Statistical Papers manuscript

Detection of a Change-point in Student-*t* Linear Regression Models

Felipe Osorio, Manuel Galea[1] [⋆]

Departamento de Estadística, Universidad de Valparaíso, Chile.

Received: June 5, 2003; revised version: June 14, 2004

Abstract In this work the Schwarz Information Criterion (SIC) is used in order to locate a change-point in linear regression models with independent errors distributed according to the Student-*t* distribution. The methodology is applied to data sets from the financial area.

Key words: Change-point, Student-t model, regression model, Schwarz information criterion.

1 Introduction

The problem of change-points has been a topic of permanent interest in Statistical literature. Chernoff and Zacks (1964), Gadner (1969), Hawkins (1992), Sen and Srivastava (1975) and Worsley (1979) have studied the problem of change points in the mean of a normal distribution. Horváth (1993) and Chen and Gupta (1995) studied the problem of simultaneous change-points in the mean and variance, also of a normal distribution. For a review and many further results along these lines we refer to Csörgő and Horváth (1997).

The corresponding problem of change-points in the coefficients of a linear regression model has also been analyzed, under the assumption of normality, by several authors. Quant (1958, 1960) discusses estimates and hypotheses tests for a regression model in two phases. Brown, Durbin and Evans (1975) use recursive residuals to detect change-points in regression models. Hawkins (1989) uses the union-intersection principle and Kim (1994) the likelihood ratio test to detect change-points. Csörgő and Horváth (1997)

⋆ *Address:* Departamento de Estadística, Universidad de Valparaíso, Casilla 5030, Valparaíso, Chile. E-mail: `manuel.galea@uv.cl`

discuss some asymptotic results about this topic. Recently, Chen (1998) proposes the use of the Schwarz information criterion, SIC, to detect a change-point in linear regression models, also under the assumption of normality. See also Chen and Gupta (2001).

Let us consider a sequence of observations, $(\mathbf{x}_1^T, Y_1), (\mathbf{x}_2^T, Y_2), \dots, (\mathbf{x}_n^T, Y_n)$. The objective of this paper is to test hypothesis of the form:

$$H_0 : Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, 2, \dots, n, \tag{1}$$

i.e, regression coefficients does not change, against

$$\begin{aligned} H_1 : Y_i &= \mathbf{x}_i^T \boldsymbol{\beta}_1 + \epsilon_i, \quad i = 1, \dots, k, \\ Y_i &= \mathbf{x}_i^T \boldsymbol{\beta}_2 + \epsilon_i, \quad i = k+1, \dots, n. \end{aligned} \tag{2}$$

where,

$$\boldsymbol{\beta}_1 = (\beta_0, \beta_1, \dots, \beta_{p-1})^T, \quad \boldsymbol{\beta}_2 = (\beta_0^*, \beta_1^*, \dots, \beta_{p-1}^*)^T,$$

that is, a change exists (in the regression coefficients) in an unknown position k, denominated *change-point*.

In this work Chen's (1998) results are extended to the independent Student-t linear regression model. The use of the t-distribution as an alternative to the normal distribution, has frequently been suggested in literature, for example, Lange, Little and Taylor (1989) propose the t-distribution for robust modeling in linear regression models. Firstly, the methodology described by Chen (1998) is presented. Later on, the procedure of detection of change-points is presented for independent Student-t linear regression models. Finally, this methodology is applied to a group of data from the literature and to data from Chilean stock market. A comparison with the normal model is carried out in both cases.

2 Detection of a change-point in Normal linear regression models

Consider the linear regression model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \dots, n,$$

where $\mathbf{x}_i$, $i = 1, \dots, n$, corresponds to the i-th line of the design, $\mathbf{X}$ matrix $n \times p$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$ is the vector of unknown parameters, and ϵ_i indicates a random error. In this section we suppose that the errors ϵ_i are independent random variables, each one with a distribution $N(0, \sigma^2)$, where σ^2 is an unknown parameter ($\sigma^2 > 0$). In this way, we have that responses, Y_i, $i = 1, \dots, n$, are independent random variables distributed as $N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$.

For the change-point hypothesis (equations (1) and (2)) consider the following notation, where $k = p, \dots, n-p$,

$$\boldsymbol{Y}_1 = (Y_1, Y_2, \dots, Y_k)^T, \qquad \boldsymbol{Y}_2 = (Y_{k+1}, \dots, Y_n)^T,$$

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_k^T \end{pmatrix}, \qquad \mathbf{X}_2 = \begin{pmatrix} \mathbf{x}_{k+1}^T \\ \mathbf{x}_{k+2}^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix}.$$

The work by Chen (1998) proposes transforming the process of hypothesis testing in a procedure of model selection using the Schwarz Information Criterion, (SIC) defined by:

$$SIC = -2\,L(\widehat{\boldsymbol{\theta}}) + s\log n,$$

where $L(\widehat{\boldsymbol{\theta}})$ corresponds to the log-likelihood function evaluated on the maximum likelihood estimate of the parameters, s is the number of model parameters and n is the sample size. Note that maximizing the log-likelihood function is equivalent to minimizing the Schwarz information criterion.

Under H_0, there is a model that does not present any change in the regression coefficients; on the other hand, under H_1 there is a collection of models with change-points at the positions p or $p+1$ or $\dots$ or $n-p$. The objective is, therefore, to select a model from the previous group.

