Département de sciences économiques

9912

Jumps in the Volatility of Financial Markets

PERRON, Benoît

Université de Montréal

Département de sciences économiques
Université de Montréal
Faculté des arts et des sciences
C.P. 6128, succursale Centre-Ville
Montréal (Québec) H3C 3J7
Canada
http://www.sceco.umontreal.ca
SCECO-information@UMontreal.CA
Téléphone : (514) 343-6539
Télécopieur : (514) 343-7221

Ce cahier a également été publié par le Centre interuniversitaire de recherche en économie quantitative (CIREQ) sous le numéro 0999.

This working paper was also published by the Center for Interuniversity Research in Quantitative Economics (CIREQ), under number 0999.

ISSN 0709-9231

CAHIER 9912

JUMPS IN THE VOLATILITY OF FINANCIAL MARKETS

Benoit PERRON[1]

[1] Centre de recherche et développement en économique (C.R.D.E.) and Département de sciences économiques, Université de Montréal

October 1999

The author would like to thank Peter Phillips, Oliver Linton, Donald Andrews, seminar participants at Yale, and participants at the 1998 Canadian Econometric Study Group for their helpful comments. Inmaculada Fiteni has gratefully shared some of her programs. The author assumes full responsibility for any remaining errors. Financial assistance from an Alfred P. Sloan Dissertation Fellowship and from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

RÉSUMÉ

Des études récentes suggèrent que la variance conditionnelle des rendements financiers est sujette à des sauts. Ce papier étend une procédure non paramétrique de détection de sauts développée par Delgado et Hidalgo (1996) à la détection de sauts dans les moments conditionnels supérieurs, en particulier la variance conditionnelle. Les résultats de simulation démontrent que cette procédure estime de façon raisonnable le nombre de sauts ainsi que leurs emplacements. L'application de cette procédure aux rendements journaliers sur l'indice S&P 500 révèle la présence de plusieurs sauts dans la variance conditionnelle.

Mots clés : sauts, variance conditionnelle, noyau, fenêtre unidirectionnelle

ABSTRACT

Recent work suggests that the conditional variance of financial returns may exhibit sudden jumps. This paper extends a non-parametric procedure to detect discontinuities in otherwise continuous functions of a random variable developed by Delgado and Hidalgo (1996) to higher conditional moments, in particular the conditional variance. Simulation results show that the procedure provides reasonable estimates of the number and location of jumps. This procedure detects several jumps in the conditional variance of daily returns on the S&P 500 index.

Key words : jumps, conditional variance, kernel, one-sided windows

1 Introduction

In the past fifteen years, a large amount of attention has been paid to the properties of second moments of financial data. Proper account of the conditional variance of financial data is important for inference purposes as well as for empirical implementation of options pricing models and optimal hedge portfolios. To capture the apparent volatility clustering found in financial data, Engle (1982) suggested in a seminal paper the ARCH model in which the conditional variance, h_t, is a linear function of past squared returns. Since then, numerous extensions of the ARCH model have been suggested, such as the very popular GARCH model (Bollerslev (1986)):

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + ... + \alpha_p y_{t-p}^2 + \beta_1 h_{t-1} + \ldots + \beta_q h_{t-q} \tag{1}$$

Other variants were developed to capture phenomena such as asymmetry , kinks, and discontinuities for which the original ARCH model is poorly suited.

As an alternative to this wide variety of parametric models, Pagan and Schwert (1990), Pagan and Ullah (1988) and Pagan and Hong (1991) have used various non-parametric estimators of the conditional variance. These estimators assume that the conditional variance process is smooth, and they will be inconsistent at any point of discontinuity. Early evidence for the existence of discontinuities in the conditional variance was provided by Lamoureux and Lastrapes (1990); Hamilton and Susmel (1994) and Cai (1994) provided simple parametric forms by adding a Markov chain to an ARCH model. This framework has been extended recently by Dueker (1997) and Francq and Zakoïan (1999) by allowing for GARCH processes and by Maheu and McCurdy (1999) by allowing duration-dependent transition probabilities. The advantage of these models relative to standard GARCH formulations is that they can adapt much more quickly to periods of high or low volatility. GARCH models are too persistent to capture a sudden increase in volatility; similarly, shocks to the conditional variance die out too slowly to capture certain historical episodes.

