A Bayesian Algorithm for Detecting Multiple Changes of Variance in a Time Series

Enrique de Alba[*†] and Michelle Boue[*]
[*]ITAM and †The University of Waterloo

March 2000

Abstract

Changes in variance are frequently found in economic or financial time series. Detecting variation or sudden changes in these time series is central to understanding and proper interpretation of their behavior. In particular the topic of changes in the variance (volatility) of the series has been the subject of much research. In fact it has given rise to whole areas of research such as ARCH, stochastic volatility , switching regime models, and others. Research in this area has considered detecting multiple changes through the application of diverse procedures. The majority of Bayesian literature considers only the possibility of a single structural break. When more than one break point is considered it is usually done by focusing on the detection of points of change one at a time because of the heavy computational burden involved in looking for more than one simultaneously. In this paper we study the problem of detecting multiple changes of variance in a time series. We first consider a sequence of independent observations and later we apply the same procedure to an AR(1) series. The break points are estimated endogenously and sequentially. We propose a procedure to detect these variance changes based on a linear programming algorithm. The problem of finding the most likely set of break points is equivalently represented as that of finding the shortest path connecting two nodes on a suitable graph. An algorithm for solving this latter problem is presented. The method can be extended to detect changes either in its level, trend or variance of a time series, or structural breaks in econometric models. We provide examples.

1 Introduction

The study of variability is important in many areas, but in finance it is crucial because it is a measure of risk. Changes in variance are frequently found in economic or financial time series. It is important to understand variability and how it changes in order to price financial instruments and make sound financial decisions. Detecting variation or sudden changes in these time series is central to understanding and proper interpretation of their behavior. In particular the topic of changes in the variance (volatility) of the series has been the subject of much research. In fact it has given rise to whole areas of research such as ARCH, stochastic volatility, switching regime models, and others. In some cases, as in modeling the short-term interest rate, the main shortcoming of some existing models is their inability to incorporate changing volatility structures, Chan et. al. (1992). In order to solve this problem, Pai and Pedersen (1999) have proposed to use regime switching models, which allow for sporadic changes in variance. A similar approach is used by Hardy (1999) to model long term stock returns.

It appears that the statistical literature on changes of variance started with attempts to provide an alternative to the Pareto distribution to model stock returns, Hsu, Miller, and Wichern (1974). Some of the initial studies proposed a normal probability model with nonstationary variance subject to step changes at irregular time points in order to explain the non-normality and heavy tails observed in the distribution of these series. Research in this area has considered detecting multiple changes through the application of diverse procedures. The majority of Bayesian literature considers only the possibility of a single structural break. When more than one break point is considered it is usually done by focusing on the detection of points of change one at a time because of the heavy computational burden involved in looking for several points of change simultaneously. One difficulty when looking for multiple change points is that the calculations grow exponentially with the number of change points. Another potential problem is the possibility of masking effects, i.e. the statistics used cannot find evidence of a change point because there is more than one, and their combined impact on the statistic offsets any evidence of change. Inclán and Tiao (1994) provide an algorithm to solve both problems based on the iterated cumulative sum of squares (ICSS).

Other models for multiple structural breaks in economic and financial time series come from the results of unit root tests that allow for structural

change or from procedures designed to estimate multiple break models. Tests for unit roots, from the classical point of view, and allowing for multiple structural breaks in the trend can be found in: Park and Sung (1994), Bai (1997), Bai and Perron (1997) and Lumsdaine and Papell (1997). Among other applications in finance Aggarwal, Inclán and Leal (1999) apply the ICSS algorithm to the analysis of stock returns. They study several stock markets in developing countries to identify the events leading to periods of high volatility in stock returns and conclude that most periods of instability are associated with local events. Their ICSS algorithm provided a means to find the most likely points in time when the variability increases or decreases. In Wilson, Aggarwal and Inclán (1996) they also used the ICSS algorithm to identify periods of high variability in stock returns, specifically in the oil sector. They verify that returns on a portfolio of oil and gas companies and returns of oil futures present sudden changes in variance. These changes need to be included in the model before making any predictions of future variances because their presence in the model has serious implications for the estimation of the other parameters of that model. Inclán (1993) has also found that if there are changes in the variance of a series and these are not taken into account when modelling the researcher will find evidence of an Integrated GARCH when in fact there is no such structure in the data. Bayesian results on detecting changes in variance are given in Chib (1996), Holbert (1982), Inclán (1993), Stephens (1994) and Broemeling and Tsurumi (1987). Useful summaries can be found in Maddala and Kim (1997b) as well as in Wang and Zivot (2000). The latter use a very general framework, which includes changes in the trend, level or variance of a time series, using Bayesian methods.