The maximum likelihood estimates for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2)^T$, under H_0, are given by

$$\boldsymbol{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y}, \qquad \widehat{\sigma}^2 = \frac{1}{n}\,(\boldsymbol{Y} - \mathbf{X}\boldsymbol{b})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{b}).$$

The Schwarz information criterion under H_0, denoted by $SIC(n)$, is given for:

$$\begin{aligned} SIC(n) &= -2\,L_0(\widehat{\boldsymbol{\theta}}) + (p+1)\log n \\ &= n\,\log Q(\boldsymbol{b}) + n(\log 2\pi + 1) + (p+1-n)\log n, \end{aligned}$$

where $L_0(\widehat{\boldsymbol{\theta}})$ corresponds to the maximum of the log-likelihood function under H_0 and $Q(\boldsymbol{b}) = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{b})^T(\boldsymbol{Y} - \mathbf{X}\boldsymbol{b})$.

Consider the model under the alternative hypothesis, i.e, that with a change-point at the position k, ($k = p, \dots, n-p$). In this case, $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \sigma^2)^T$, and the maximum likelihood estimates turn out to be

$$\begin{aligned} \boldsymbol{b}_1 &= (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\boldsymbol{Y}, \qquad \boldsymbol{b}_2 = (\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T\boldsymbol{Y}, \\ \widehat{\sigma}^2 &= \frac{1}{n}\,\{(\boldsymbol{Y}_1 - \mathbf{X}_1\boldsymbol{b}_1)^T(\boldsymbol{Y}_1 - \mathbf{X}_1\boldsymbol{b}_1) + (\boldsymbol{Y}_2 - \mathbf{X}_2\boldsymbol{b}_2)^T(\boldsymbol{Y}_2 - \mathbf{X}_2\boldsymbol{b}_2)\}. \end{aligned}$$

then the Schwarz information criterion under H_1, denoted by $SIC(k)$, for $k = p, \dots, n-p$, is

$$\begin{aligned} SIC(k) &= -2\,L_k(\widehat{\boldsymbol{\theta}}) + (2p+1)\log n \\ &= n\log\{Q(\boldsymbol{b}_1) + Q(\boldsymbol{b}_2)\} + n(\log 2\pi + 1) + (2p+1-n)\log n, \end{aligned}$$

where $L_k(\widehat{\boldsymbol{\theta}})$ corresponds to the maximum of the log-likelihood function under H_1.

Selection Criteria

The selection criteria is to choose a model with a change-point in the k position, if for some k

$$SIC(n) > SIC(k).$$

When the null hypothesis is rejected, the maximum likelihood estimate of the change-point in the regression coefficients, denoted by $\widehat{k}$, must satisfy,

$$\begin{aligned} SIC(\widehat{k}) &= \min\{SIC(k) : p \leq k \leq n-p\}, \\ &= \max\{L_k(\boldsymbol{\theta}) : p \leq k \leq n-p\}. \end{aligned}$$

3 Detection of a change-point in independent Student-t linear regression models

Consider the independent Student-t linear regression model, where the errors $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are independent and identically random variables with $t(0, \phi, \nu)$ distribution. Next, the maximum likelihood estimates under both hypotheses are described: one without changes in the regression coefficients (H_0) and the other with a change in a given position k. The EM algorithm is used to estimate the parameters due to its simple implementation.

In the first place the estimates are presented under H_0. In this case the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ is

$$L_0(\boldsymbol{\theta}) = n \log K(\nu) - \frac{n}{2} \log \phi - \frac{\nu+1}{2} \sum_{i=1}^{n} \log\{1 + d_i^2(\boldsymbol{\beta}, \phi)/\nu\},$$

where

$$\begin{aligned} d_i^2 &= d_i^2(\boldsymbol{\beta}, \phi) = \frac{(Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2}{\phi}, \quad i = 1, \dots, n, \\ K(\nu) &= \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\ \Gamma(\frac{\nu}{2})} \end{aligned}$$

The score functions are given by

$$\begin{aligned} \boldsymbol{U}(\boldsymbol{\beta}) &= \frac{1}{\phi} \sum_{i=1}^{n} v_i (Y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \mathbf{x}_i = \frac{1}{\phi} \mathbf{X}^T \mathbf{V} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}), \\ \boldsymbol{U}(\phi) &= -\frac{n}{2\phi} + \frac{1}{2\phi^2} \sum_{i=1}^{n} v_i (Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 = -\frac{n}{2\phi} + \frac{1}{2\phi^2} Q_{\mathbf{V}}(\boldsymbol{\beta}), \end{aligned}$$

where $\mathbf{V} = diag(v_1, v_2, \dots, v_n)$, with

$$v_i = v_i(\boldsymbol{\theta}) = \frac{\nu+1}{\nu + d_i^2}, \quad i = 1, \dots, n,$$

$$Q_{\mathbf{V}}(\boldsymbol{\beta}) = (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}).$$

The likelihood equations correspond to a nonlinear system of equations which should be solved via iterative methods.

In this work we suppose that the degrees of freedom ν is known. Fernández and Steel (1999), alert on the estimate ν and they notice that in this case the function of log-likelihood is unbounded and that indeed it corresponds to an nonregular estimation problem. Due to this, it is suggested (see Lange, Little and Taylor, 1989) to estimate $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$ considering a set of acceptable values for ν and to choose the one that maximizes the function of log-likelihood. In our case, the Schwarz information criterion will be used to choose among some values of ν.