This paper will thus test for the presence of jumps in the volatility of financial returns. The methodology used is a suitably modified version of a non-parametric kernel estimator suggested by Delgado and Hidalgo (1996). In view of the numerous parametric models that have been suggested to model time-varying volatility, the non-parametric approach seems promising. Misspecification of the continuous component of the conditional variance may lead to erroneous inference on the presence of jumps. The test is derived by using one-sided windows, first introduced by Müller (1992) , in estimating the conditional variance. At points where a jump occurs, the left-hand side and right-hand side estimates will converge to the left and right limit respectively at that point. The difference between these estimates provides the basis for the detection of a jump.

The paper is organized as follows: section 2 will describe the non-parametric procedure to detect discontinuities in a regression function as developed by Delgado and Hidalgo (1996) and its application to the detection of jumps in the conditional variance. The finite-sample performance of this procedure is analyzed in section 3 through a simulation experiment. Section 4 will present empirical results using weekly returns on the Standard and Poor's 500 index, and section 5 will conclude. All proofs are relegated to a mathematical appendix.

2 Non-parametric jump detection in the conditional variance

In this section, we present results on the estimation of jumps in the conditional variance in a nonparametric fashion. The reader is referred to Delgado and Hidalgo (1996) for more detail on the procedure.

For simplicity, assume that y_t is a random variable with zero mean. We will estimate the conditional moments of y_t by using the functional representation:

$$G\left(y_t\right) = m(X_t) + u_t = g\left(X_t\right) + S\left(Z_t\right) + u\left(X_t\right) \tag{2}$$

where G and g are continuous functions, $X_t = (X_{1,t}, \ldots, X_{p+1,t})'$ is a $p+1$ vector whose last element is time defined as a fraction of the sample $X_{p+1,t} = \frac{t}{T}$ where T is the sample size, $S\left(Z_t\right)$ is a step function with finite jumps whose argument is a member of X_t, and $u\left(X_t\right)$ is a stochastic disturbance term with $E\left(u\left(X_t\right)|X_t\right) = 0$. For notational simplicity, the dependence of S on Z_t and u on X_t will be suppressed in the following and simply written as u_t and S_t. The leading candidates for inclusion in X_t are lagged values of the dependent variable, $y_{t-1}, ..., y_{t-p}$. Delgado and Hidalgo (1996) analyzed the case where $G(y_t) = y_t$.

The model suggested is general in terms of the mechanism triggering changes in S_t. The changes, however, are assumed to be explained by the k^{th} member of X which we denote by Z, possibly "time". Using "time" to identify the occurrence of jumps does not mean that we actually believe that the passage of time is the determining factor in leading to a jump. Rather, any variable that causes the jumps would have to be consistent with the findings obtained by assuming that "time" causes them.

The non-parametric estimator of the regression function that we will analyze is the usual Nadaraya-Watson kernel estimator of $m(x)$:

$$\widehat{m}(x) = \frac{\sum_{t=1}^{T} K\left(\frac{X_t - x}{b}\right) G(y_t)}{\sum_{t=1}^{T} K\left(\frac{X_t - x}{b}\right)} \tag{3}$$

where $K : \mathbb{R}^{p+1} \to \mathbb{R}$ is a kernel and b is the usual bandwidth parameter. For simplicity, the class of kernels we will analyze is of multiplicative form:

$$K\left(\frac{X_t - x}{b}\right) = k\left(\frac{X_{1t} - x_1}{b}\right) \cdots k\left(\frac{X_{p+1t} - x_{p+1}}{b}\right) \tag{4}$$

where $k : \mathbb{R} \to \mathbb{R}$.

The Delgado and Hidalgo (1996) approach for detecting discontinuities is to look at the difference between kernel estimates with one-sided windows along the values of Z_t:

$$\widehat{m}^{+}(z) = \frac{\sum_{t=1}^{T} K^{+}\left(\frac{\widetilde{Z}_t - z}{b}\right) G(y_t)}{\sum_{t=1}^{T} K^{+}\left(\frac{\widetilde{Z}_t - z}{b}\right)} = \frac{\widehat{P}^{+}(z)}{\widehat{f}(z)} \tag{5}$$

and

$$\widehat{m}^{-}(z) = \frac{\sum_{t=1}^{T} K^{-}\left(\frac{\widetilde{Z}_t - z}{b}\right) G(y_t)}{\sum_{t=1}^{T} K^{-}\left(\frac{\widetilde{Z}_t - z}{b}\right)} = \frac{\widehat{P}^{-}(z)}{\widehat{f}(z)} \tag{6}$$

