* = 1$ is the minimum number of prior degree of freedoms.

In the next theorem we will derive the marginal likelihood for the informative model. As special cases we list subsequently the results for the non-informative and the fractional marginal likelihood cases

Theorem 2.1 The marginal likelihood for informative priors
Consider the regression model with outliers as in (1) with informative (conjugate) prior distribution (see also Polasek 1997 or Poirier 1996)

$$(\tilde{\beta}, \sigma^2) \sim N\Gamma(\tilde{\beta}_*, \tilde{H}_*, s_*^2, n_*).$$

Then the marginal likelihood for the event "outlier at point j" is given by

$$f(y|j) = \pi^{-\frac{T}{2}} \frac{|\tilde{H}_{**}|^{1/2}}{|\tilde{H}_*|^{1/2}} \cdot \frac{\Gamma(n_{**}/2)}{\Gamma(n_*/2)} \cdot \frac{(n_* s_*^2)^{\frac{n_*}{2}}}{(n_{**} s_{**}^2)^{\frac{n_{**}}{2}}}, \tag{5}$$

where

$$n_{**} = n_* + T,$$
$$\tilde{H}_{**}^{-1} = \tilde{X}'\tilde{X} + \tilde{H}_*^{-1},$$
$$n_{**} s_{**}^2 = n_* s_*^2 + ESS + (\tilde{\beta}_* - \hat{\beta})'((\tilde{X}'\tilde{X})^{-1} + \tilde{H}_*)^{-1}(\tilde{\beta}_* - \hat{\beta}),$$
$$ESS = (y - \tilde{X}\hat{\tilde{\beta}})'(y - \tilde{X}\hat{\tilde{\beta}}),$$

and $\hat{\beta}_j$ is the OLS estimate given in (3).
Proof:
See Polasek and Ren (1997).

Lemma 2.1.a Regression models and outliers with non-informative prior
The marginal likelihood for the model (1) is given by

$$f(y|j) = |X'X|^{-\frac{1}{2}} h_j^{-\frac{1}{2}} ((ESS - \frac{e_j^2}{h_j})\pi)^{-\frac{T-p-2}{2}} \Gamma\left(\frac{T-p-2}{2}\right)/2 \tag{6}$$

with

$$h_j = 1 - x_j'(X'X)^{-1}x_j, \quad x_j = X'D_j : (k \times 1), \tag{7}$$
$$e_j = y_j - x_j'\tilde{b}, \quad \tilde{b} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y, \quad \tilde{X} = (X : D_j),$$
$$ESS_j = (y - \tilde{X}\tilde{b})'(y - \tilde{X}\tilde{b}), \quad j = 1, \ldots, T,$$

$$ESS = (y - X\tilde{b})'(y - X\tilde{b}),$$

and the OLS estimate is

$$b = (X'X)^{-1}X'y.$$

Proof:
Insert the stacked model (2) into the marginal likelihood of the linear regression model given in Polasek and Ren (1997), and note that

$$|\tilde{X}'\tilde{X}| = |X'X||1 - x_j'(X'X)^{-1}x_j| = |X'X|h_j \tag{8}$$

with h_j given as in (7) and

$$ESS_j = ESS - \frac{e_j^2}{h_j}, \quad j = 1, \dots, T. \tag{9}$$

Lemma 2.1.b The fractional marginal likelihood for AR(p) model with outliers
Consider the univariate outlier model as in (1), then the fractional marginal likelihood is given by

$$f_b^1(y|j) = b^{\frac{Tb-k}{2}}(\pi ESS_j)^{\frac{T(b-1)}{2}}\Gamma\left(\frac{T-k}{2}\right)/\Gamma\left(\frac{Tb-k}{2}\right). \tag{10}$$

Proof:
O'Hagan (1995) considers for the regression model the non-informative class of priors $p(\beta, \sigma^2) \propto \sigma^{2t}$ where t can be any integer. In this paper we will use $t = 1$ for all fractional priors. For details see also Polasek and Ren (1997).

2.2 The homoskedastic AR(p) break point models with outlier

Consider the homoskedastic break point model with outlier

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_m\beta_1 \\ X^m\beta_2 \end{pmatrix} + D_j\delta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}, \quad j = 1, \dots, T, \tag{11}$$

where D_j is a $T \times 1$ vector with a 1 in the $j-th$ position.

Theorem 2.2 The marginal likelihood for informative priors

For the homoskedastic break point regression model we assume the following conjugate prior distribution

$$\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{pmatrix} | \sigma^2 \sim N \left[\begin{pmatrix} \tilde{b_{1*}} \\ \tilde{b_{2*}} \end{pmatrix}, \sigma^2 \tilde{H}_* \right],$$

and

$$\sigma^{-2} \sim \Gamma(s_*^2, n_*).$$

Then the marginal likelihood is given by

$$\begin{aligned} f(y) &= (2\pi)^{-\frac{T}{2}} \frac{(n_* s_*^2)^{n_*/2}}{|\tilde{H}_{*j}|^{\frac{1}{2}} \Gamma(n_*/2)} \\ &\cdot \sum_{m=k+1}^{T-k} p_m |\tilde{H}_{mj}|^{\frac{1}{2}} |\tilde{H}_j^m|^{\frac{1}{2}} (n_{**} s_{**}^2)^{-\frac{n_{**}}{2}} \cdot \Gamma(n_{**}/2) \end{aligned} \qquad (12)$$

with parameters

$$n_{**} = n_* + T,$$

$$\tilde{H}_{mj}^{-1} = \tilde{H}_{*j}^{-1} + \tilde{X}_{mj}' \tilde{X}_{mj},$$

$$(\tilde{H}_j^m)^{-1} = \tilde{H}_{*j}^{-1} + \tilde{X}_j^{m\prime} \tilde{X}_j^m,$$

$$\begin{aligned} n_{**} s_{**}^2 &= n_* s_*^2 + E\tilde{S}S_{mj} + E\tilde{S}S_j^m + (\tilde{b}_{1*j} - \tilde{\hat{\beta}}_{1j})'((\tilde{X}_{mj}' \tilde{X}_{mj})^{-1} + \tilde{H}_{*j})^{-1} (\tilde{b}_{1*j} - \tilde{\hat{\beta}}_{1j}) + \\ &+ (\tilde{b}_{2*j} - \tilde{\hat{\beta}}_{2j})'((\tilde{X}_j^{m\prime} \tilde{X}_j^m)^{-1} + \tilde{H}_{*j})^{-1} (\tilde{b}_{2*j} - \tilde{\hat{\beta}}_{2j}), \end{aligned}$$

and

$$\tilde{\hat{\beta}}_{1j} = (\tilde{X}_{mj}' \tilde{X}_{mj})^{-1} \tilde{X}_{mj}' y_m, \quad \tilde{\hat{\beta}}_{2j} = (\tilde{X}_j^{m\prime} \tilde{X}_j^m)^{-1} \tilde{X}_j^{m\prime} y^m. \qquad (13)$$

Lemma 2.2.a The marginal likelihood with non-informative prior
The marginal likelihood for the homoskedastic model with non-informative prior

$$\begin{aligned} f(y|j) &= \sum_{m=k+1}^{T-k} p_m |X_m' X_m|^{-\frac{1}{2}} |X^{m\prime} X^m|^{-\frac{1}{2}} h_{jm}^{-\frac{1}{2}} \\ &\Gamma(\frac{T}{2} - k) \pi^{-\frac{T}{2}+k} (ESS_m + ESS^m - \frac{e_j^2}{h_{jm}})^{-\frac{T}{2}+k}, \end{aligned} \qquad (14)$$

where the leverage points are given by

$$h_{jm} = \begin{cases} 1 - x_j'(X_m{}'X_m)^{-1}x_j & if \quad j \leq m, \\ 1 - x_j'(X^m{}'X^m)^{-1}x_j & if \quad j > m, \end{cases} \tag{15}$$

and $x_j' = X'D_j$ is the j-th vector of the regression matrix and the residuals are

$$e_j = \begin{cases} y_j - x_j'\tilde{\beta}_1 & if \quad j \leq m, \\ y_j - x_j'\tilde{\beta}_2 & if \quad j > m. \end{cases} \tag{16}$$

Lemma 2.2.b The fractional marginal likelihood (for $0 < b < 1$) is given by

$$\begin{aligned} f_b^1(y|j) &= \sum_{m=k+1}^{T-k} p_m b^{\frac{Tb-2k}{2}} (\pi E\tilde{S}S_{mj} + \pi E\tilde{S}S_j^m)^{-\frac{T-Tb}{2}} \\ &\cdot \Gamma\left(\frac{T-2k}{2}\right) / \Gamma\left(\frac{Tb-2k}{2}\right). \end{aligned} \tag{17}$$

For $b = 2k/T$ we have

$$f_b^1(y|j) = \sum_{m=k+1}^{T-k} p_m (\pi E\tilde{S}S_{mj} + \pi E\tilde{S}S_j^m)^{-\frac{T-2k}{2}} \cdot \Gamma\left(\frac{T-2k}{2}\right). \tag{18}$$

2.3 The heteroskedastic AR(p) break point model with outlier

For the heteroskedastic break point regression model with outlier

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left[\begin{pmatrix} X_1\beta_1 \\ X_2\beta_2 \end{pmatrix} + D_j\delta_j, \begin{pmatrix} \sigma_1^2 I_m & 0 \\ 0 & \sigma_2^2 I_{T-m} \end{pmatrix}\right]. \tag{19}$$