The estimation using the EM algorithm will now be considered. The observed log-likelihood function under the model without changes, turns out to be

$$L_0(\boldsymbol{Y}|\boldsymbol{v};\boldsymbol{\theta}) = -\frac{n}{2}\,\log 2\pi\phi + \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2\phi}\,Q_{\mathbf{V}}(\boldsymbol{\beta}).$$

The EM algorithm maximizes the log-likelihood iteratively through two stages. The $(r+1)$-th step of the algorithm is summarized as follows:

E Step: Starting from initial estimates for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \phi)^T$, weight estimates $v_i^{(r)}$ are obtained through the conditional expectations

$$E(U_i|Y_i;\boldsymbol{\theta}^{(r)}) = v_i^{(r)} = \frac{\nu+1}{\nu + d_i^{2(r)}(\boldsymbol{\beta}^{(r)}, \phi^{(r)})},$$

with

$$d_i^{2\,(r)} = \frac{(Y_i - \mathbf{x}_i^T\boldsymbol{\beta}^{(r)})^2}{\phi^{(r)}}, \quad i = 1, 2, \ldots, n.$$

and where the independent random variables, $U_i \sim Gamma((\nu+1)/2, (\nu + d_i^2)/2)$.

M Step: Using the weights obtained at the E step of the algorithm, the maximum likelihood estimates are obtained,

$$\boldsymbol{\beta}^{(r+1)} = (\mathbf{X}^T\mathbf{V}^{(r)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{(r)}\boldsymbol{Y}, \quad \phi^{(r+1)} = \frac{1}{n}\,Q_{\mathbf{V}^{(r)}}(\boldsymbol{\beta}^{(r)}),$$

with

$$\mathbf{V}^{(r)} = diag(v_1^{(r)}, v_2^{(r)}, \ldots, v_n^{(r)})$$

and the algorithm proceeds between E and M step until the sequence $\boldsymbol{\theta}^{(r)}$ converge.

Under a change model on the regression coefficients in a given k position, it is possible to show that the log-likelihood function turns out to be

$$\begin{aligned} L_k(\boldsymbol{\theta}) \;&=\; n\log K(\nu) - \frac{n}{2}\,\log\phi - \frac{\nu+1}{2}\sum_{i=1}^{k}\log\{1 + d_i^2(\boldsymbol{\beta}_1,\phi)/\nu\} \\ &\quad - \frac{\nu+1}{2}\sum_{i=k+1}^{n}\log\{1 + d_i^2(\boldsymbol{\beta}_2,\phi)/\nu\}, \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \phi)^T$ and the associated score functions are given by:

$$\boldsymbol{U}(\boldsymbol{\beta}_1) = \frac{1}{\phi}\ \mathbf{X}_1^T \mathbf{V}_1(\boldsymbol{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}_1), \qquad \boldsymbol{U}(\boldsymbol{\beta}_2) = \frac{1}{\phi}\ \mathbf{X}_2^T \mathbf{V}_2(\boldsymbol{Y}_2 - \mathbf{X}_2\boldsymbol{\beta}_2),$$

$$\boldsymbol{U}(\phi) = -\frac{n}{2\phi} + \frac{1}{2\phi^2}\ \{Q_{\mathbf{V}_1}(\boldsymbol{\beta_1}) + Q_{\mathbf{V}_2}(\boldsymbol{\beta_2})\},$$

with

$$\mathbf{V}_1 = diag(v_1, v_2, \dots, v_k), \qquad \mathbf{V}_2 = diag(v_{k+1}, \dots, v_n).$$

To develop the estimate procedure using the EM algorithm, the observed log-likelihood function under the model with a change in a k position (given) takes the form:

$$\begin{aligned} L_k(\boldsymbol{Y}|\boldsymbol{v};\boldsymbol{\theta}) \ = \ & -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\phi + \frac{1}{2}\log|\mathbf{V}_1| - \frac{1}{2\phi}\ Q_{\mathbf{V}_1}(\boldsymbol{\beta}_1) \\ & + \frac{1}{2}\log|\mathbf{V}_2| - \frac{1}{2\phi}\ Q_{\mathbf{V}_2}(\boldsymbol{\beta}_2) \end{aligned}$$

The $(r+1)$-th iteration of the algorithm of expectation maximize consists in two steps, described as follows:

E Step: The conditional expectations $v_i^{(r)}$ should be obtained, this are given by

$$E(U_i|Y_i;\boldsymbol{\theta}^{(r)}) = v_i^{(r)} = \frac{\nu+1}{\nu + d_i^2(\boldsymbol{\theta}^{(r)})},$$

which are based on the following weights function,

$$d_i^2(\boldsymbol{\theta}^{(r)}) = \begin{cases} \dfrac{(Y_i - \mathbf{x}_i^T\boldsymbol{\beta}_1^{(r)})^2}{\phi^{(r)}} &, \ i = 1, 2, \dots, k, \\ \dfrac{(Y_i - \mathbf{x}_i^T\boldsymbol{\beta}_2^{(r)})^2}{\phi^{(r)}} &, \ i = k+1, \dots, n. \end{cases}$$

where the independent random variables U_i follow a Gamma distribution with parameters $(\nu+1)/2$ and $(\nu + d_i^2)/2$.

M Step: Using the estimates obtained at the E step, likelihood estimates are calculated by:

$$\begin{aligned} \boldsymbol{\beta}_1^{(r+1)} \ &= \ (\mathbf{X}_1^T\mathbf{V}_1^{(r)}\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{V}_1^{(r)}\boldsymbol{Y}_1, \\ \boldsymbol{\beta}_2^{(r+1)} \ &= \ (\mathbf{X}_2^T\mathbf{V}_2^{(r)}\mathbf{X}_2)^{-1}\mathbf{X}_2^T\mathbf{V}_2^{(r)}\boldsymbol{Y}_2, \\ \phi^{(r+1)} \ &= \ \frac{1}{n}\Big\{Q_{\mathbf{V}_1^{(r)}}(\boldsymbol{\beta}_1^{(r)}) + Q_{\mathbf{V}_2^{(r)}}(\boldsymbol{\beta}_2^{(r)})\Big\}. \end{aligned}$$

with

$$\mathbf{V}_1^{(r)} = diag(v_1^{(r)}, v_2^{(r)}, \dots, v_k^{(r)}), \qquad \mathbf{V}_2^{(r)} = diag(v_{k+1}^{(r)}, \dots, v_n^{(r)}).$$

The implementation of the EM algorithm under H_1, requires a slight modification of the iteratively reweighted least square, because the scale parameter ϕ remains *fixed* along the sequence of observations.