with $K^{-}\left(\frac{\widetilde{Z}_t - z}{b}\right) = k^{-}\left(\frac{Z_t - z}{b}\right) \prod_{s \neq k} k\left(\frac{X_{st} - x_s}{b}\right)$ where $k^{-}(\cdot)$ has domain $\mathbb{R}^{-}$ and $K^{+}\left(\frac{\widetilde{Z}_t - z}{b}\right) = k^{+}\left(\frac{Z_t - z}{b}\right) \prod_{s \neq k} k\left(\frac{X_{st} - x_s}{b}\right)$ where $k^{+}(\cdot)$ has support on $\mathbb{R}^{+}$. This means that $\widehat{m}^{+}(z)$ averages values of $G(y_t)$ with Z_t greater than z, and $\widehat{m}^{-}(z)$ averages points where Z_t is less than z. Often, we will refer to these estimators as right-side and left-side estimators respectively for obvious reason. These estimators were introduced by Müller (1992) with $X_t = Z_t = \frac{t}{T}$.

At points of continuity of $S(Z)$, all three estimators $(\widehat{m}(z), \widehat{m}^{+}(z), \widehat{m}^{-}(z))$ will converge to the same value, $m(z)$. At points where $S(Z)$ is not continuous, all three estimators will converge to different values with $\widehat{m}(z)$ inconsistent in this case.

Decompose the conditional variance as:

$$h_t = var(y_t|\mathcal{F}_{t-1}) = E\left[y_t^2|\mathcal{F}_{t-1}\right] - \{E[y_t|\mathcal{F}_{t-1}]\}^2 \tag{7}$$

where $\mathcal{F}_{t-1}$ is the sigma-field generated by past information and estimate each term by left-sided and right-sided kernel estimates. The appropriate estimator is:

$$\begin{aligned}\widehat{h}^{\pm}(z) &= \frac{\sum_{t=1}^{T} K^{\pm}\left(\frac{\widetilde{Z}_t - z}{b}\right) y_t^2}{\sum_{t=1}^{T} K^{\pm}\left(\frac{\widetilde{Z}_t - z}{b}\right)} - \left\{\frac{\sum_{t=1}^{T} K^{\pm}\left(\frac{\widetilde{Z}_t - z}{b}\right) y_t}{\sum_{t=1}^{T} K^{\pm}\left(\frac{\widetilde{Z}_t - z}{b}\right)}\right\}^2 \\ &= \frac{\widehat{P}_2^{\pm}(z)}{\widehat{f}(z)} - \left[\frac{\widehat{P}_1^{\pm}(z)}{\widehat{f}(z)}\right]^2 \\ &= \widehat{m}_2^{\pm}(z) - \left[\widehat{m}_1^{\pm}(z)\right]^2\end{aligned} \tag{8}$$

For the remainder of the paper, the $\pm$ notation refers to the right-side or left-side estimate as appropriate. The two-sided version of this estimator has been analyzed by Masry and Tjostheim (1995). The behavior of each of $\widehat{m}_2^{\pm}(z)$ and $\widehat{m}_1^{\pm}(z)$ is obtained directly from Delgado and Hidalgo (1996). The asymptotic distribution of $\widehat{h}^{\pm}(z)$ however is complicated by the correlation between the two terms and will be derived below.

Define the process:

$$\widehat{\Delta}(z) = \widehat{h}^{+}(z) - \widehat{h}^{-}(z). \tag{9}$$

The framework is applicable to multiple jumps, and these can be estimated sequentially. Suppose that there are M jumps, with M known, an assumption we will relax later. That is,

$$S(Z) = \Delta_0 + \Delta_1 (Z_t \geq z_1) + \ldots + \Delta_M (Z_t \geq z_M)$$

with $\Delta_1 > \ldots > \Delta_M$ without loss of generality. Define the estimate the first jump point as $\widehat{z}_1 = \arg\max_{z \in Q} \left[\widehat{\Delta}(z)\right]^2$ where $Q = [\underline{z}, \overline{z}]$ for some $\underline{z}, \overline{z} \in int(\chi)$, the sample space of Z. After the first jump is estimated by $\widehat{z}_1$, the estimate of the second jump point is obtained in a similar fashion, but due to the uncertainty in the exact value of z_1, it is necessary to trim the interval over which we search for the next jump. An almost sure bound derived by Yin (1988) is used to prevent the presence of another jump in the vicinity of an estimated jump. When estimating the k^{th} jump, we search over $Q^k = Q - \bigcup_{j=1}^{k-1} Q_j$ with $Q_j = [\widehat{z}_j - 2b, \widehat{z}_j + 2b]$. The distribution of the resulting estimates is summarized in the following theorem:

Theorem 1 *Under assumptions A1-A3 and B1-B5 of Delgado and Hidalgo* (1996), *with* $\widehat{z}_k \in int\left(Q^k\right)$,

$$(i)\ \left(Tb^{p-1}\right)^{\frac{1}{2}}\left(\widehat{z}_k - z_k\right) \xrightarrow{d} N\left(0, \frac{\gamma_{(1)} S\left(z_k\right)}{\Delta_h\left(z_k\right)^2\left[k_{(1)}^{-}(0)\right]^2 f\left(z_k\right)}\right)$$

$$(ii)\ \left(Tb^{p+1}\right)^{\frac{1}{2}}\left(\widehat{\Delta}_h\left(\widehat{z}_k\right) - \Delta_h\left(z_k\right)\right) \xrightarrow{d} N\left(0, \frac{\gamma_{(0)} S\left(z_k\right)}{f\left(z_k\right)}\right)$$

and $\widehat{z}_1, \widehat{z}_2, \ldots, \widehat{z}_M$ *are independent; where*

$k_{(1)}^{-}(\nu)$ *is the first derivative of the left-hand kernel at* ν,
$f\left(z_k\right)$ *is the density of* X *with* $Z = z_k$,
$\gamma_{(n)} = \int \left[K_{(n)}^{-}(v)\right]^2 dv$, *the integral of the* n^{th} *derivative of the left-hand kernel,*
$S\left(z_k\right) = \left[\left(\sigma_{2+}^2\left(z_k\right) + \sigma_{2-}^2\left(z_k\right)\right) + 4m_1^{+}\left(z_k\right)^2\left(\sigma_{1+}^2\left(z_k\right) + \sigma_{1-}^2\left(z_k\right)\right) - 4m_1^{+}\left(z_k\right)H\left(z_k\right)\right]$,
$\sigma_{\pm}^2\left(z_k\right)$ *the left-side and right-side variance of* u_t *at* z_k, *and*

$$H\left(z_k\right) = E\left[u_1^{+}\left(z_k\right)u_2^{+}\left(z_k\right)\right] + E\left[u_1^{-}\left(z_k\right)u_2^{-}\left(z_k\right)\right] + E\left[u_1^{+}\left(z_k\right)m_2^{+}\left(z_k\right)\right] + E\left[u_1^{-}\left(z_k\right)m_2^{-}\left(z_k\right)\right] + E\left[u_2^{+}\left(z_k\right)m_1^{+}\left(z_k\right)\right] + E\left[u_2^{-}\left(z_k\right)m_1^{-}\left(z_k\right)\right].$$

Note that the rate of convergence in the previous theorem is slower than the $\sqrt{T}$ rate obtained in parametric models. [1] However, in contrast to parametric models (see Bai (1997) and Bai and Perron (1998)), the absence or presence of other jumps does not change the behavior of the break point estimates. This is due to the local nature of kernel estimation.

Notice that if we are ready to assume that the mean function does not have any discontinuity, the problem simplifies since the estimates of the mean cancel out from $\widehat{\Delta}_h(z)$ if

[1] Loader (1996) develops an alternative procedure which achieves convergence at rate $\sqrt{T}$. However, this requires putting non-zero weight on the current observation which is undesirable in this case. See Pagan and Hong (1991) for this point.

$\widehat{m}_1^+ = \widehat{m}_1^- = \widehat{m}_1$. In this situation, we can look for jumps in the conditional variance by looking for jumps in y_t^2 only since $\widehat{\Delta}_h(z) = \widehat{m}_2^+(z) - \widehat{m}_2^-(z)$ which does not depend on the mean and use the results in Delgado and Hidalgo (1996) directly. An alternative approach to dealing with the mean is to use a parametric model, such as one from the ARMA class or a Markov-switching model. In this case also, the distribution of $\widehat{\Delta}_h(z)$ and $\widehat{z}_k$ does not depend on the estimated mean since the parametric estimates of the mean will converge at a faster rate than the non-parametric estimates of the higher moments.