Theorem 2.3 The marginal likelihood for informative priors we assume the following conjugate prior distribution

$$\begin{pmatrix} \tilde{\beta}_{1j}|\sigma_1^2 \\ \tilde{\beta}_{2j}|\sigma_2^2 \end{pmatrix} \sim N\left[\begin{pmatrix} \tilde{b}_{1*j} \\ \tilde{b}_{2*j} \end{pmatrix}, \tilde{H}_{*j} = \begin{pmatrix} \sigma_1^2\tilde{H}_{1*j} & 0 \\ 0 & \sigma_2^2\tilde{H}_{2*j} \end{pmatrix}\right],$$

and

$$\sigma_i^{-2} \sim \Gamma(s_{i*}^2, n_{i*}), \quad \text{i=1,2.}$$

Then the marginal likelihood (conditional that point j is an outlier) is given by

$$f(y|j) = (2\pi)^{-\frac{T}{2}} \prod_{i=1}^{2} \frac{(n_{i*}s_{i*}^2)^{n_{i*}/2}}{|\tilde{H}_{i*j}|^{\frac{1}{2}}\Gamma(n_{i*}/2)} \cdot \sum_{m=k+1}^{T-k} p_m \prod_{i=1}^{2} \frac{|\tilde{H}_{i**j}|^{\frac{1}{2}}\Gamma(n_{i**}/2)}{(n_{i**}s_{i**}^2/2)^{\frac{n_{i**}}{2}}} \quad (20)$$

with parameters

$$n_{1**} = n_{1*} + m,$$

$$n_{2**} = n_{2*} + T - m,$$

$$\tilde{H}_{i**j}^{-1} = \tilde{H}_{i*j}^{-1} + \tilde{X}_{ij}'\tilde{X}_{ij},$$

$$n_{i**}s_{i**}^2 = n_{i*}s_{i*}^2 + E\tilde{S}S_{ij} + (\tilde{b}_{i*j} - \tilde{\hat{\beta}}_{ij})'((\tilde{X}_{ij}'\tilde{X}_{ij})^{-1} + \tilde{H}_{i*j})^{-1}(\tilde{b}_{i*j} - \tilde{\hat{\beta}}_{ij}),$$

$$E\tilde{S}S_{ij} = (y_i - \tilde{X}_{ij}\tilde{\hat{\beta}}_{ij})'(y_i - \tilde{X}_{ij}\tilde{\hat{\beta}}_{ij}),$$

and

$$\tilde{\hat{\beta}}_{ij} = (\tilde{X}_{ij}'\tilde{X}_{ij})^{-1}\tilde{X}_{ij}'y_i, \quad \text{i=1,2.}$$

Lemma 2.3.a The marginal likelihood with non-informative prior
The marginal likelihood for the heteroskedastic model with non-informative prior is given by

$$f(y|j) = \sum_{m=k+1}^{T-k} p_m |X_1'X_1|^{-\frac{1}{2}}|X_2'X_2|^{-\frac{1}{2}} \Gamma\left(\frac{m-k}{2}\right)\Gamma\left(\frac{T-m-k}{2}\right) h_{jm}^{-\frac{1}{2}}(\pi ESS_{1j})^{-\frac{m-k}{2}}(\pi ESS_{2j})^{-\frac{T-m-k}{2}} \quad (21)$$

with the leverage point h_{jm} given as in (15). Note that we need for $j = 1,\ldots,T$ the following expressions for the error sum of squares

$$if \quad j \le m : \begin{cases} ESS_{1j} = (y_1 - X_1\tilde{\beta}_1)'(y_1 - X_1\tilde{\beta}_1) - e_j^2/h_{jm}, \\ ESS_{2j} = (y_2 - X_2\tilde{\beta}_2)'(y_2 - X_2\tilde{\beta}_2), \end{cases} \quad (22)$$

or

$$if \quad j > m : \begin{cases} ESS_{1j} = (y_1 - X_1\tilde{\beta}_1)'(y_1 - X_1\tilde{\beta}_1), \\ ESS_{2j} = (y_2 - X_2\tilde{\beta}_2)'(y_2 - X_2\tilde{\beta}_2) - e_j^2/h_{jm}, \end{cases}$$

and the residual e_j given as in (16) with the OLS estimates

$$\tilde{\beta}_1 = (\tilde{X}_1{}'\tilde{X}_1)^{-1}\tilde{X}_1{}'y_1 \quad and \quad \tilde{X}_1 = (X_1 : D_{1j}), \tag{23}$$

or

$$\tilde{\beta}_2 = (\tilde{X}_2{}'\tilde{X}_2)^{-1}\tilde{X}_2{}'y_2 \quad and \quad \tilde{X}_2 = (X_2 : D_{2j})$$

with $e_{1j} : m \times 1$ and $e_{2j} : (T-m) \times 1$ being the dummy variable or the j-th unity vector, respectively.

Lemma 2.3.b The fractional marginal likelihood for the heteroskedastic break point model (with $b \,\epsilon\, (m^{-1}, 1 - m^{-1})$) is given by

$$\begin{aligned} f_b^1(y|j) &= \sum_{m=k+1}^{T-k} p_m b^{Tb-k+1} \Gamma\left(\frac{m-k}{2}\right) / \Gamma\left(\frac{mb-k}{2}\right) \\ &\quad \cdot (\pi E\tilde{S}S_{1j})^{-\frac{m-mb}{2}} (\pi E\tilde{S}S_{2j})^{-\frac{(T-m)(1-b)}{2}} \frac{\Gamma\left(\frac{T-m-k}{2}\right)}{\Gamma\left(\frac{(T-m)b-k}{2}\right)}. \end{aligned} \tag{24}$$

2.4 A Bayes test for outliers

Let $H_j = \{\text{outlier at } j\}$ be the hypothesis that there is an outlier at point $j = 1, \ldots, T$. The hypothesis $H_A = \{\text{outlier in the time series}\}$ is given by the union of events

$$H_A = \bigcup_{j=1}^{T} H_j.$$

The posterior odds for the test of $H_0 = \{\text{no outlier}\}$ over H_A are given by

$$\frac{P_r(H_0|y)}{P_r(H_A|y)} = \frac{f_0(y)}{f_A(y)} \cdot \frac{P_r(H_0)}{P_r(H_A)}, \tag{25}$$

and the marginal likelihood of the alternative is

$$\begin{aligned} f_A(y) &= \sum_{j=1}^{T} f(y \mid j) \cdot p_j \\ &= \frac{1}{T} \sum_{j=1}^{T} f(y \mid j) = \bar{f}(y), \end{aligned} \tag{26}$$

and for the null-hypothesis

$$f_0(y) = |X'X|^{-\frac{1}{2}}(\pi ESS)^{-\frac{T-p-1}{2}}\Gamma\left(\frac{T-p-1}{2}\right), \tag{27}$$

where we assume equal probability for the presence of an outlier at any time point: $p_j = 1/T$. Then the Bayes factor is

$$B = \frac{f_0(y)}{f_A(y)} = \frac{T \cdot (\pi ESS)^{-\frac{T-p-1}{2}}\Gamma(\frac{T-p-1}{2})}{\sum_{j=1}^{T}(\pi(ESS - \frac{e_j^2}{h_j}))^{-\frac{T-p-2}{2}} \cdot h_j^{-\frac{1}{2}}\Gamma(\frac{T-p-2}{2})}. \tag{28}$$

2.5 Bayes test for break points in an AR(p) process

Denote the following hypotheses as $H_m = \{\text{break point at } m\}$ and $H_p = \{\text{order of the AR process is } p\}$. We want to test the hypotheses that there is a break point across different orders of the AR process.
Thus, we consider the null-hypotheses and the modified null hypothesis
$\tilde{H}_0 = H_0 \bigcup \{\text{outlier at point j}\}$, i.e.
$\tilde{H}_0 = \{\text{The AR model is of order 1, or 2, or } \ldots, \text{ p}_{\text{max}}\}$.

$$H_0 = \bigcup_{p=1}^{p_{max}} H_p, \quad (\tilde{H}_0 = \bigcup_{p=1}^{p_{max}} H_p \bigcup_{j=1}^{T} H_j), \tag{29}$$

and the alternative is that there is additionally a break point:

$$H_A = \bigcup_{p=1}^{p_{max}} H_p \bigcup_{m=k+1}^{T-k} H_m, \quad (\tilde{H}_A = H_A \bigcup_{j=1}^{T} H_j). \tag{30}$$

The marginal likelihood for null-hypothesis is

$$f_0(y) = \sum_{p=1}^{p_{max}} f(y \mid p) \cdot p_p, \tag{31}$$

where p_p is the prior probability of an $AR(p)$ model with lag p and

$$f(y \mid p) = \Gamma\left(\frac{T-p-1}{2}\right)|X_p'X_p|^{-\frac{1}{2}}(\pi ESS_p)^{-\frac{T-p-1}{2}} \tag{32}$$

with $X_p = (1, x_1, \ldots, x_p)$, $ESS_p = (y - X_p\hat{\beta}_p)'(y - X_p\hat{\beta}_p)$.
Also, we find for the alternative hypothesis

$$f_A(y) = \frac{1}{p_{max}} \sum_{i=1}^{p_{max}} \overline{f}(y) \tag{33}$$

with the average marginal likelihood given as

$$\overline{f}(y \mid p) = \frac{1}{T-2k} \sum_{m=k+1}^{T-k} f(y \mid p, m), \tag{34}$$

and $f(y \mid p, m)$ is the marginal likelihood of an $AR(p)$ model with breakpoint m given by

$$f(y \mid p, m) = \Gamma\left(\frac{T-p-2}{2}\right) |X'_{p,m} X_{p,m}|^{-\frac{1}{2}} (\pi ESS_{p,m})^{-\frac{T-p-2}{2}}, \tag{35}$$

where $X_{p,m}$ and $ESS_{p,m}$ are a function of p as well.
Then Bayes factor is

$$B = \frac{f_0(y)}{f_A(y)}.$$

2.6 Bayes test for outliers and break points

Let $H_j = \{\text{outlier at } j\}$ and $H_m = \{\text{break point at } m\}$ be the hypotheses for an outlier and a break point at a particular point in time, respectively.
Then the hypothesis $H_A = \{\text{outlier and break point in the time series}\}$ is given by the union of events

$$H_A = \bigcup_{j=1}^{T} H_j \bigcup_{m=k+1}^{T-k} H_m. \tag{36}$$

The marginal likelihood for this hypothesis is

$$\begin{aligned} f_A(y) &= \sum_{j=1}^{T} \sum_{m=k+1}^{T-k} f(y \mid j, m) \cdot p_j p_m \\ &= \frac{1}{T-2k} \sum_{m=k+1}^{T-k} \overline{f}(y \mid m) \end{aligned} \tag{37}$$

with the marginal likelihood conditional on a break point m is given by

$$\overline{f}(y \mid m) = \frac{1}{T}\sum_{j=1}^{T} f(y \mid m, j). \tag{38}$$

Alternatively, the marginal likelihood can be calculated as

$$f_A(y) = \frac{1}{T}\sum_{j=1}^{T} f(y \mid j), \tag{39}$$

and $f(y \mid j)$ is given by (14) or (21), and $f_0(y)$ given by

$$f_0(y) = \frac{1}{T}\sum_{j=1}^{T} \Gamma\left(\frac{T-p-2}{2}\right) |\tilde{X}_j' \tilde{X}_j|^{-\frac{1}{2}} (\pi \widetilde{ESS}_j)^{-\frac{T-p-2}{2}}. \tag{40}$$