Now the Schwarz information criterion is presented under the model without changes in the regression coefficients and under the model with a change in the k given position. Under H_0, we have that

$$\begin{aligned} SIC(n) &= -2\,L_0(\widehat{\boldsymbol{\theta}}) + (p+1)\log n \\ &= -2\,n\log K(\nu) + n\,\log\widehat{\phi} + (p+1)\log n \\ &\quad +(\nu+1)\sum_{i=1}^{n}\log\{1+d_i^2(\widehat{\boldsymbol{\beta}},\widehat{\phi})/\nu\} \end{aligned} \tag{3}$$

and under H_1 (for given k), the Schwarz information criterion is given by

$$\begin{aligned} SIC(k) &= -2\,L_k(\widehat{\boldsymbol{\theta}}) + (2p+1)\log n \\ &= -2\,n\log K(\nu) + n\,\log\widehat{\phi} + (2p+1)\log n \\ &\quad +(\nu+1)\Big[\sum_{i=1}^{k}\log\{1+d_i^2(\widehat{\boldsymbol{\beta}}_1,\widehat{\phi})/\nu\} + \sum_{i=k+1}^{n}\log\{1+d_i^2(\widehat{\boldsymbol{\beta}}_2,\widehat{\phi})/\nu\}\Big]. \end{aligned} \tag{4}$$

Note that when $\nu\to\infty$ the expressions corresponding to the normal case (Chen, 1998) are obtained.

Selection Criteria

Using the Schwarz information criterion (equations (3) and (4)) the hypothesis of no change on the regression coefficients H_0 is rejected if:

$$SIC(n) > \min\{SIC(k) : p \le k \le n-p\}.$$

Or equivalently if, $\triangle_n < 0$, where

$$\triangle_n = \min\{SIC(k) - SIC(n) : p \le k \le n-p\}.$$

If H_1 is accepted, i.e, there is a change-point at the regression coefficients, then the change position is estimated via maximum likelihood, in a way that $\widehat{k}$ has to satisfy.

$$SIC(\widehat{k}) = \min\{SIC(k) : p \le k \le n-p\}.$$

To make conclusions about change-points statistically significant, the level of significance α and its associated critical value c_α ($c_\alpha \ge 0$) are introduced. So instead of rejecting H_0 when $\triangle_n < 0$, the hipothesis of no change on the regression coefficients is rejected if

$$\triangle_n + c_\alpha < 0 \tag{5}$$

where c_α satisfies the relationship:

$$1-\alpha = P[\triangle_n + c_\alpha < 0\,|H_0] = P[Z_n < p\log n + c_\alpha\,|H_0]$$

where $Z_n = \max\{-2(L_0(\widehat{\boldsymbol{\theta}}) - L_k(\widehat{\boldsymbol{\theta}})) : p \leq k \leq n-p\}$ and using the theorem 1.3.1 in Csörgő and Horváth (1997), we have that

$$c_\alpha \approx \frac{1}{a^2(\log n)}\{d_p(\log n) - \log\log[1-\alpha+\exp(-2e^{d_p(\log n)})]^{-1/2}\}^2 - p\log n \tag{6}$$

with

$$a(x) = (2\log x)^{1/2}, \quad d_p(x) = 2\log x + \frac{p}{2}\log\log x - \log\Gamma\left(\frac{p}{2}\right)$$

Therefore we propose to use (6) to calculate approximate c_α values, for a level of significance α.

If it is suspected that a change-point exists in the k position, when carrying out the test of change by means of Z_n, there are $n-2p$ tests being done out, one for each k. A useful technique for finding approximates critical values is based on the inequality of Bonferroni. Using this method one has the following criteria, to reject H_0 at a nominal α level if

$$Z_n > \chi^2_{1-\alpha/(n-2p)}(p).$$

For the normal case, LRT has an exact distribution, in which case, H_0 is rejected if $Z_n > F_{1-\alpha/(n-2p)}(p, n-2p)$.

Some Considerations

It is evident that implementing a routine for change-points detection by means of the SIC procedure is quite simple; it is also possible to observe that the methodology of detection for changes in the regression coefficients considered here is equivalent to the procedure for testing a change-point using the Likelihood ratio test given in Csörgő and Horváth (1997). However, they notice that the region of rejection of the hypothesis is conservative and indeed large samples should be used in (6) to test H_0 against H_1.

The methodology for detection of change-points delineated here corresponds to a test of a *unique* change. Chen and Gupta (1997) mention a method proposed by Vostrikova (1981) who outlines a procedure of binary segmentation, in which the problem of detection of *multiple* changes is subdivided in to equivalent simple detection processes.

The estimation of parameters for the problem of detection of changes, in the regression coefficients and the approach of model selection conforms the base of an S-Plus routine, created for developing these calculations.

4 Applications

In this section, the methodology derived in this work is applied to data frequently analyzed in the literature (see Chen (1998) and Chang and Huang (1997)). Data of returns from a share belonging the Chilean stock market is also considered.