In the case of financial returns, we can obtain more precise formulas. Decompose returns as usual as:

$$y_t = m_1(X_t) + h_t^{\frac{1}{2}} \varepsilon_t \tag{10}$$

$$\equiv m_1(X_t) + u_{1t} \tag{11}$$

where ε_t is assumed *i.i.d.* $(0,1)$ and independent of $h_t = E_{t-1}[u_{1t}^2] = E[u_{1t}^2 | \mathcal{F}_{t-1}]$, the conditional variance of y_t. The formulas from theorem 1 in this special case are contained in the following corollary:

Corollary 2 *In model* (10),

$$S(z_k) = \left[h^+(z_k)^2 + h^-(z_k)^2\right] E\left(\varepsilon_t^4 - 1\right) - 4\widetilde{\Delta}_1(z_k)\, h^-(z_k)^{\frac{3}{2}} E\left(\varepsilon_t^3\right) - 4\widetilde{\Delta}_1^2(z_k)\, h^-(z_k)$$

where $\widetilde{\Delta}_1(z_k) = m_1^+(z_k) - m_1^-(z_k)$.

This corollary allows us to construct plug-in estimates of the standard errors of both the jump points and the jump sizes.

In the previous results, the number of jumps, M, was assumed to be known, a situation unlikely in practice. Theorem 1 leads to a sup test for the presence of j jumps against only $j-1$. Define the statistic

$$\widehat{\theta}_j = \sup_{q \in Q^j} \left| \frac{\widehat{\Delta}_h(q)\, f(q)}{S(q)\, \gamma_{(0)}} \right| \tag{12}$$

This is simply the largest values of the t-statistic for the hypothesis $\Delta_h(z_j) = 0$. The distribution of the test is:

Theorem 3 *Under the same assumptions as in theorem 1 and* $\int k_+(u)\, k_+^{(1)}(u)\, du = 0$, *if* $\Delta_h(z_k) = 0$, *as* $T \to \infty$

$$\Pr\left[\left(Tb^{p+1}\right)^{\frac{1}{2}} \widehat{\theta}_k < \vartheta\right] \to \prod_{j=1}^{k} \exp\left[-2\exp\left(-B_T^j\left(\vartheta - C_T^j\right)\right)\right]$$

where $B_T^j = \left[2\ln\left(\frac{\lambda(Q_j)}{b}\right)\right]^{\frac{1}{2}}$, $\lambda(\cdot)$ *is Lebesgue measure, and* $C_T^j = B_T^j + \frac{1}{B_T^j}\ln\left[\frac{1}{2\pi}\left(\frac{\int k_{(1)}^{+}(u)^2 du}{\int k^{+}(u)^2 du}\right)^{\frac{1}{2}}\right]$.

A stopping rule must then be chosen. A possibility is to stop if the p–value of the above test is greater than a predetermined significance level. Unfortunately, this procedure will not consistently estimate M, since at each step there is always a positive probability of rejecting the null hypothesis regardless of sample size. Thus, when the procedure arrives at stage $M+1$, the test of $M+1$ versus M jumps gives a non-zero probability of not rejecting the additional jump, no matter how large the sample size is. Thus, this procedure will overstate the number of jumps selected. However, in practice, the probability of overspecifying the number of jumps will go down very quickly. For a significance level α for the tests, the probability of overspecifying the number of jumps by j is α^j. At usual significance levels, this goes to zero very quickly. An alternative would be to let α approach zero at an appropriate rate.

The technique presented here suffers from obvious drawbacks. Because it requires the deletion of a large number of observations around the estimated jump points, the procedure will not be able to detect jumps that are close to each other or that are quickly reversed. It should be more apt at detecting long-lasting regime changes. In addition, the procedure assumes that $G(y_t)$ can be represented by a function of few conditioning variables. This might not be realistic since certain parametric models, such as GARCH, imply that y_t^2 is a function of its infinite past. If such a model is correct, then our non-parametric estimator may not provide a satisfactory description of the behavior of the conditional variance. If a few lags are not sufficient to reduce u_t to a martingale difference sequence, the asymptotic theory on which inferences are based will be invalid.