Then the Bayes factor is given by the ratio of (40) and (2.6):

$$B = \frac{f_0(y)}{f_A(y)}. \tag{41}$$

2.7 Bayes test for outliers and heteroskedastic break points

Let H_j and H_m be the hypotheses as before and consider the additional hypotheses $H_{h=0} = \{\text{homoskedastic} \quad \text{model}\}$ and $H_{h=1} = \{\text{heteroskedastic} \quad \text{model}\}$. Then the alternative hypothesis can be constructed as the union of simpler hypotheses:

$$H_A = \bigcup_{j=1}^{T} H_j \bigcup_{m} H_m \bigcup_{h=0}^{1} H_h. \tag{42}$$

Note that the null-hypothesis is simply given by $H_0 = \bigcup_{j=1}^{T} H_j$, since heteroskedasticity is only connected with the break point model. The marginal likelihood for this problem is

$$f(y \mid j, h=0) \rightarrow \ (\mathrm{H_{h=0}}), \qquad \text{and} \qquad f(y \mid j, h=1) \rightarrow \ (H_{h=1}). \tag{43}$$

Assuming equal probability, the marginal likelihood is given as the average

$$\overline{f}_A(y \mid j) = \frac{1}{2}[f(y \mid j, h=0) + f(y \mid j, h=1)]. \tag{44}$$

Furthermore the marginal likelihood for the alternative is given by the average

$$f_A(y) = \frac{1}{T}\sum_{j=1}^{T} \overline{f}_A(y \mid j). \tag{45}$$

The marginal likelihood under the null hypothesis is

$$f_0(y) = \frac{1}{T}\sum_{j=1}^{T} f_0(y \mid j) \tag{46}$$

with

$$f_0(y) = \frac{1}{T}\sum_{j=1}^{T} \Gamma\left(\frac{T-p-2}{2}\right) |\tilde{X}_j' \tilde{X}_j|^{-\frac{1}{2}} (\pi \widetilde{ESS}_j)^{-\frac{T-p-2}{2}}. \tag{47}$$

Finally the Bayes factor is given by

$$B = \frac{f_A(y)}{f_0(y)}. \tag{48}$$

2.8 Bayes tests for outlier with informative priors

Assuming an AR(p) process with conjugate prior, we test the models

$$H_0 : y = X\beta + \epsilon \tag{49}$$

against

$$H_A : y = \tilde{X}\tilde{\beta} + \epsilon = X\beta + D_j\vartheta + \epsilon, \quad j = 1, \ldots, T. \tag{50}$$

The marginal likelihood under H_0 is given by

$$f_0(y) = \pi^{-\frac{T}{2}} \frac{|H_{**}|^{\frac{1}{2}}}{|H_*|^{\frac{1}{2}}} \frac{\Gamma(\frac{n_{**}}{2})}{\Gamma(\frac{n_*}{2})} \frac{(n_* s_*)^{n_*/2}}{(n_{**} s_{**})^{n_{**}/2}}, \tag{51}$$

and for H_A we have to average

$$\overline{f}_A(y) = \frac{1}{T}\sum_{j=1}^{T} f_A(y \mid j), \tag{52}$$

where

$$f_A(y \mid j) = \pi^{-\frac{T}{2}} \frac{|\tilde{H}_{**}|^{\frac{1}{2}}}{|\tilde{H}_*|^{\frac{1}{2}}} \frac{\Gamma(\frac{n_{**}}{2})}{\Gamma(\frac{n_*}{2})} \frac{(n_* \tilde{s}_*)^{n_*/2}}{(n_{**} \tilde{s}_{**})^{n_{**}/2}}, \tag{53}$$

Note that $|\tilde{H}_*| = s_*^2 |H_*|$. Therefore we have for the Bayes factor of H_0 against H_A

$$
\begin{aligned}
B_A^0 &= \frac{|H_{**}|^{\frac{1}{2}}(n_* s_*)^{n_*/2}|H_*|^{-\frac{1}{2}}(n_{**}s_{**})^{-n_{**}/2}}{\frac{1}{T}\sum_{j=1}^T |\tilde{H}_{**}|^{\frac{1}{2}}(n_*\tilde{s}_*)^{n_*/2}|\tilde{H}_*|^{-\frac{1}{2}}(n_{**}\tilde{s}_{**})^{-n_{**}/2}} \\
&= \frac{|H_{**}|^{\frac{1}{2}}(n_* s_*)^{n_*/2}(n_{**}s_{**})^{-n_{**}/2}}{\frac{1}{T}\sum_{j=1}^T |\tilde{H}_{**}|^{\frac{1}{2}}(n_*\tilde{s}_*)^{n_*/2}s_*^{-1}(n_{**}\tilde{s}_{**})^{-n_{**}/2}}. \qquad (54)
\end{aligned}
$$

3 Multivariate outlier models

3.1 Multivariate outlier models

In this section we consider the following multivariate regression model with outliers

$$Y = XB + D_j\delta + U, \quad j = 1, \ldots, T. \qquad (55)$$

Again D_j is a dummy variable defined as the j-th unity vector of dimension T, and δ is a $(1 \times M)$ row vector of outliers (a different location shift for the M regressions).

The errors are multivariate normally distributed:

$$U \sim N_{T\times M}[0, \Sigma \otimes I_T].$$

The multivariate model can be written in compact form as

$$Y = \tilde{X}\tilde{B} + U, \quad or \quad Y \sim N[\tilde{X}\tilde{B}, \Sigma \otimes I_T] \qquad (56)$$

with $\tilde{X} = (X : D_j)$ a $T \times (k+1)$ regressor matrix, and $\tilde{B} = (B : \theta_j)'$ the $(k+1) \times M$ matrix of regression coefficients. We derive the marginal likelihood for 3 different priors.

Theorem 3.1 Multivariate break points with outliers
For the multivariate break point model (56) we assume a (conjugate) normal-Wishart prior

$$p(\tilde{B}, \tilde{\Sigma}) = NW[\tilde{B}_*, \tilde{H}_*, \tilde{\Sigma}_*, n_*]. \qquad (57)$$

In the context of a multivariate VAR model one can use the univariate outlier model as in section 2.1 for all the M regressions.

The marginal likelihood for a break point at j and informative prior is given by

$$f(Y|j) = (2\pi)^{-\frac{MT}{2}} \frac{c_{n_{**}}}{c_{n_*}} \frac{|\Sigma_*|^{\frac{n_*}{2}}}{|\Sigma_{**}|^{\frac{n_{**}}{2}}} \frac{|\tilde{H}_{**}|^{M/2}}{|\tilde{H}_*|^{M/2}}, \tag{58}$$

with the parameters

$$\Sigma_{**} = \Sigma_* + \tilde{U}'\tilde{U} + \Delta, \quad \tilde{U} = Y - \tilde{X}\hat{\tilde{B}},$$

$$hat\tilde{B} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y,$$

$$\tilde{H}_{**}^{-1} = \tilde{X}'\tilde{X} + H_*^{-1} = \tilde{X}'\tilde{X} - x_j x_j' + H_*^{-1},$$

$$\Delta = (\hat{\tilde{B}} - \tilde{B}_*)'[(\tilde{X}'\tilde{X})^{-1} + H_*]^{-1}(\hat{\tilde{B}} - \tilde{B}_*),$$

$$c_{n*} = 2^{\frac{Mn_*}{2}} \pi^{\frac{M(M-1)}{4}} \prod_{j=1}^{M} \Gamma[\frac{n_* + 1 - j}{2}], \tag{59}$$

and $n_{**} = n_* + T,\ x_j = X'D_j, \quad j = 1, \dots, T.$
Proof: Since the model is structurally equivalent to the ordinary normal-Wishart regression model, the result follow from Polasek and Ren (1997).

Lemma 3.1.a Multivariate regression with outliers (non-informative prior)
The multivariate model for outliers with non-informative prior is given as in (55). Then the marginal likelihood for the event { break point in the time series } is given by

$$f(Y|j) = |X'X|^{-\frac{M}{2}} h_j^{-\frac{M}{2}} \Gamma_M \left(\frac{T-p-2}{2}\right) \pi^{-\frac{M}{2}(T-k)} |\hat{U}'\hat{U} - \hat{u}_j \hat{u}_j' h_j^{-1}|^{-\frac{M(T-k)}{2}} \tag{60}$$

with $\hat{u}_j = \hat{U}'D_j$ and

$$\hat{U} = Y - X\hat{B} - D_j\hat{\delta} = Y - \tilde{X}\hat{\tilde{B}},$$

where $\tilde{X} = (X : D_j)$ and $\hat{\tilde{B}} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y$ is the multivariate OLS estimator.
Proof: See Polasek and Ren (1997).

Lemma 3.1.b Multivariate regression with outliers (fractional prior)

For the homoskedastic multivariate regression model (55) the fractional marginal likelihood for bϵ(0,1) is given by

$$f_b^1(Y) = b^{\frac{Tb}{2}} |\pi \hat{U}'\hat{U}|^{-\frac{T(1-b)}{2}} \Gamma_M \left(\frac{T-k}{2}\right) / \Gamma_M \left(\frac{Tb-k}{2}\right) \tag{61}$$

with

$$\hat{\tilde{U}} = Y - \tilde{X}\hat{\tilde{B}}, \quad and \quad \hat{\tilde{B}} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y.$$

Note that for $b = \frac{k+1}{2}$ (and $t = 0$) we get the simpler formula

$$f_b^1(Y) = \left(\frac{k+1}{T}\right)^{\frac{k+1}{2}} |\pi \hat{U}'\hat{U}|^{-\frac{T-k-1}{2}} \Gamma_M \left(\frac{T-k+1}{2}\right).$$

Proof: Use the results of Polasek and Ren (1997) for model (56).