In this paper we study the problem of detecting multiple changes of variance in a time series. We first consider a sequence of independent observations and later we apply the same procedure to an AR(1) series. We follow the general framework of Inclán (1993) and Wang and Zivot (2000). The number of changes as well as their location is determined from their posterior distributions. The emphasis is on detecting the changes. This paper utilizes a Bayesian approach for identifying multiple endogenous changes in the variance of a time series. The break points are estimated endogenously and sequentially. We propose a procedure to detect these variance changes based on a dynamic programming algorithm. The problem of finding the most likely set of break points is equivalently represented as that of finding the shortest path connecting two nodes on a suitable graph. An algorithm for

solving this latter problem is given. The method can be extended to detect changes either in its level, trend or variance of a time series, or structural breaks in econometric models. We use the well known series of daily IBM stock price, May 1961-November 1962, from Box and Jenkins (1976), as well as simulated series to illustrate their application.

2 Bayesian Model for Independent Observations

The motivation for the use of Bayesian models comes from two arguments, Zivot and Phillips (1994). First, classical change point estimation procedures produce a pretesting bias in any subsequent inference. Under a Bayesian approach conditional inferences concerning the parameters of the model can be carried out by detecting the points with greatest probability of being a change point. Alternatively, however, they can be avoided by placing a prior distribution over the location of the structural break, which produces posterior distributions from which unconditional inferences can be made. Second, the resulting limit theory under the classical approach can be quite complicated and the finite sample distributions of the unit root test statistics can be very different from their asymptotic counterparts. Under the Bayesian approach, the distribution theory is valid in finite samples.

In this section we provide a Bayesian model for multiple changes in variance of a series of independent observations that will allow us to determine the number of breaks and their location. Let $\{y_t\}$,be a series of independent observations, such that

$$y_t \sim N(0, \sigma_t^2) \quad , t = 1, ..., T, \quad \text{where} \quad \sigma_t^2 = \sigma_\ell^2 \quad \text{if} \quad (T_{\ell-1}+1) < t < T_\ell \quad (1)$$

for $\ell = 1, ..., m+1$ and where we use the convention that $T_1 = 0$ and $T_{m+1} = T$. That is we have a series with m variance changes, or equivalently $m+1$ regimes, with the changes at $1 \leq T_1 < T_2 < ... < T_m \leq T$. The number of changes, m, as well as the indices $(T_1, ..., T_m)$, of the break points, are explicitly treated as unknown parameters.

The purpose is to use a Bayesian approach to "estimate" the unknown number of change points together with their location when T observations on $\{y_t\}$ are available. Let $\underline{y} = (y_1, ..., y_T)'$ and consider the m-partition $(T_1, ..., T_m)$. Let $\underline{T}_j^{(m)}$ denote the j-th m-partition $(T_1, ..., T_m)$, $j = 1, ..., \binom{T-1}{m}$.

We are assuming the first point is not considered a potential change point, since if it is, it would be so with respect to an unobserved portion of the series and is not relevant for our analysis. The joint density of $\underline{y}$ given a number of changes and a specific partition can be written as

$$f(\underline{y} \mid \underline{\sigma}_m, m, \underline{T}_j^{(m)}) = \frac{1}{(2\pi)^{-T/2}} \prod_{\ell=1}^{m_j+1} \left[\sigma_\ell^{-(T_\ell - T_{\ell-1})} \exp\left\{ -\frac{1}{2\sigma_\ell^2} \sum_{t=T_{\ell-1}+1}^{T_\ell} y_t^2 \right\} \right]. \tag{2}$$