3 Simulation Results

In this section, we will present results from a limited simulation experiment. Specifically, we want to look at the number of jumps found, both correctly and incorrectly, the sensitivity of the results to the various choices of the parameters, notably the bandwidth and the lag length, and the behavior of the estimated jumps.

To make the experiment as realistic as possible, we use the commonly-used GARCH $(1,1)$ model:

$$\begin{aligned} y_t &= \mu + h_t^{\frac{1}{2}} \varepsilon_t \\ h_t &= \omega_t (1 - \alpha - \beta) + \alpha (y_{t-1} - \mu)^2 + \beta h_{t-1}. \end{aligned}$$

The values of the parameters are set at those estimated from the data set used in the empirical application in the next section. Specifically, the data set is made up of 4299 daily continuously–compounded returns on the Standard and Poors 500 index between January 2nd, 1980 and December 31st, 1996. The data is plotted in figure 1.

***** Insert figure 1 here *****

We allow two sets of parameters, one to verify the size of the procedure and one to verify its power. In particular, the first set of parameters makes ω_t a constant with no jump and removes all dynamics in h_t $(\alpha = \beta = 0)$. The second set of parameters allows for two jumps in ω_t:

$$\omega_t = \Delta_0 + \Delta_1 1\left(\frac{t}{T} \geq \tau_1\right) + \Delta_2 1\left(\frac{t}{T} \geq \tau_2\right)$$

The values of the estimated parameters are in table 1.

***** Insert table 1 here *****

With each of these sets of parameter values, the condition for y_t to have finite fourth moment is satisfied. The distribution of ε_t is taken as $N(0,1)$, and the sample size is $T = 1000$. We replicate the experiments 500 times.

The bandwidth is chosen as $b_j = c\widehat{\sigma}_j T^{\frac{-1}{p+5}}$ where c is the bandwidth constant, $\widehat{\sigma}_j$ is the estimated standard deviation of variable j, T is the sample size, and p is the lag length. We allow for three values of c : 0.8, 1, and 1.2. Moreover, we analyze a data-determined selection rule by minimizing a variation of the cross-validation criterion:

$$CV(c) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t^2 - \widehat{m}_{2t}^{+}\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\left(y_t^2 - \widehat{m}_{2t}^{-}\right)^2$$

where $\widehat{m}_{2t}^{+}$ (respectively $\widehat{m}_{2t}^{-}$) is the right side (respectively left side) estimate of $E_{t-1}\left(y_t^2\right)$. This criterion uses only the fit on the second moment of y_t to choose the bandwidth. We could generalize this criterion to include the fit in the first moment as well, but the difference between the two criteria was usually not very important. The bandwidth constant is allowed to vary between 0.8 and 1.2 with a step of 0.1. Finally, the one-sided kernel is $k_+(x) = x(3-x)e^{-x}1(x \geq 0)$, while the two-sided kernel is Gaussian. All tests are carried out at the 5% significance level with normality imposed and ten percent of the observations are deleted at the beginning and end of the sample.

The results for the model with constant variance are presented in table 2 and figure 2. Each column of table 2 presents the frequency distribution of the estimated number of jumps for each combination of lag length and bandwidth constant. Each cell gives the fraction of replications where the estimated number of breaks equaled the number in the left margin. The results are excellent for a small number of lags in that few jumps are erroneously detected for any value of lag length and bandwidth constant. However, the table stresses the importance of not overspecifying the information set as this deteriorates the results dramatically as the information set is expanded. The second row from the bottom shows that the number of incomplete replications is very small. Two reasons explain an "Incomplete": first, after deleting observations around the estimated jumps, either no or too little data is left for the test to be computed, or secondly, $\widehat{f}$ is too small for all observations. Note also that cross-validation helps in improving the results at all lags, but in particular for the larger numbers of lags.

***** Insert table 2 here *****

The results for the second data-generating process are presented in table 3 and figures 3 and 4. In this case, the first feature to notice in table 3 is the large number of replications that were "incomplete" for $p > 0$. However, with at least one lagged return included in the conditioning set, the number of jumps found is relatively close to the true one on average. Note also that cross validation generally reduces the number of incomplete replications. Figure 3 shows the same information in graphical form with the bandwidth chosen by cross validation. For $p > 1$, the distribution is well-centered as the dynamics of the conditional variance are better captured.