3.2 Multivariate regression with break points and outliers

In this section we are deriving the marginal likelihood with 3 priors for the multivariate break point and outlier model given by

$$\begin{aligned} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} &\sim N\left[\begin{pmatrix} X_1 B_1 \\ X_2 B_2 \end{pmatrix} + D_j \delta, \Sigma \otimes I_T\right] \\ &= N\left[\begin{pmatrix} \tilde{X}_1 \tilde{B}_1 \\ \tilde{X}_2 \tilde{B}_2 \end{pmatrix}, \Sigma \otimes I_T\right], \end{aligned} \tag{62}$$

where the unknown break point can lie in $k+1 \leq m \leq T-k$.

Theorem 3.2 Homoskedastic multivariate break point model with outlier (informative priors)
Consider model (62) with the conjugate prior distribution

$$f(\tilde{B}_1, \tilde{B}_2, \Sigma^{-1}) = N[\tilde{B}_{1*}, \Sigma \otimes \tilde{H}_{1*}] N[\tilde{B}_{2*}, \Sigma \otimes \tilde{H}_{2*}] W[\Sigma_*, n_*]. \tag{63}$$

Then the marginal likelihood for the multivariate regression model is given by

$$f(Y) = (2\pi)^{-\frac{MT}{2}} \sum_{m=k+1}^{T-k} p_m \frac{c_{n**}}{c_{n*}} |H_{1**}|^{M/2} |H_{2**}|^{M/2} |\Sigma_{**}|^{n_{**}/2} \tag{64}$$

with the constants c_{n**} and c_{n*} given as in (59), $p_m = \frac{1}{T-k}$,

$$H_{i**}^{-1} = H_{i*}^{-1} + X_i' X_i, \quad i = 1, 2, \tag{65}$$

$$\Sigma_{**} = \Sigma_* + \sum_{i=1}^{2} \{\hat{U}_i' \hat{U}_i + (\hat{B}_i - B_{i*})' [(X_i' X_i)^{-1} + H_{i*}^{-1}]^{-1} (\hat{B}_i - B_{i*})\}, \tag{66}$$

and

$$\hat{U}_i = Y_i - X_i \hat{B}_i, \quad \hat{B}_i = (\tilde{X}_i' \tilde{X}_i)^{-1} \tilde{X}_i' Y_i, \quad i = 1, 2. \tag{67}$$

Proof:
Use the results of the break point model in Polasek and Ren (1997).

Lemma 3.2.a Homoskedastic break point model with outliers (non-informative prior)
We consider model (62) with non-informative prior

$$f(B_1, B_2, \Sigma) \propto |\Sigma|^{-\frac{M+1}{2}}.$$

Then the marginal likelihood for the event { outlier in Y } is given by

$$\begin{aligned} f(Y) \quad = \quad & \sum_{m=k+1}^{T-k} p_m |\hat{U}_1' \hat{U}_1 + \hat{U}_2' \hat{U}_2|^{k-\frac{T}{2}} \Gamma_M(\frac{T-k}{2}) \\ & \pi^{-\frac{M}{2}(T-2k)} |X_1' X_1|^{-\frac{M}{2}} |X_2' X_2|^{-\frac{M}{2}} h_{jm}^{-1/2} \end{aligned} \tag{68}$$

with the rows of U_1 and U_2 given as

$$\begin{cases} \hat{u}_{j1}' = y_{j1}' - x_{j1}' \hat{\beta}_1, & for \quad j \leq m, \\ \hat{u}_{j2}' = y_{j1}' - x_{j1}' \hat{\beta}_2, & for \quad j > m, \end{cases} \tag{69}$$

and

$$\hat{U}'_1 \hat{U}_1 = \begin{cases} (Y_1 - X_1 \hat{B}_1)'(Y_1 - X_1 \hat{B}_1) - \hat{u}_{j1} \hat{u}_{j1}' h_{j1}^{-1}, & for \quad j \leq m, \\ (Y_1 - X_1 \hat{B}_1)'(Y_1 - X_1 \hat{B}_1), \quad for \quad j > m, \end{cases} \tag{70}$$

$$\hat{U}'_2 \hat{U}_2 = \begin{cases} (Y_2 - X_2 \hat{B}_2)'(Y_2 - X_2 \hat{B}_2) - \hat{u}_{j2} \hat{u}_{j2}' h_{j2}^{-1}, & for \quad j > m, \\ (Y_2 - X_2 \hat{B}_2)'(Y_2 - X_2 \hat{B}_2), \quad for \quad j \leq m, \end{cases} \tag{71}$$

The leverage points h_{ji} are given as in (15), and the y_{ji}' and x_{ji}' are j-th rows of Y_i and X_i, $i = 1, 2$, respectively.

Proof: In analogy to theorem 3.1 and observe the location of the outliers.

Lemma 3.2.b Homoskedastic break point model with outliers (fractional prior)
We consider model (62) and use the fractional prior approach of O'Hagan (1995). Then the marginal likelihood for the event { outlier in Y } is given by

$$\begin{aligned} f_b^1(Y) &= \sum_{m=k+1}^{T-k} p_m |\pi(\hat{U}_1'\hat{U}_1 + \hat{U}_2'\hat{U}_2)|^{-\frac{T-Tb}{2}} b^{\frac{Tb}{2}} \\ &\quad \Gamma_M(\frac{T-2k}{2})/\Gamma_M(\frac{Tb-2k}{2}) \end{aligned} \quad (72)$$

with the rows of $\hat{U}_1$ and $\hat{U}_2$ defined as in (70) and (71).
For $b = \frac{2k+2}{2}$ we obtain a slight simplification

$$\begin{aligned} f_b^1(Y) &= \sum_{m=k+1}^{T-k} p_m \left(\frac{2k+2}{T}\right)^{k+1} |\pi(\hat{U}_1'\hat{U}_1 + \hat{U}_2'\hat{U}_2)|^{-\frac{T}{2}-k-1} \\ &\quad \Gamma_M(\frac{T-2k}{2}). \end{aligned} \quad (73)$$

3.3 Heteroskedastic multivariate break points and outliers

In this section we extend the previous discussion of the multivariate break point model and we analyse the heteroskedastic multivariate regression model with break points and outliers.
The model is given as an extension of the homoskedastic model (62)

$$\begin{aligned} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} &\sim N\left[\begin{pmatrix} X_1B_1 \\ X_2B_2 \end{pmatrix} + D_j\delta, \begin{pmatrix} \Sigma_1 \otimes I_{n_1} & 0 \\ 0 & \Sigma_2 \otimes I_{n_2} \end{pmatrix}\right] \\ &= N\left[\begin{pmatrix} \tilde{X}_1\tilde{B}_1 \\ \tilde{X}_2\tilde{B}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 \otimes I_{n_1} & 0 \\ 0 & \Sigma_2 \otimes I_{n_2} \end{pmatrix}\right] \end{aligned} \quad (74)$$

with $n_1 = m$ and $n_2 = T - m$.

Theorem 3.3 Heteroskedastic break point model with outliers (informative prior)
We consider the model (74) with the following conjugate prior distribution

$$f(\tilde{B}_1, \tilde{B}_2, \tilde{\Sigma}_1^{-1}, \tilde{\Sigma}_2^{-1}) = \prod_{i=1}^{2} N[\tilde{B}_{i*}, \Sigma_{i*} \otimes \tilde{H}_{i*}] W[\Sigma_{i*}, n_{i*}],$$

where the $* - Index$ denotes known prior parameter matrices.
Then the marginal likelihood for the multivariate model is given by

$$f(Y) = \sum_{m=k+1}^{T-k} p_m f(Y_1) f(Y_2), \tag{75}$$

where the marginal likelihoods of the upper and lower models are given by

$$f(Y_i) = (2\pi)^{-Mn_i} \frac{c_{n**i}}{c_{n*i}} \frac{|\tilde{\Sigma}_{**i}|^{\frac{n_{**i}}{2}}}{|\tilde{\Sigma}_{*i}|^{\frac{n_{*i}}{2}}} \frac{|\tilde{H}_{**i}|^{\frac{M}{2}}}{|\tilde{H}_{*i}|^{\frac{M}{2}}}, \quad i = 1, 2$$

with c_{n*i} and c_{n**i} given as in (59).

Lemma 3.3.a Heteroskedastic break point model with outliers (non-informative prior)
Consider the heteroskedastic multivariate break point model with outliers as in (74). The marginal likelihood for the event { outlier in Y } with non-informative prior is

$$\begin{aligned} f(Y) \quad = \quad & \sum_{m=k+1}^{T-k} p_m |X_1' X_1|^{-\frac{M}{2}} |X_2' X_2|^{-\frac{M}{2}} h_{jm}^{-M/2} \\ & \Gamma_M\left(\frac{m-k}{2}\right) \Gamma_M\left(\frac{T-m-k}{2}\right) \\ & |\pi \hat{U}_{1j}' \hat{U}_{1j}|^{-\frac{M(m-k)}{2}} |\pi \hat{U}_{2j}' \hat{U}_{2j}|^{-\frac{M(T-m-k)}{2}} \end{aligned} \tag{76}$$

with h_{jm} given as in (15), and the sum of squares matrix of the residuals are given by

$$if \quad j \le m : \begin{cases} \hat{U}_{1j}' \hat{U}_{1j} = \hat{U}_1' \hat{U}_1 - \hat{u}_j \hat{u}_j' h_{jm}^{-1}, \\ \hat{U}_{2j}' \hat{U}_{2j} = \hat{U}_2' \hat{U}_2, \end{cases} \tag{77}$$

$$if \quad j > m : \begin{cases} \hat{U}_{1j}'\hat{U}_{1j} = \hat{U}_1'\hat{U}_1, \\ \hat{U}_{2j}'\hat{U}_{2j} = \hat{U}_2'\hat{U}_2 - \hat{u}_j\hat{u}_j'h_{jm}^{-1}. \end{cases} \tag{78}$$

Proof: Apply lemma 3.1.a twice.