Let the vector of all the parameters in the model be $\underline{\eta}' = (\underline{\sigma}_m, m, \underline{T}_j^{(m)})$, where $\underline{\sigma}_m = (\sigma_1, \sigma_2, ...\sigma_m)'$ is the vector of standard deviations. corresponding to each segment of the j-th m-partition. We need to specify the prior distribution for $\underline{\eta}$. To do this assume $0 \leq m \leq M$ for some M known and large enough to include all possible break points, as well as prior independence. We take $f(\underline{\eta}) \propto f(\underline{\sigma}_m)\ f(m, \underline{T}^{(m)})$ and, Inclán (1993), $f(\underline{\sigma}_m) = \prod_{\ell=1}^{m} f(\sigma_\ell)$, where

$$f(\sigma_\ell) \propto \sigma_\ell^{-(\nu+1)} \exp(-c\sigma_\ell^{-2}), \quad \sigma_\ell > 0, \quad \ell = 1, ..., m, \tag{3}$$

where c and ν are known. The prior for the number of breaks is taken as a discrete uniform distribution $f(m) = (M+1)^{-1}$, $m = 0, ..., M$. This is different from the prior used by Inclán (1993), and even though it is non-informative with respect to the number of changes it is still proper. We also assume that, given the number of breaks, a priori all m-partitions have the same probability of occurring so that

$$f(\underline{T}_j^{(m)} \mid m) = \frac{1}{\binom{T-1}{m}}, \text{ all } \quad j = 1, ..., \binom{T-1}{m}. \tag{4}$$

The joint posterior distribution is then

$$f(\underline{\sigma}_m, m, \underline{T}_j^{(m)} \mid \underline{y}) \propto \binom{T-1}{m}^{-1} \prod_{\ell=1}^{m_j+1} \left[\sigma_\ell^{-(T_\ell - T_{\ell-1})-(\nu+1)} \exp\left\{ -\frac{1}{2\sigma_\ell^2} \left(\sum_{t=T_{\ell-1}+1}^{T_\ell} y_t^2 + 2c \right) \right\} \right], \tag{5}$$

with ν and c known and where m_j indicates that the $j-th$ m-partition has been taken. Upon integrating with respect to $\underline{\sigma}_m$ we obtain the marginal posterior distribution

$$f(m, \underline{T}_j^{(m)} \mid \underline{y}) = K \binom{T-1}{m}^{-1} \prod_{\ell=1}^{m_j+1} \left\{ \Gamma\left[(d_\ell + \nu)/2\right] \left(\sum_{t=T_{\ell-1}+1}^{T_\ell} y_t^2 + 2c \right)^{-[(d_\ell+\nu]/2} \right\}, \tag{6}$$

where $d_\ell = T_\ell - T_{\ell-1}$. In order to derive the posterior for m and $\underline{T}_j^{(m)}$ we express (6) as

$$f(m, \underline{T}_j^{(m)} \mid \underline{y}) = K \binom{T-1}{m}^{-1} SSD(\underline{T}_j^{(m)}), \tag{7}$$

with

$$SSD(\underline{T}_j^{(m)}) = \prod_{\ell=1}^{m_j+1} \left\{ \Gamma\left[(d_\ell + \nu)/2\right] \left(\sum_{t=T_{\ell-1}+1}^{T_\ell} y_t^2 + 2c \right)^{-[(d_\ell+\nu]/2} \right\}.$$

From this equation we can obtain the marginal distribution $f(m \mid \underline{y})$ by adding $f(m, \underline{T}_j^{(m)} \mid \underline{y})$ over the possible values of $\underline{T}_j^{(m)}$. Hence for $m = 0$ we have $f(m = 0 \mid \underline{y}) = K \cdot SSD(\underline{T}_j^{(0)})$. Now for $m = 1$

$$f(m = 1 \mid \underline{y}) = K \binom{T-1}{1}^{-1} \sum_{j=1}^{\binom{T-1}{1}} SSD(\underline{T}_j^{(m)}),$$

and for $m = 2$

$$f(m = 2 \mid \underline{y}) = K \binom{T-1}{2}^{-1} \sum_{j=1}^{\binom{T-1}{2}} SSD(\underline{T}_j^{(m)}),$$

and so on for $m = 3, ...M$. To find K we take

$$\sum_{m=0}^{M} f(m \mid \underline{y}) = 1 = K \sum_{m=0}^{M} \binom{T-1}{m}^{-1} \sum_{j=1}^{\binom{T-1}{m}} SSD(\underline{T}_j^{(m)}),$$

thus if we make

$$RMSD(\underline{T}^{(m)}) = \binom{T-1}{m}^{-1} \sum_{j=1}^{\binom{T-1}{m}} SSD(\underline{T}_j^{(m)})$$

then

$$K^{-1} = \sum_{m=0}^{M} RMSD(\underline{T}^{(m)})$$

and

$$f(m \mid \underline{y}) = K \cdot RMSD(\underline{T}^{(m)}) = \frac{RMSD(\underline{T}^{(m)})}{\sum_{m=0}^{M} RMSD(\underline{T}^{(m)})}, \tag{8}$$

for $m = 0, 1, ..., T$. From (6) and (8) The conditional posterior distribution of $\underline{T}_j^{(m)}$ given m is

$$f(\underline{T}_j^{(m)} \mid \underline{y}, m) = \frac{SSD(\underline{T}_j^{(m)})}{\sum_{j=1}^{\binom{T-1}{m}} SSD(\underline{T}_j^{(m)})}, \tag{9}$$