***** Insert table 3 here *****

***** Insert figure 3 here *****

***** Insert figure 4 here *****

Figure 4 displays a histogram of the estimated location of the jumps with cross-validated bandwidth. The jumps are located at fractions 0.3626 and 0.7046 of the sample. The second jump is the dominant one and should be estimated first. The figure supports this assertion, but the jump at 0.3626 is poorly estimated. There is much mass near the boundary of the sample space (0.1) which may reflect boundary effects associated with kernel estimation.

Our conclusion from these two small experiments is that the procedure performs reasonably well in estimating the number of jumps in a data set. It, however, has a tendency to run out of observations quickly and to estimate the second (and presumably subsequent) jumps imprecisely. The use of cross-validation improves the results noticeably and should be used in practical implementation of the procedure.

4 Empirical Results

In this section, we will apply our jump detection procedure to the series of daily returns on the S&P 500 index between January 2nd, 1980 and December 31st, 1986 used to calibrate the simulation exercise above. The data set comprises 4299 observations.

The detection procedure is implemented in the same way as for the simulation experiment with up to 3 lags in the information set. The moments $E\left(\varepsilon_t^3\right)$ and $E\left(\varepsilon_t^4-1\right)$ are estimated by using a two-sided kernel over the data remaining after deleting twice the bandwidth on each side of the currently estimated jump. The use of the two-sided kernel provided more reliable estimates of these moments.

Table 4 gives the results for 4 lag lengths considered with the bandwidth chosen by our modified version of cross validation. Full results for other choices of bandwidth are available upon request from the author. As in the simulation experiment, we allowed the constant to vary between 0.8 and 1.2 in increments of 0.1. The results are quite interesting in two ways: first, the procedure did not run out of observations at all lags, and some jump dates seem to be recurring, in particular late 1982-mid 1983, late 1986, and August 1990.

***** Insert table 4 here *****

The jump in late 1986 occurs a few months before the October 1987 crash and seems quite natural, as does the August 1990 jump which coincides with Iraq's invasion of Kuwait.

5 Conclusion

This paper has developed a non-parametric procedure to test for the presence of discontinuities in the conditional variance. Simulation evidence showed that the procedure performs well in detecting jumps when some exist and will not find many when none exist if the dynamics are approximately well-specified.

Results from an application to daily stock returns show that some sudden changes in volatility occurred over the sample period. The findings that the volatility of financial markets cannot be described as a smooth functions of lagged returns casts doubts on the practice of fitting GARCH-type models. It remains to be evaluated how much is lost by neglecting to model this feature of the data. It may very well be that GARCH models still provide a good first approximation to the behavior of the data. This will be left for future research efforts.

References

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

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

[3] Bollerslev, T., 1986, Generalized Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

[4] Bollerslev, T., R. F. Engle, and D. B. Nelson, 1994, ARCH Models, in: Engle, R. F. and D. L. McFadden, eds, *Handbook of Econometrics, vol. 4*, (Elsevier Science), 2959-3038.

[5] Cai, J., 1994, A Markov Model of Switching-Regime ARCH, *Journal of Business and Economic Statistics*, 12, 309-316.

[6] Davidson, J., 1994, *Stochastic Limit Theory*, (New York: Oxford University Press).

[7] Delgado, M. A. and J. Hidalgo, 1996, Nonparametric Inference on Structural Break Points, mimeo.

[8] Ding, Z., C. W. J. Granger, and R. F. Engle, 1993, A Long-Memory Property Of Stock Market Returns and a New Model, *Journal of Empirical Finance*, 1, 86-106.

[9] Dueker, M. J., 1997, "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility", *Journal of Business and Economics Statistics*, 15, 26-34.

[10] Engle, R. F., 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica*, 50, 987-1008.

[11] Francq, C. and J.-M. Zakoïan, 1999,"Linear-representations Based Estimation of Switching-regime GARCH Models", mimeo.

[12] Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993, On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, 48, 1779-1801.

[13] Hamilton, J. D. and R. Susmel, 1994, Autoregressive Conditional Heteroskedasticity and Changes in Regime, *Journal of Econometrics*, 64, 307-333.

[14] Lamoureux, C. G. and W. D. Lastrapes, 1990, Persistence In Variance, Structural Change, and the GARCH Model, *Journal of Business and Economic Statistics*, 8, 225-234.

[15] Li, C.W. and W.K. Lee, 1996, On a Double Threshold Autoregressive Heteroscedastic Time Series Model, *Journal of Econometrics*, 11, 253-274.