Lemma 3.3.b Heteroskedastic break point model with outliers (fractional prior)

We consider the heteroskedastic break point model with outliers (74) and use the fractional prior approach of O'Hagan (1995). Then the marginal likelihood for fractional prior is

$$
\begin{aligned}
f(Y) &= \sum_{m=2k+1}^{T-2k} p_m b^{Tb} |\pi \hat{U}_1' \hat{U}_1|^{-\frac{n_1(1-b)}{2}} |\pi \hat{U}_2' \hat{U}_2|^{-\frac{n_2(1-b)}{2}} \\
&\quad \frac{\Gamma_M\left(\frac{n_1-k+1}{2}\right)}{\Gamma_M\left(\frac{n_1 b-k+1}{2}\right)} \frac{\Gamma_M\left(\frac{n_2-k+1}{2}\right)}{\Gamma_M\left(\frac{n_2 b-k+1}{2}\right)} \\
&= \sum_{m=2k+1}^{T-2k} p_m b^{Tb} \prod_{i=1}^{2} |\pi \hat{U}_i' \hat{U}_i|^{-\frac{n_i(1-b)}{2}} \\
&\quad \frac{\Gamma_M\left(\frac{n_i-k+1}{2}\right)}{\Gamma_M\left(\frac{n_i b-k+1}{2}\right)}, \qquad (79)
\end{aligned}
$$

where the sum of squares matrices for $\hat{U}_1$ and $\hat{U}_2$ are given as in (77) and (78).

3.4 The Bayes factor for the multivariate outlier models

We consider the two multivariate regression models with the matrices $Y : n \times M$ and $X : n \times k$

$$
Y \sim N[XB, \Sigma \otimes I_n], \qquad (80)
$$

$$
Y \sim N[XB + D_\delta A, \Sigma \otimes I_n]. \qquad (81)
$$

The fractional marginal likelihood for the model (81) is

$$
f_b^1(Y) = b^{\frac{nb}{2}} |\pi \hat{U}'_1 \hat{U}_1|^{-\frac{n(1-b)}{2}} \Gamma_M\left(\frac{n-k}{2}\right) / \Gamma_M\left(\frac{nb-k}{2}\right). \qquad (82)
$$

The fractional Bayes factor for any fraction $0 < b < 1$ is given by

$$
B_b = c_b |\hat{U}'_0 \hat{U}_0 (\hat{U}'_1 \hat{U}_1)^{-1}|^{-\frac{n(1-b)}{2}} \qquad (83)
$$

with

$$
c_b = \frac{\Gamma_M\left(\frac{n-k}{2}\right)}{\Gamma_M\left(\frac{n-k-1}{2}\right)} \frac{\Gamma_M\left(\frac{nb-k}{2}\right)}{\Gamma_M\left(\frac{nb-k-1}{2}\right)},
$$

$$\hat{U}_1 = Y - X\hat{B}, \quad \hat{B} = (X'X)^{-1}X'Y,$$

$$\hat{U}_2 = Y - X\hat{B} - D_\delta\hat{A}, \quad (\hat{B}, \hat{A})' = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y,$$

and

$$\tilde{X} = (X : D_\delta).$$

4 Examples

Table 1 analyses the (real) Swiss GNP and (real) Swiss Consumption for the period 1970 Q1 to 1993 Q2 for fractional prior where the Bayes factor is calculated by formula (53). For both time series the marginal likelihood selects the lag order $p = 2$ and in both cases the outlier model is selected also for lag 2. If we adopt the 9:19:99 rule for quantifying evidence with Bayes factors and the log equivalent (i.e. $\ln 9 = 2.2$, $\ln 19 = 2.9$ and $\ln 99 = 4.6$) then we see that differences in the log marginal likelihood are quite close. In the last row we have calculated the averages of the log likelihood over the 5 periods.

In Table 2 we used the informative Bayes model to test for stationarity/non-stationarity in the time series. The first two columns evaluates the non-stationarity of the GNP and consumption series. Lag order 5 is found for GNP and lag 2 for consumption and the unit root model (non-stationarity) is accepted.

Table 3 is an extended analysis for the outlier model in Table 1. Now we include a break point model as an alternative and we discover that the break point model with outliers fits the data best. Also the average marginal likelihood over all lag length indicate that there is decisive evidence for the AR break point model.

While Table 3 tests the homoskedastic case, Table 4 shows the break point comparison for marginal likelihoods. Now we see that the marginal likelihood picks up lag length 2 in both models and the differences in consumption is larger than the model without trend.

The bivariate AR regression break point model is shown in Table 5 and favours slightly VAR outlier model for order $p = 1$.

5 Conclusions

This paper shows how the marginal likelihood approach for model selection of Polasek and Ren (1997) can be applied to the univariate and multivariate linear model in the presence of outliers and break points. The model assumes a normal-gamma or a normal-Wishart conjugate prior distribution which allows a closed form integration for calculating the Bayes factor and the marginal likelihood. The fractional Bayes factor of O'Hagan (1995) can be fitted nicely in this framework and lead to quite simple formulars in the case of non-informative priors. The approach is demonstrated for a macroeconomic example involving Swiss income and consumption. We find that the AR break point model and VAR outlier model is favored over the simple AR models. A further extension of this approach to AR model selection in presence of unit roots can be found in Pelloni and Polasek (1998).

	GNP		
Order	AR	AR-outlier	Bayes factor
1	-28.8602	-27.6991	0.3131
2	-25.2707*	-23.2414*	0.1314
3	-26.2865	-24.2599	0.1318
4	-26.1485	-24.6281	0.2186
5	-26.7735	-24.8610	0.1477
Ave.	-26.6679	-24.9379	0.1885
	Consumption		
Order	AR	AR-outlier	Bayes factor
1	-86.9461	-81.5249	0.0044
2	-83.6903*	-77.2270*	0.0029
3	-83.9526	-76.5808	0.0019
4	-84.8609	-77.2905	0.0005
5	-85.6335	-77.3346	0.0002
Ave.	-85.0167	-77.9916	0.0019

Table 1: The log marginal likelihood of AR model, the log of average marginal likelihood (6) and of AR outlier model for Swiss GNP and consumption from 1970 Q1 to 1993 Q2 with fractional prior and Bayes factor (28)

(* maximum marginal likelihood)

Order	Nonstationarity		Stationarity	
p	GNP	Consumption	GNP	Consumption
1	-26.7873	-81.3930	-26.9553	-84.1907
2	-26.1211	-80.0339*	-27.7769	-84.0320*
3	-25.4228	-81.0418	-27.1269	-85.2720
4	-24.6705	-81.3218	-27.6295	-85.9959
5	-22.1161*	-82.3220	-26.2795*	-86.7633
6	-22.6250	-83.4973	-26.5631	-87.9805

Table 2 AR model: The unit root (stationarity) test with the log of marginal likelihood for Swiss GNP and consumption (original data) from 1970 Q1 to 1993 Q2 and fractional prior
(* maximum marginal likelihood)

	GNP		
Order	AR break point	AR break point with outlier	Bayes factor
1	-26.4391	-23.5739	0.0569
2	-24.8772*	-20.1199*	0.0085
3	-26.7824	-20.8784	0.0027
Ave.	-26.0329	-21.5241	0.0227

	Consumption		
Order	AR break point	AR break point with outlier	Bayes factor
1	-78.7601	-79.6437	2.4195
2	-75.2243*	-77.3578	8.4443
3	-76.6213	-72.4543*	0.0154
4	-79.1109	-73.6674	0.0043
Ave.	-77.4292	-75.7808	2.7208

Table 3 : The log marginal likelihood of the AR break point model and (6) AR break point (homoskedastic) with outlier for Swiss GNP and consumption from 1970 Q1 to 1993 Q2 and fractional prior

(* maximum marginal likelihood)

	GNP		
Order	AR break point	AR break point with outlier	Bayes factor
1	-27.7853	-26.9401	0.4294
2	-26.2791*	-24.8473**	0.2388
3	-28.6702	-25.3060	0.0345
Ave.	-27.5782	-25.6978	0.2342

	Consumption		
Order	AR break point	AR break point with outlier	Bayes factor
1	-78.7653	-71.8447	0.0009
2	-74.7624*	-69.8956**	0.0076
3	-75.4322	-72.9279	0.0817
4	-75.0132	-73.7331	0.2780
Ave.	-75.9932	-72.1003	0.0920

Table 4 : The log marginal likelihood of the AR break point model and (6) AR break point (heteroskedastic) with outlier for Swiss GNP and consumption from 1970 Q1 to 1993 Q2 and fractional prior

(* maximum marginal likelihood)

p	VAR	VAR-outlier	Bayes factor
1	-60.6871	-60.3824	0.7373
2	-53.6207*	-53.4474**	0.8408
3	-56.3326	-56.1885	0.8658
3	-58.3634	-58.2465	0.8896
Ave.	-57.2509	-57.0662	0.8333

Table 5 VAR(p) and VAR(p) with outlier: The log marginal likelihood of VAR model and the log of average marginal likelihood of VAR outlier model for Swiss GNP and consumption from 1970 Q1 to 1993 Q2 with fractional prior

p	Nonstationarity	Stationarity
1	-50.4197*	-56.9198*
2	-52.9816	-59.8293
3	-52.3332	-61.3383
4	-56.9765	-66.2083

Table 6 VAR model: The unit root (stationarity) test with the log of marginal likelihood for Swiss GNP and consumption from 1970 Q1 to 1993 Q2 and fractional prior

(* maximum marginal likelihood)

References

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

[2] Andrews, D. W. K., Lee, I. and Ploberger, W. (1996), Optimal change point tests for normal linear regression, Jou. of Econometrics 70, 9-38.