$j = 1, ..., \binom{T-1}{m}$.

These formulas are analytically very simple, but can be very computing intensive. In the following section we discuss how they may be applied sequentially to detect the break points in a series, while avoiding some of the calculations. Given the number of change points we can use (9) to determine the specific change-points. However this would imply computing $SSD(\underline{T}_j^{(m)})$ for all $\binom{T-1}{m}$ partitions with m changes.

3 Autoregressive Model

In this section we extend the previous results to other types of models. First we consider the case of an AR(1) model with changes in the variance, but constant autoregressive parameter, i.e. let

$$y_t = \phi y_{t-1} + u_t, \qquad t = 1, ..., T,$$

with u_t independent $N(0, \sigma_t^2)$, where σ_t^2 behaves as indicated in (1). The likelihood function in this case is

$$f(\underline{y} \mid \phi, \underline{\sigma}_m, m, \underline{T}_j^{(m)}) = \frac{1}{(2\pi)^{-T/2}} \prod_{\ell=1}^{m_j+1} \left[\sigma_\ell^{-(T_\ell - T_{\ell-1})} \exp\left\{ -\frac{1}{2\sigma_\ell^2} \sum_{t=T_{\ell-1}+1}^{T_\ell} (y_t - \phi y_{t-1})^2 \right\} \right]. \tag{10}$$

In this model we let $\underline{\eta}' = (\phi, \underline{\sigma}_m, m, \underline{T}_j^{(m)})$ and use the prior distributional assumption $f(\underline{\eta}) \propto f(\phi)\ f(\underline{\sigma}_m)\ f(m, \underline{T}^{(m)})$, where $f(\underline{\sigma}_m)$ and $f(m, \underline{T}^{(m)})$ are defined as above and we use a proper (non-informative) prior $f(\phi) = 1/2$, $-1 < \phi < 1$. The posterior distribution is then

$$f(\phi, \underline{\sigma}_m, m, \underline{T}_j^{(m)} \mid \underline{y}) \propto \frac{1}{(2\pi)^{-T/2}} \binom{T-1}{m}^{-1} \prod_{\ell=1}^{m_j+1} \left[\sigma_\ell^{-(T_\ell - T_{\ell-1})-(\nu+1)} \exp\left\{ -\frac{1}{2\sigma_\ell^2} \sum_{t=T_{\ell-1}+1}^{T_\ell} (y_t - \phi y_{t-}\right.\right.$$

Integrating with respect to $\underline{\sigma}_m$ yields

$$f(\phi, m, \underline{T}_j^{(m)} \mid \underline{y}) = K \binom{T-1}{m}^{-1} \prod_{\ell=1}^{m_j+1} \left[\Gamma\left[(d_\ell + \nu)/2 \right] \left\{ \sum_{t=T_{\ell-1}+1}^{T_\ell} (y_t - \phi y_{t-1})^2 + 2c \right\}^{-(d_\ell+\nu)/2} \right],$$

so that in order to carry out the analysis as above we now need to integrate with respect to ϕ and thus obtain an expression for $f(m, \underline{T}_j^{(m)} \mid \underline{y})$. This requires computing

$$SSD(\underline{T}_j^{(m)}) = \int_{-1}^{1} \prod_{\ell=1}^{m_j+1} \left[\Gamma\left[(d_\ell + \nu)/2 \right] \left\{ \sum_{t=T_{\ell-1}+1}^{T_\ell} (y_t - \phi y_{t-1})^2 + 2c \right\}^{-(d_\ell+\nu)/2} \right] d\phi, \tag{11}$$

which must be done numerically, Inclán (1993). We can now use (7) with this definition of $SSD(\underline{T}_j^{(m)})$ and so the argument used to obtain equations (8) and (9) follows in a straightforward manner. It is clear once more that these formulas are analytically simple, but can be very computing intensive, since in this case they involve numerical integration.