4.1 Holbert's data (Chen, 1998)

Holbert (1982) use a simple linear regression model with changes in the coefficients from a Bayesian point of view. Later on, Chen (1998) analyzed this data using the SIC method to detect changes in the mean for Normal linear regression models. In this section, Holbert data will be used to illustrate the SIC procedure, for the detection of change-points in independent Student-t linear regression models.

Sales volume (in millions) of the Boston stock market (BSE) is considered as a response variable and the sales volume of the New York stock market (NYAMSE) as a regressive variable. The data corresponds to the period between January 1967 and November 1969.

Figure 1 shows the scatter plot of Holbert's data.

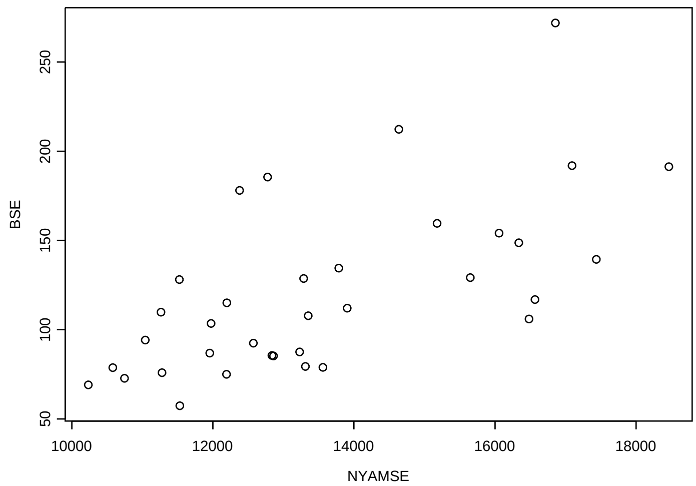

Fig. 1 Holbert's data scatter plot

The SIC procedure was applied for the detection of changes, for the Student-t linear regression model with a set of values for the degrees of freedom, as well as for the normal model. The summary of results is in the following table.

Table 1 Results of SIC Procedure

t Model	$SIC(n)$	$\min SIC(k)$	$\widehat{k}$
$\nu = 1$	363.358	357.416	9
$\nu = 4$	358.082	355.035	23
$\nu = 8$	359.332	356.669	23
$\nu = 30$	360.855	358.074	23
$\nu \to \infty$	361.453	358.475	23

A possible change-point is detected in the period $\widehat{k} = 23$, and this result is in concordance with the conclusion obtained by Chen (1998).

Using the Schwarz Information Criterion, the Student-t model with $\nu = 4$ was chosen to develop some later analysis. The parameter estimates for the model with a change-point in position 23 is presented. The results for the normal model are also presented.

There is an appreciable difference on the coefficients estimates for the Student-t model with $\nu = 4$ degrees of freedom and the normal model; in particular, the difference among the intercepts of both models is evident. However, the scatter plot with fitted regression lines for the model with a change-point in position 23, do not reveal a remarkable difference between the normal model and the Student-t model with $\nu = 4$ (see Table 2 and Figure 2).

Table 2 Estimates for the model with a change in the 23rd position.

t Model	Cases	Estimate	
		β_0	β_1
$\nu = 4$	1 to 23	-96.0781	0.0162
	24 to 35	15.2759	0.0065
$\nu \to \infty$	1 to 23	-110.5475	0.0178
	24 to 35	11.0747	0.0067

In accordance with the above-mentioned, the condition (5) was verified, obtaining that detected changes do not turn out to be significant. The analysis by means of the Bonferroni method reveals that the Student-t model for Holbert's data does not detect changes in the coefficients. Note however, that the sample size for these data is small, so that this decision can be conservative.

$\nu = 4$ $\nu \to \infty$

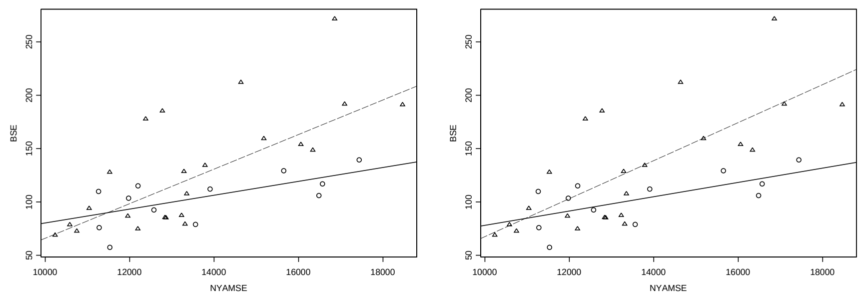

Fig. 2 Scatter plots and adjust regression

Observations in $\triangle$, (- - -) correspond to previous periods to observation 24 (from where a change was detected in the sequence). Notice that, in spite of perceiving a graphic difference in the fitted regression lines, these changes are not significant.

$\nu = 4$ $\nu \to \infty$

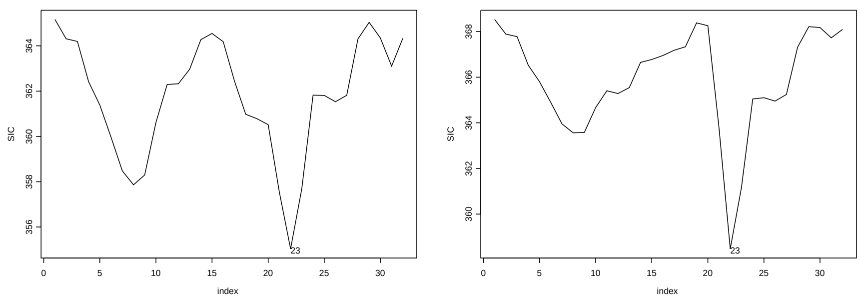

Fig. 3 Diagram $SIC(k)$ vs. k

The graph of the Schwarz information criterion against the index has been suggested, i.e, $SIC(k)$ vs. k, because it is possible to obtain the maximum likelihood estimate of k from this graph. In both diagrams the 23rd position may be as a possible maximum likelihood estimate for the change-point. Note, however, that for the normal model, the $SIC(k)$ values are bigger than these for the Student-t model with $\nu = 4$ and, therefore, it is suggested to use this last one as an alternative to the normal distribution.