[3] Andrews, D. W. K. and Ploberger, W. (1994), Optimal tests when a nuisance parameter is present only under the alternative, Econometrica 62, 1383-1414.

[4] Barbieri, M. and Conigliani, C. (1997), Bayesian analysis of autoregressive time series with break points, working paper, University of Roma and Nottingham.

[5] Basci, S., Mukhopadhyay, C. and Zaman, A. (1996), Detecting structural change when the change point is unknown, Bilkent University, Turkey.

[6] Berger, J. O., and Pericchi, L. R. (1996), The intrinsic Bayes factor for model selection and prediction, Journal of the American Statistical Association 91, 109-122.

[7] Chib, S. (1995), Marginal likelihood from the Gibbs output, JASA 90, 1313-1321.

[8] Chow, G. (1960), Tests of the equality between two sets of coefficients in two linear regressions, Econometrica, 65, 561-605.

[9] Litterman R.B. (1986), A statistical approach to economic forecasting, Journal of Business and Economic Statistics 4, 1-24.

[10] O'Hagan, A. (1995), Fractional Bayes factors for model comparison (with discussions), JRSSB, 57, 99-138.

[11] Pelloni, G. and Polasek, W. (1998), The sectoral volatility hypothesis: A comparison of sectoral employment in the US and Germany, University of Basel, discussion paper.

[12] Polasek, W. and Pötzelberger, K. (1989), Robust Bayesian analysis of a parameter change in linear regression, Empirical Economics, Physica-Verlag, Heidelberg, 14, 123-137.

[13] Polasek, W. (1989), Local autoregression for detection of change in causality, Hackl, P. (ed.) Statistical analysis and forecasting of econometric structural change, Springer-Verlag, 407-440.

[14] Polasek W., Ren L. (1997), Structural breaks and model selection with marginal likelihoods, Proceedings of the Workshop on Model Selection in Racugno W. (ed.), Cagliari 223-274 (with discussion), Pitagora Ed. Bologna.

[15] Smith, A.F.M. (1975), A Bayesian approach to inference about a change-point in a sequence of random variables, Biometrika 62, 407-416.

[16] Verdinelli I., Wasserman L. (1990), Bayesian analysis of outlier problems using the Gibbs sampler, Statistics and Computing 1991-1, 105-117.

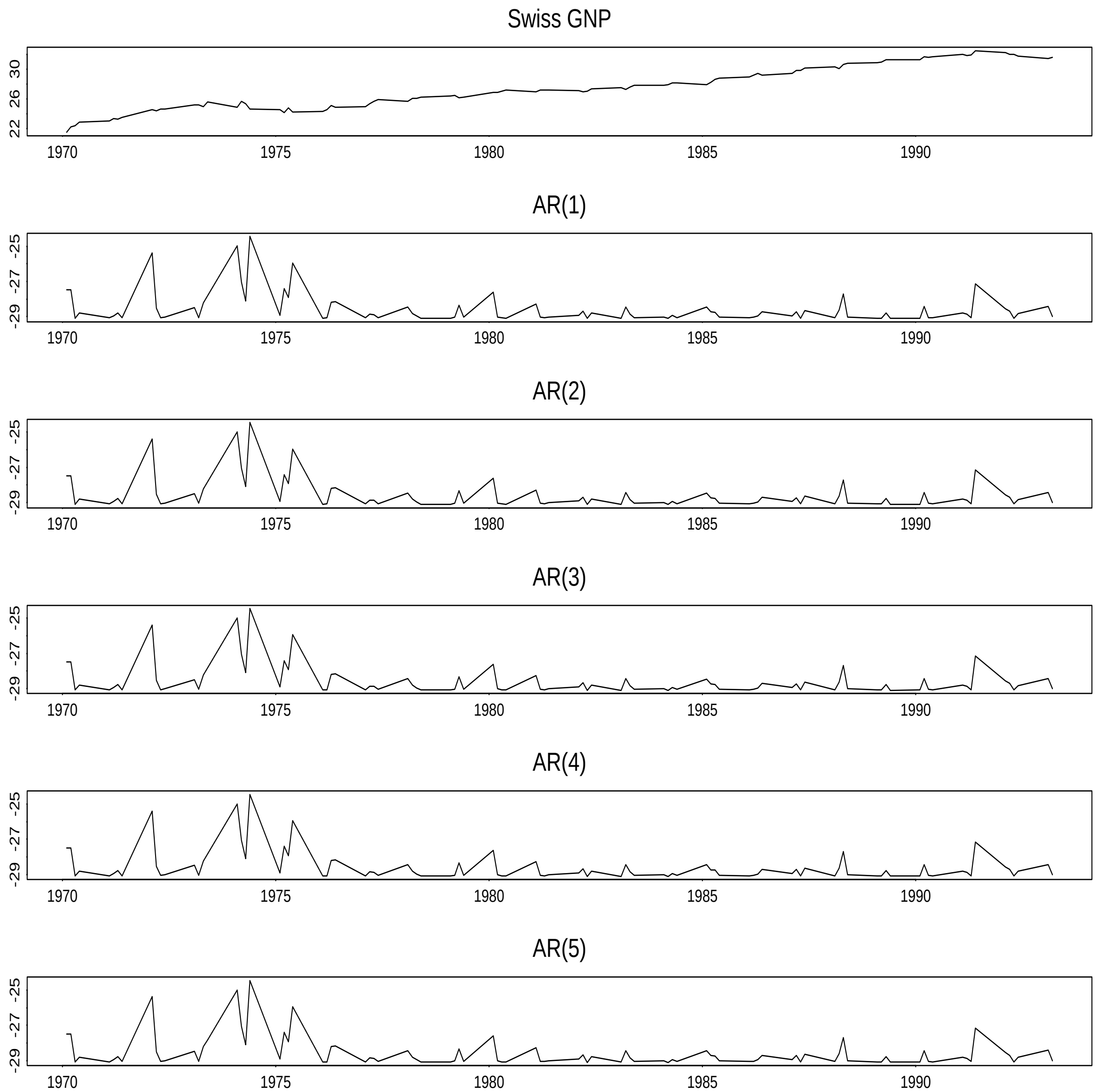

Figure 1: The log marginal likelihood (6) of the Swiss GNP from 1970 Q1 to 1993 Q2 for AR(p) outlier model

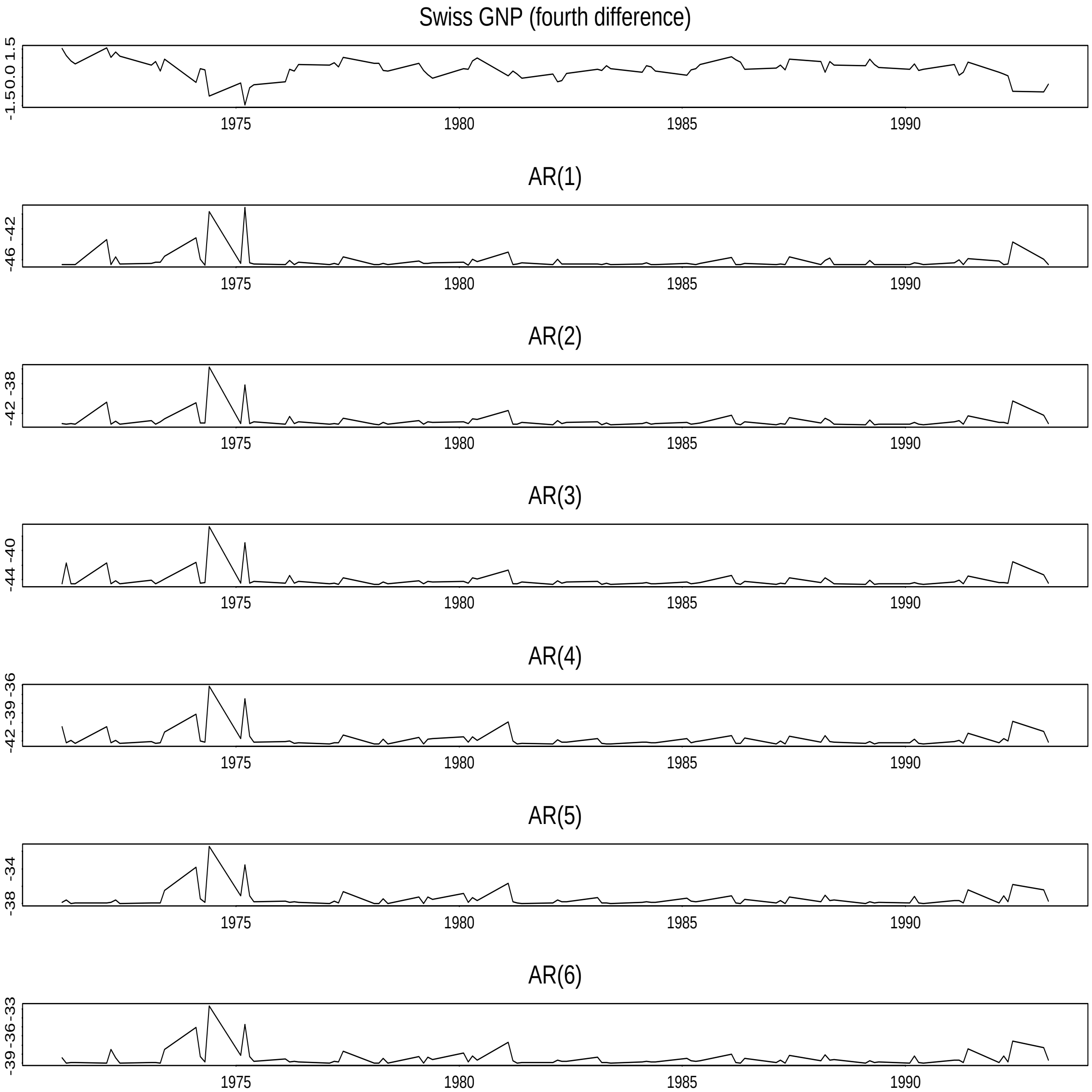

Figure 2: Fourth difference: Marginal likelihood (6) of the Swiss GNP (fourth difference) from 1970 Q1 to 1993 Q2 for AR(p) outlier model