4 Prior-Posterior Analysis

To carry out the estimation of the number of break points and their specific location using our Bayesian approach we need to specify the parameters of the prior distributions involved, i.e. c and ν in (3) and M in the prior for m. Following Inclán (1993) we use the sample standard deviation of $\{y_t\}$ as an approximation to σ_{mode} and $c = \sigma^2_{\mathrm{mode}}(\nu+1)/2$, with $\nu = 1$. This definition of c implies that the prior mean is $E(\sigma_\ell) = \sigma_{\mathrm{mode}}$. The choice of M can be made based on the analysts knowledge of the probable number of breaks in the series.

Let M_m be the model with m changes and let $P(M_m)$ be the prior probability that this model is correct. From the definition of the prior for m we know that $P(M_m) = 1/(M+1)$. With the values of the hyper-parameters given we can select the model using posterior odds as in Inclán(1993). Alternatively we can select the model with the highest posterior probability $P(M_m \mid \underline{y})$. This would imply computing (8) for all $m = 1, ..., M$ and then, conditional on the chosen value of m computing (9) for all $m-$partitions. We propose to do the computations sequentially, as will be described in Section 5 of this paper.

With these, or other, values for the hyper-parameters in hand and with the formulas determined for the posterior distributions $f(m \mid \underline{y})$ and $f(\underline{T}_j^{(m)} \mid \underline{y}, m)$ we must proceed to 'estimate' the number of variance changes and, conditional on it, the actual points where the changes took place. We use as an estimate of m the mode of $f(m \mid \underline{y})$, say $\widetilde{m}$, and then find the specific $m-$partition that maximizes $f(\underline{T}_j^{(m)} \mid \underline{y}, \widetilde{m})$, denote it by $\widetilde{\underline{T}_j^{(m)}}$. To carry out inference on ϕ we can then use $f(\phi \mid \widetilde{m}, \widetilde{\underline{T}_j^{(m)}}, \underline{y})$. Notice that to do this we need to compute $SSD(\underline{T}_j^{(m)})$, for $m = 0, ..., M$ and $j = 1, ..., \binom{T-1}{m}$. We propose to do it sequentially. If we know $m = m_0$, as is frequently assumed in the literature, we would then only need to obtain $SSD(\underline{T}_j^{(m_0)})$ for $j = 1, ..., \binom{T-1}{m_0}$. In the case that m is not known we could also do the estimation sequentially, as follows. Let $m = 1$ and find the change point that maximizes $f(\underline{T}_j^{(m)} \mid \underline{y}, m = 1)$. Assuming this change takes place at b_1, given, now maximize $f(\underline{T}_j^{(m)} \mid \underline{y}, m = 2, b_1)$. The procedure can stop whenever the conditional probability of an additional change point is less than a certain pre-specified value, say 0.5 so that the posterior odds are less than one. However, it is clear that this process may not be optimal. In the following

sections we propose an optimal algorithm, which, nevertheless, reduces the number of computations.

If the number of breaks is unknown we could use an approach similar to the one mentioned in Wang and Zivot (2000), who determine this number by use the Schwarz Bayesian Information Criterion (BIC) defined as

$$BIC(m) = -21nL(\widehat{\underline{\eta}}|\underline{y}) + q1n(T),$$

where $L(\widehat{\underline{\eta}}|\underline{y})$ denotes the likelihood function (2) (or (10) in the AR(1) case) evaluated at $\widehat{\underline{\eta}}$ and q denotes the total number of estimated parameters in a model with m breaks. The Schwarz criterion indicates that the model with the highest posterior probability is the one that minimizes the BIC. Yao (1988) has shown that minimizing $BIC(m)$ is a consistent criterion for determining the number of structural changes in a normal sequence of random variables with an unknown number of shifts in mean. Liu et. al. (1997) have extended his result to segmented regression models with exogenous regressors. Instead of using $\widehat{\underline{\eta}}$, we compute $BIC(m)$ using the posterior modes of $\underline{T}_j^{(m)}$ for $j = 1, ..., m$ and use the posterior means of the remaining parameters, ϕ and $\underline{\sigma}_m$.

5 Sequential Detection of Break Points

The process of computing the posterior distribution of all possible $m-$partitions, $m = 0, 1, ..., M$, to determine