In Osorio (2001) the vulnerability of the normal model to atypical observations is shown, indeed, some diagrams of local influence are considered indicating that for the t with $\nu = 4$ this influence is smaller.

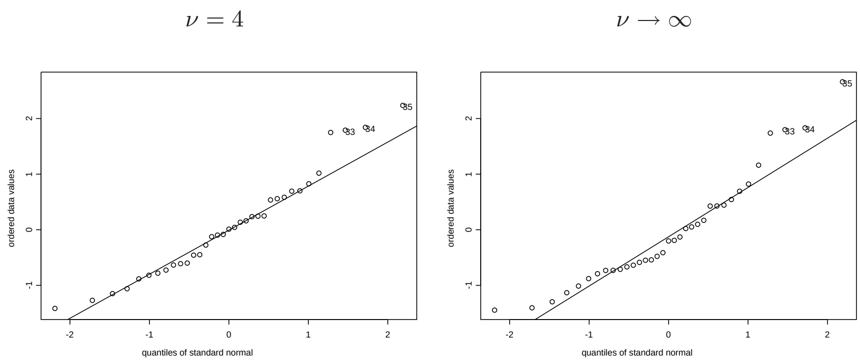

Fig. 4 Quantile plots

From the quantile plots a slight deviation of the supposed distributional is appreciated for the t with $\nu = 4$ as for the Normal model ($\nu \to \infty$), therefore, the quantile plots do not offer conclusive information in this respect.

4.2 Detection of a change in the systematic risk

The Capital Asset Pricing Model (CAPM) establishes that the expected return on an asset is equal to the risk free rate return plus a prize for risk. This model was independently derived by Sharpe (1964), Lintner (1965), and Mossin (1966). Let r be a random variable, which denotes the asset return. According to the CAPM, the expected value of r, is given by:

$$E(r) = r_f + \beta(E(r_m) - r_f), \tag{7}$$

where r_f is the risk free rate return, r_m corresponds to the market return and β denotes the systematic risk of the asset. An extension of the model (7) given by

$$E(r) - r_f = \alpha + \beta(E(r_m) - r_f), \tag{8}$$

incorporates a coefficient α that denotes the asset return independent to the market fluctuations. It is usual to consider a linear regression model to estimate α and β, this is, given a group of n observations for the return of the asset, the market return and the risk free return, one has

$$r_t - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \epsilon_t, \quad t = 1, \dots, n, \tag{9}$$

where r_t denotes the return of the asset in the period t, r_{mt} is the market return in the period t, $t = 1, \dots, n$. Here we suppose that $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are independent random errors such that $\epsilon_t \sim t(0, \phi, \nu)$, $t = 1, \dots, n$.

The main interest in the CAPM model is to carry out inferences regarding the systematic risk β. To estimate the beta parameter from (9), the least squares method is frequently used. Also to carry out inferences regarding the parameters, we generally suppose that the returns follow a normal distribution. It is important to note that the estimate of the systematic risk on Latin American markets (emergent) has some peculiarities, such as atypical observations and points leverages. In this respect there are robust estimates for β (see Duarte and Mendes (1997)); however, changes in the systematic risk have not been considered.

The CAPM was used to data corresponding to monthly returns, adjusted by equity variations, of "Vineyards Concha y Toro". IPSA, was used as the return for the market and as risk free rate was used the interest rate in the sale of discounted bonds of the Central Bank (PDBC) based monthly. The data corresponds to the period among the months of March 1990 to April 1999. Figure 5 shows the scatter plot for the data asset Vineyards Concha y Toro.

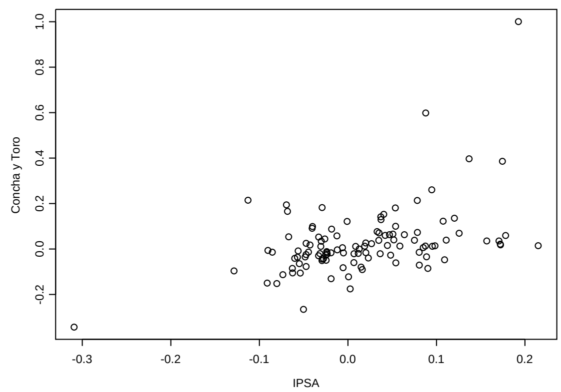

Fig. 5 Vineyard Concha y Toro data scatter plot

The procedure of detection of a point of change was developed considering a independent t linear regression model for a set of values for freedom degrees using SIC procedure. The results are as follows:

Table 3 Results of SIC Procedure.

t Model	$SIC(n)$	$\min SIC(k)$	$\hat{k}$
$\nu = 1$	-172.304	-175.924	100
$\nu = 4$	-172.136	-173.098	100
$\nu = 8$	-159.518	-157.776	(100)
$\nu = 30$	-135.362	-135.803	25
$\nu \to \infty$	-116.382	-123.829	25

Note that the possible periods detected as change-points for the normal model and for the t model, with $\nu = 1$ or 4, are different; also, for the t model with 8 degrees of freedom a change is not detected (the position among parenthesis corresponds to $\min SIC(k)$). However, the SIC for the Student-t model with $\nu = 1$ or 4 are *smaller* than for the normal model.