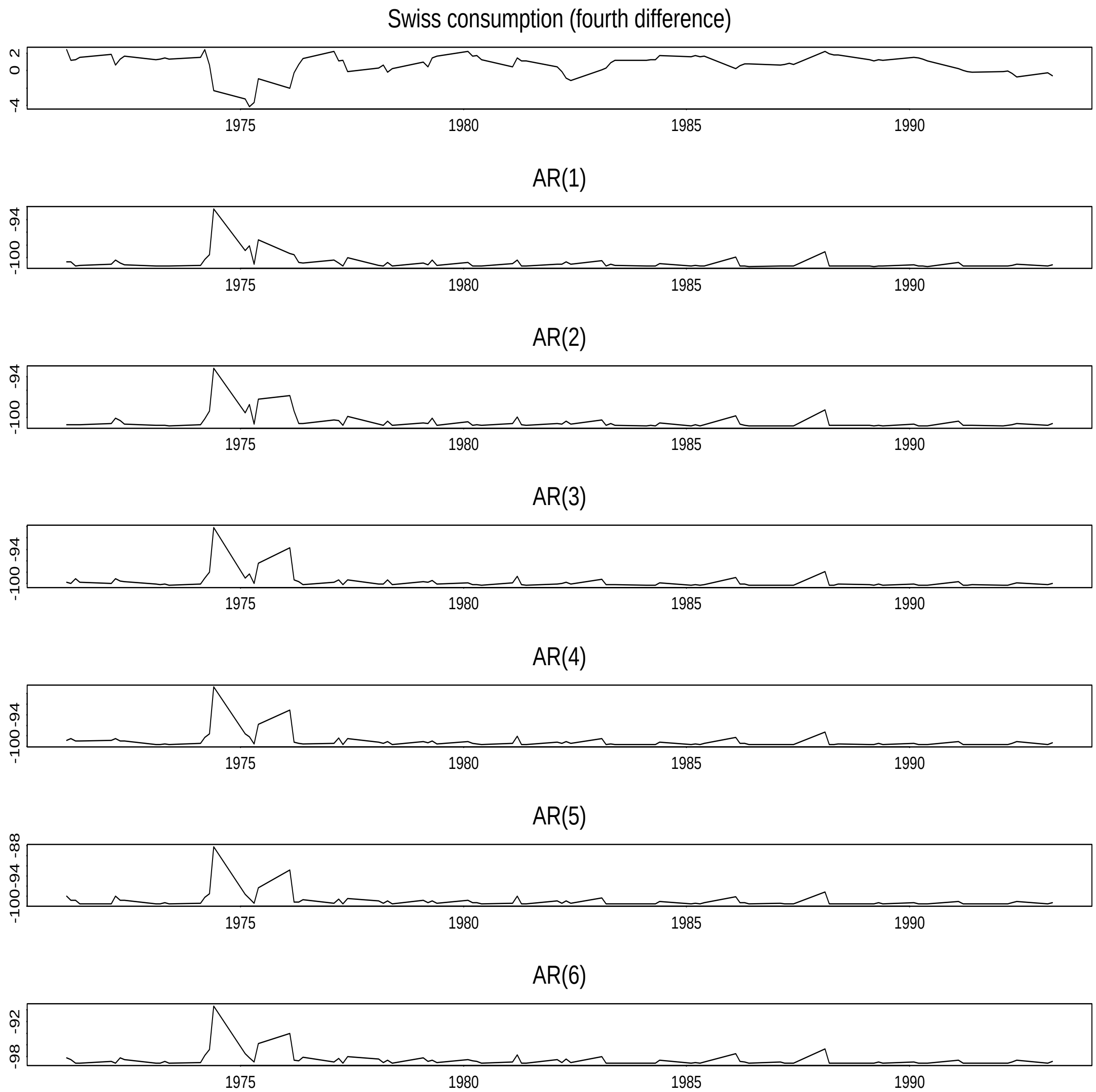

Figure 3: Fourth difference: Marginal likelihood (6) of the Swiss consumption (fourth difference) from 1970 Q1 to 1993 Q2 for AR(p) outlier model

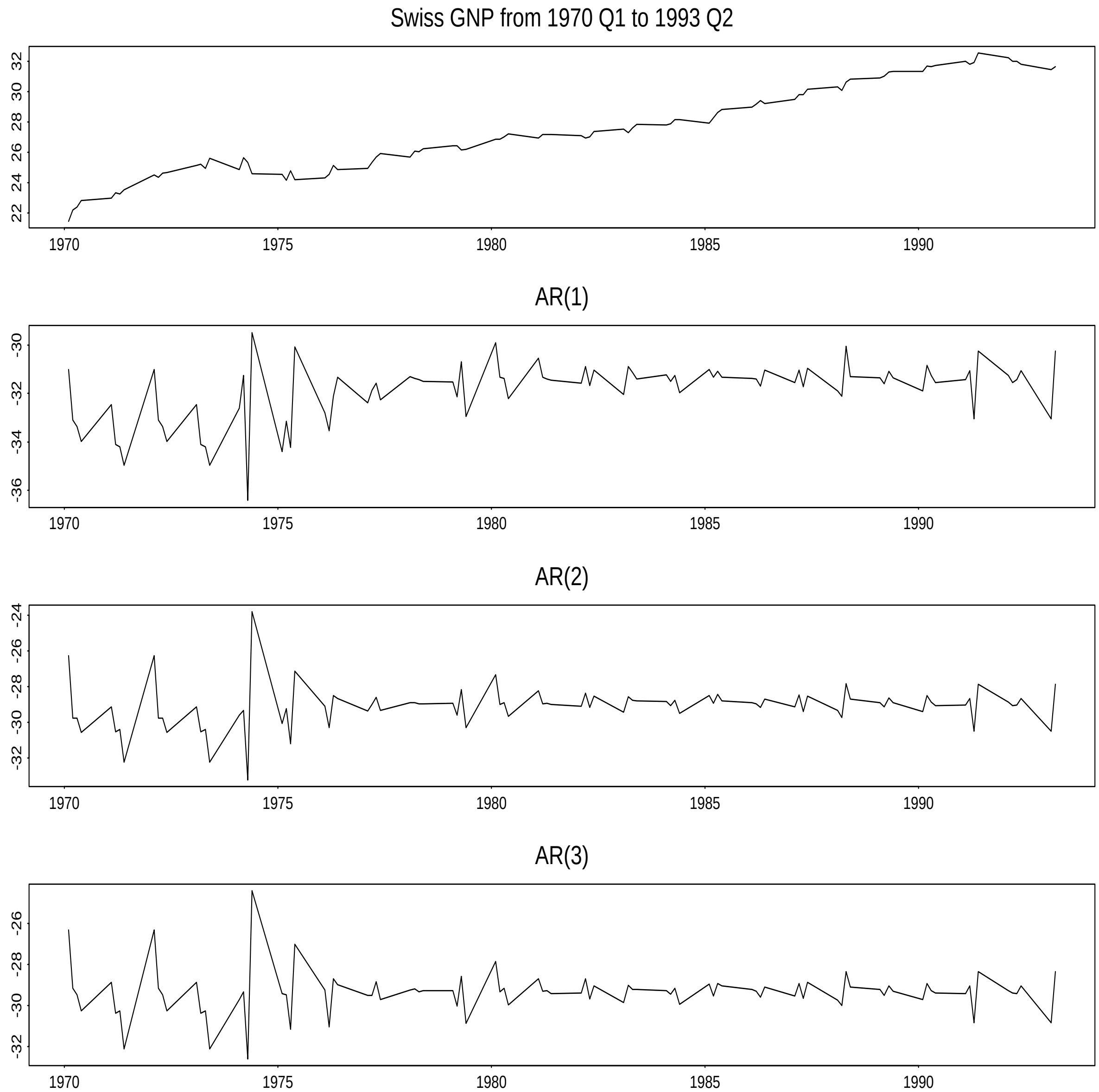

Figure 4: The log marginal likelihood (14) of the Swiss GNP from 1970 Q1 to 1993 Q2 for AR(p) break point model with outlier, equal variance and non-informative prior