Using the minimum information approach, the t model was chosen with 1 degree of freedom. In what follows this model will be used to carry out later analysis. The estimate of parameters was developed for the model with a change in position k, obtaining the following,

Table 4 Estimate for the model with a change in position k.

t Model	Cases	Estimates	
		α	β
$\nu = 1$	1 to 100	-0.0069	0.3021
	101 to 110	-0.0111	1.1096
$\nu \to \infty$	1 to 25	0.0372	1.5995
	26 to 110	-0.0019	0.4799

As a result of some interest, to verify that for the normal distribution the change detected in position 25 is significant to 5%, on the other hand for the t-distribution with any one of the considered degrees of freedom, the changes are not significant, and therefore for these data, the Student-t model does not detect any change in the regression coefficients. This decision is ratified when carrying out the procedure of Bonferroni. Indeed, a change is only detected in the coefficients when it is considered the normal model.

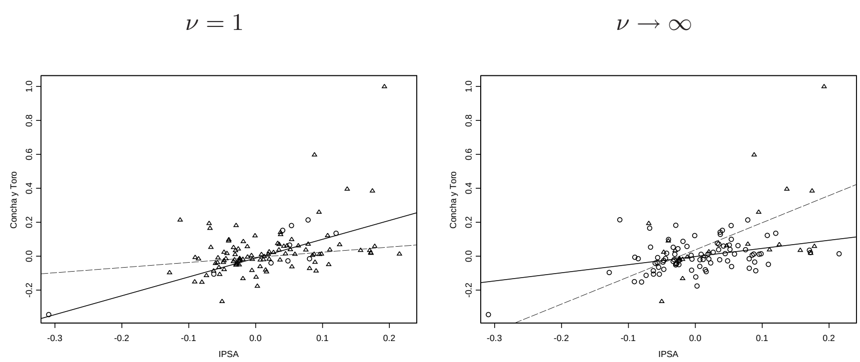

Fig. 6 Dispersion and straight line of Regression Diagram

The observations in $\triangle$, (- - -) correspond to the previous periods to the point where a change happens in the sequence. Although for the normal model a point of change is detected in the regression coefficients, as for the

t model this change is not significant. Consider the results of the estimate parameters for the model without changes that is presented in the following table:

Table 5 Estimate for the model without changes in the coefficients.

t Model	log-likelihood	$\widehat{\alpha}$	$\widehat{\beta}$
$\nu = 1$	93.2026	-0.0084	0.3462
$\nu \to \infty$	65.2417	0.0129	0.8884

Starting from this, the use the Student-t model with $\nu = 1$ is suggested, to estimate the systematic risk through the estimates for β in the model that does not present any change.

The diagram was also carried out $\widehat{\beta}_{(i)}$ vs. i (see Figure 7), where $\widehat{\beta}_{(i)}$ denotes the maximum likelihood estimator of β without the i-th observation. Here certain stability is noticed by the $\widehat{\beta}_{(i)}$, so that it is clear that the t model reduces the influence of atypical observations.

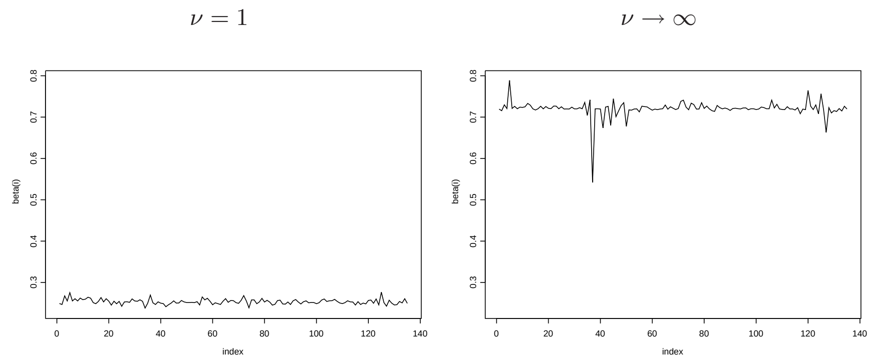

Fig. 7 $\widehat{\beta}_{(i)}$ vs. i plot

The diagram was also obtained $\widehat{\beta}_{\{k\}}$ vs. k where k indicates the number of observations included for the calculation of the estimate of β.

In spite of the natural fluctuations of this estimates when having few observations, note that this graph reveals the position of change detected by the normal model. However, a greater stability for the Student-t model with $\nu = 1$, is appreciated, by this we perceive a certain insensibility of the Student-t model in the face of changes in the coefficients in this case.

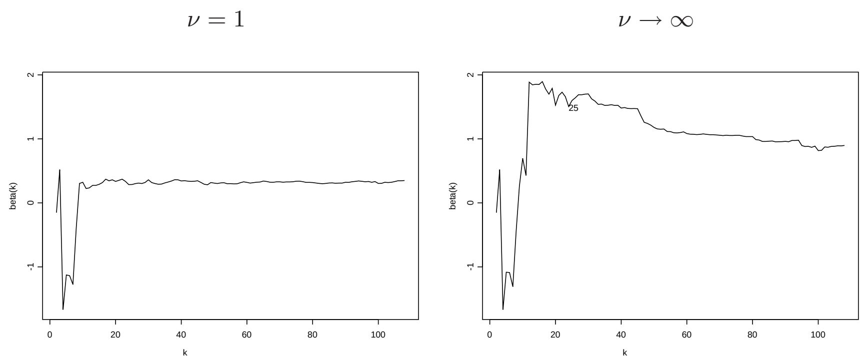

Fig. 8 Diagram $\widehat{\beta}_{\{k\}}$ vs. k

Also notice that the Student-t model with $\nu = 1$ has values $SIC(k)$ smaller than of the Normal model. Starting from the approach of minimum information the use of the Student-t distribution is suggested with $\nu = 1$ instead of the normal distribution for this data set.

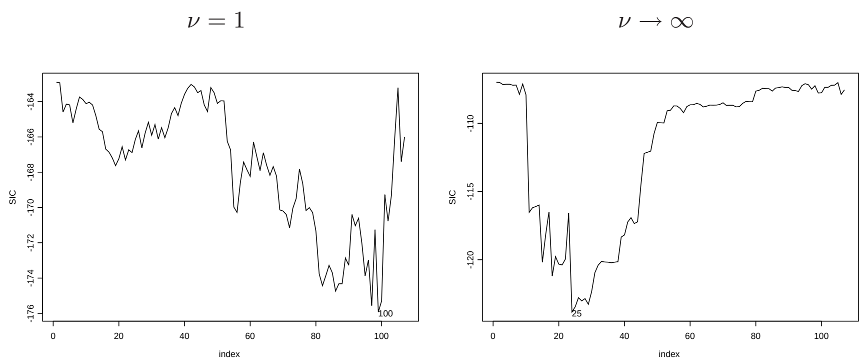

Fig. 9 $SIC(k)$ Diagram vs. k

In Osorio (2001) there are remarks on the great impact that the period of February 28 1991 (observation 12) has on the MLE for the normal model, on the contrary of the t model with $\nu = 1$ where the influence of the periods is smaller. Also, one can observe an evident deviation from the supposed distributional for the normal model (see Figure 10). Therefore, the above-mentioned suggests the use of distribution t with $\nu = 1$.

Consider tables 4 and 5, note that when using the t model with $\nu = 1$ the appreciation that one has of the systematic risk is different from that of the normal model. Indeed, when using the estimates of parameters obtained by means of the t with $\nu = 1$ (model does not present changes in the coefficients), the asset of Vineyards Concha y Toro is less *risky* than for the normal model. This information can be particularly useful for risk administrators or of stocks portfolios.

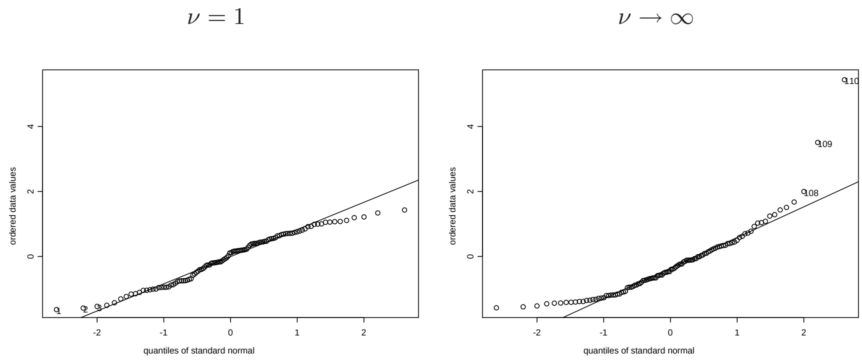

Fig. 10 Quantile Plots

Conclusion

It is evident that the detection of points of change through the SIC offers an important simplification, allowing a simple computer implementation.

It was also indicated that the approach of testing hypothesis proposed in this work is equivalent to the likelihood ratio test for locate change points, therefore an expression was presented for the calculation of approximate quantile values and this way testing hypothesis with a significant level. It is of interest to notice that this expression is valid for any distribution of the errors so that the decision can be conservative for small sample sizes.

Acknowledgements

The authors acknowledge the partial financial support from Project Fondecyt 1000424-Chile. The authors are grateful to the referees whose comments and suggestions were valuable to improve the exposition of the paper.

References

1. Chan Y. P. and Huang, W. T. (1997). Inferences for the linear errors-in-variables with changepoint models. *Journal of the American Statistical Association* **92**, 171-178.
2. Chen, J. (1998). Testing for a change point in linear regression models. *Communications in Statistics - Theory & Methods* **27**, 2481-2493.

3. Chen, J. and Gupta, A. K. (1995). Likelihood procedure for testing change points hypothesis for multivariate gaussian model. *Random Operators and Stochastic Equations* **3**, 235-244.
4. Chen, J. and Gupta, A. K. (1997). Testing and locating variance change point with application to stock prices. *Journal of the American Statistical Association* **92**, 739-747.
5. Chen, J. and Gupta, A. K. (2001). On change point detection and estimation. *Communications in Statistics - Simulation & Computation* **30**, 665-697.
6. Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, New York.
7. Duarte, A. M. and Mendes, B. V. M. (1998). Robust estimation of systematic risk in emerging stock markets. *Emerging Markets Quaterly* **1**, 85-95.
8. Fernández, C. and Steel, M. (1999). Multivariate student t regression models: pitfalls and inference. *Biometrika* **86**, 153-167.
9. Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics* **21**, 671-680.
10. Lange, K. L.; Little, J. A. and Taylor, M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* **84**, 881-896.
11. McLachlan, G. L. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.
12. McQuarrie, A. D. R. and Tsai, C. L. (1998). *Regression and Time Series Model Selection*. World Scientific Publishing, Singapore.
13. Osorio, F. (2001). *Detección de un punto de cambio en modelos de regresión lineal t independiente*, Tesis para optar al título de ingeniero en Estadística. Departamento de Estadística. Universidad de Valparaíso, Chile.