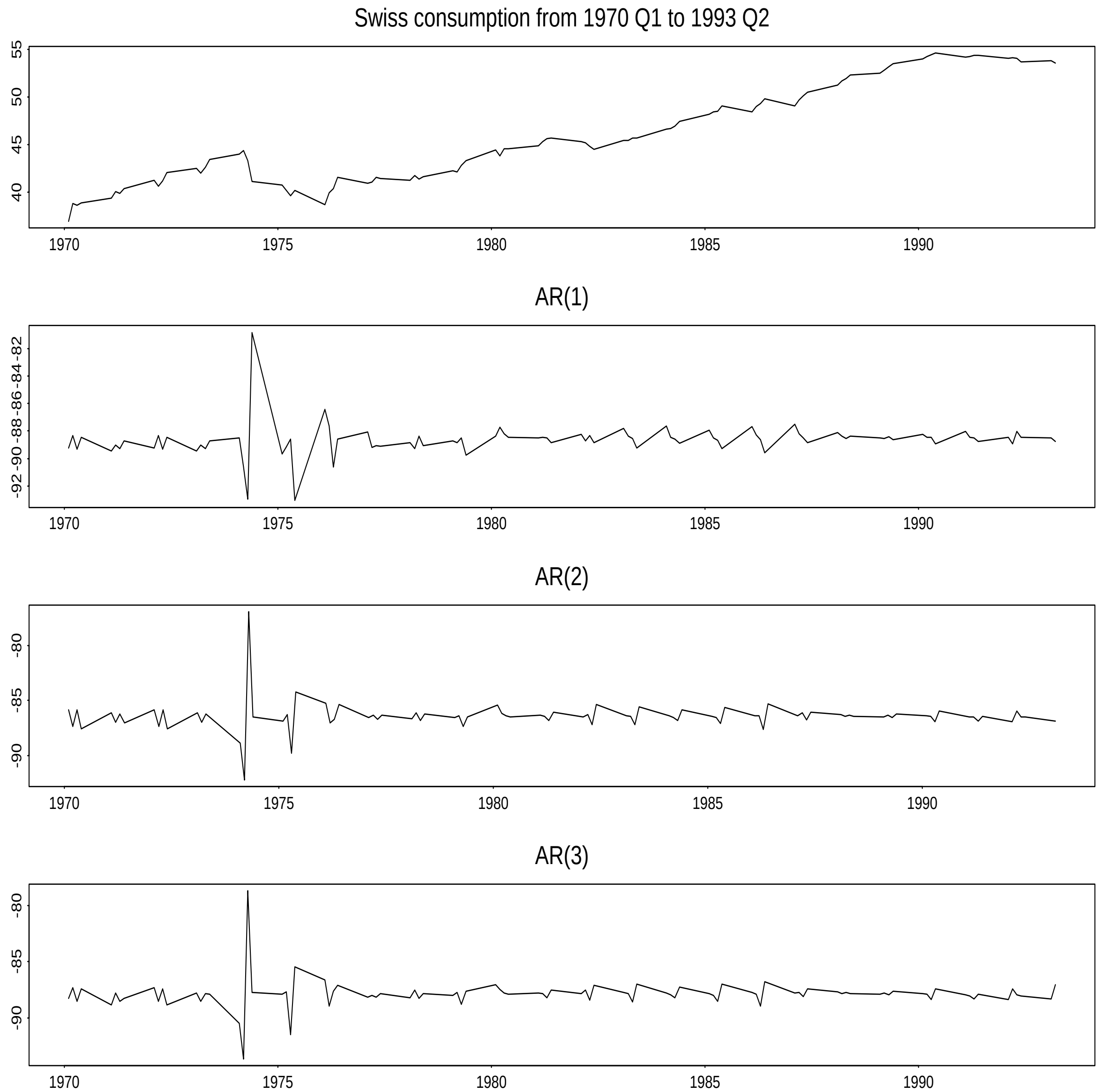

Figure 5: The log marginal likelihood (14) of the Swiss consumption from 1970 Q1 to 1993 Q2 for AR(p) break point model with outlier, equal variance and non-informative prior

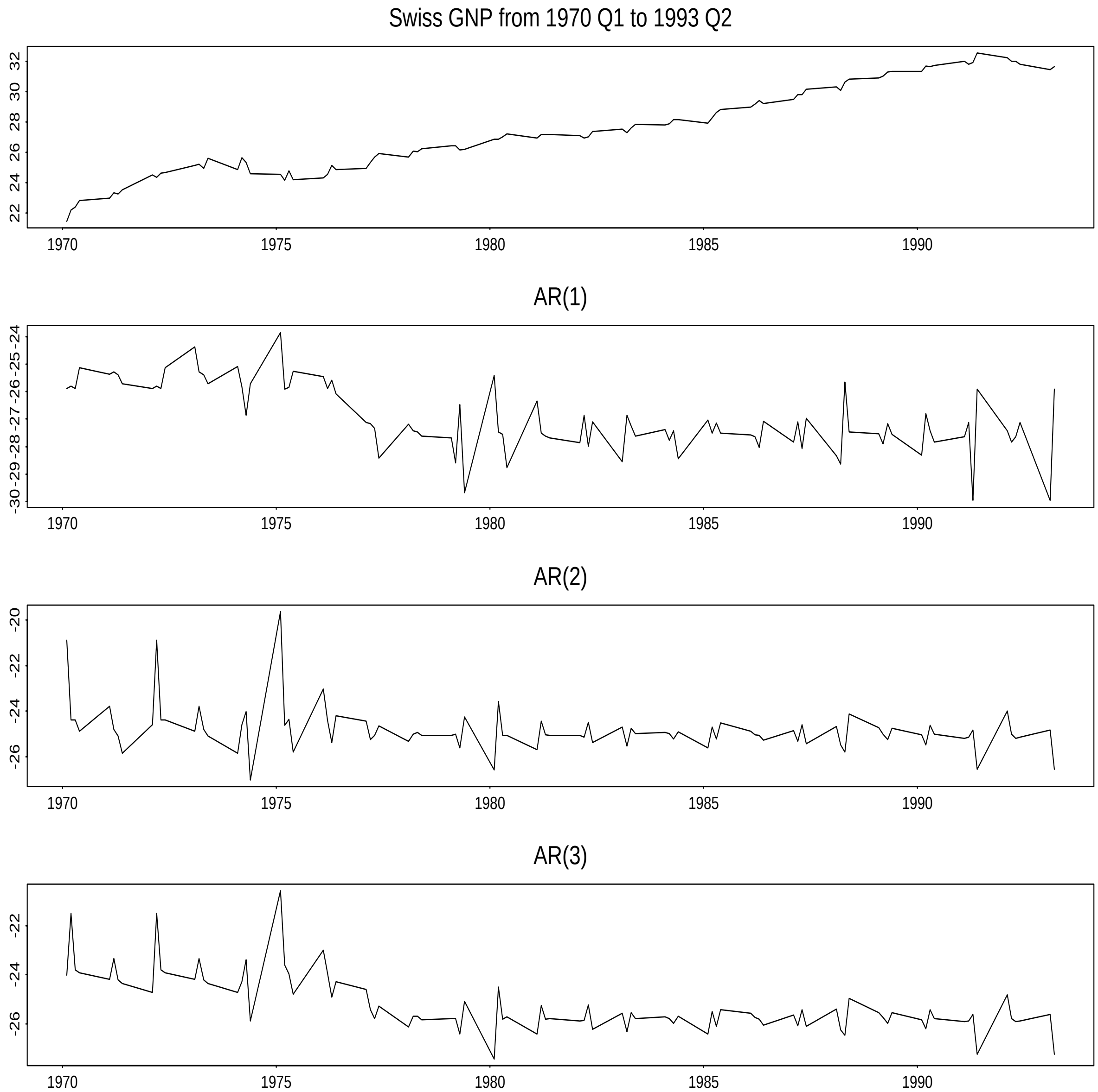

Figure 6: The log marginal likelihood (21) of the Swiss GNP from 1970 Q1 to 1993 Q2 for AR(p) break point model with outlier, unequal variance and non-informative prior

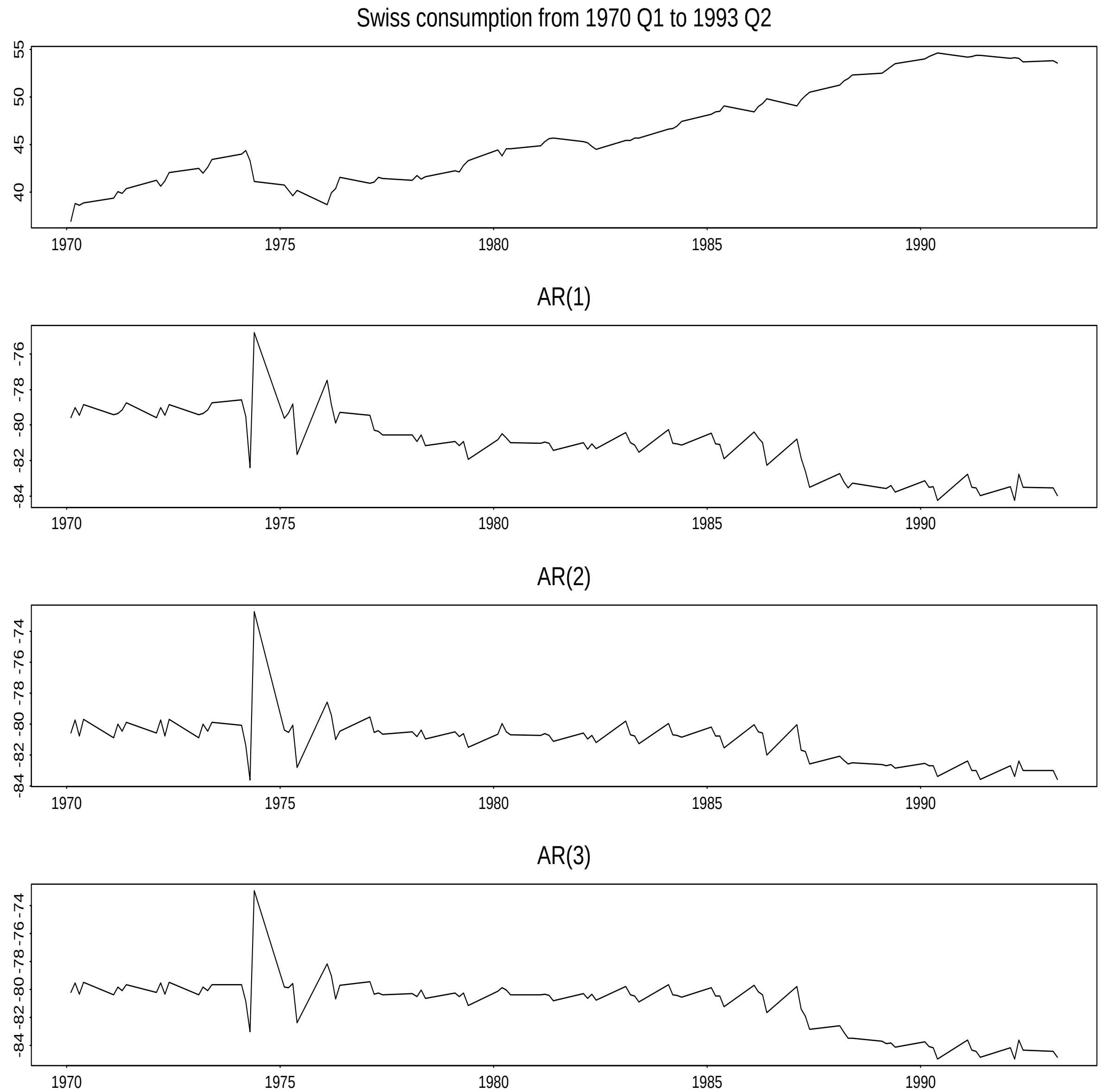

Figure 7: The log marginal likelihood (21) of the Swiss consumption from 1970 Q1 to 1993 Q2 for AR(p) break point model with outlier, unequal variance and non-informative prior