[16] Liu, M., 1995, Modeling Long Memory In Stock Market Volatility, Mimeo, Duke University.

[17] Loader, C. R., 1996, Change Point Estimation Using Nonparametric Regression, *Annals of Statistics*, 24, 1667-1678.

[18] Maheu, J. M., and T. H. McCurdy, 1999, "A New Approach to Modelling Volatility Dynamics", mimeo.

[19] Masry, Elias and Dag Tjϕstheim, 1995, Nonparametric Estimation and Identification of Nonlinear ARCH Time Series, *Econometric Theory*, 11, 258-289.

[20] Müller, H.-G., 1992, Change-Points in Nonparametric Regression Analysis, *Annals of Statistics*, 20, 737-761.

[21] Pagan, A. R., 1996, The Econometrics of Financial Markets, *Journal of Empirical Finance*, 3, 15-102.

[22] Pagan, A. R. and Y. S. Hong, 1991, Nonparametric Estimation and the Risk Premium in: Barnett, W. A., J. Powell, and G. E. Tauchen eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, (Cambridge University Press: Cambridge), 51-75.

[23] Pagan, A. R. and G. W. Schwert, 1990, Alternative Models for Conditional Stock Volatility, *Journal of Econometrics*, 45, 267-290.

[24] Pagan, A. R. and A. Ullah, 1988, The Econometric Analysis of Models with Risk Terms, *Journal of Applied Econometrics*, 3, 87-105.

[25] Perron, P., 1989, The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 57, 1361-1401.

[26] Rabemananjara, R. and J.M. Zakoian, 1993, Threshold ARCH Models and Asymmetries in Volatility, *Journal of Applied Econometrics*, 8, 31-49.

[27] Robinson, P., 1983, Nonparametric Estimators for Time Series, *Journal of Time Series Analysis*, 4, 185-207.

[28] van Norden, S., 1996, Regime Switching as a Test for Exchange Rate Bubbles, *Journal of Econometrics*, 11, 219-251.

[29] Whitt, W., 1970, Weak Convergence of Probability Measures on the Function Space $C\left[0,\infty\right)$, *Annals of Mathematical Statistics*, 41, 939-944.

[30] Yin, Y. Q., 1988, Detection of the Number, Locations and Magnitudes of Jumps, *Communications in Statistics – Stochastic Models,* 4, 445-455.

[31] Zakoian, J.-M., 1994, Threshold Heteroskedastic Models, *Journal of Economic Dynamics and Control*, 18, 931-955.

A Proofs

A.1 Proof of theorem 1

The proof of this theorem involves getting a local approximation for the process $\widehat{\Delta}_h$, obtaining a functional central limit theorem for this approximating process with a maximum which is distributed as postulated in the theorem.

We can infer directly from the proof of theorem 1 in Delgado and Hidalgo (1996) that $\widehat{z}_k - z_k = O_p\left(\left(Tb^{p-1}\right)^{\frac{1}{2}}\right)$. Thus, let $\widehat{\delta}_h\left(v\right) = \widehat{\Delta}_h\left(z_k + \frac{v}{\left(Tb^{p-1}\right)^{\frac{1}{2}}}\right)$. The key step is to show that:

$$\left(Tb^{p+1}\right)\left(\widehat{\delta}_h\left(v\right) - \widehat{\delta}_h\left(0\right)\right) \Rightarrow \eta_h\left(v\right) \tag{13}$$

on $C\left(-\infty,\infty\right)$ where $\eta_h\left(v\right) \sim N\left(\frac{k_{(1)}^{-}(0)v^2\Delta_{12}}{2}, \frac{v^2\gamma_{(1)}S(z_k)}{f(0)}\right)$ with $\Delta_{12} = \widetilde{\Delta}_2\left(z_k\right) - 2\widetilde{\Delta}_1\left(z_k\right)m_1^{+}\left(z_k\right)$.

First, define $\widehat{\delta}\left(v\right) = \begin{pmatrix} \widehat{\delta}_1\left(v\right) \\ \widehat{\delta}_2\left(v\right) \end{pmatrix} = \begin{pmatrix} \widehat{\widehat{\Delta}}_1\left(z_k + \frac{v}{\left(Tb^{p-1}\right)^{\frac{1}{2}}}\right) \\ \widehat{\widehat{\Delta}}_2\left(z_k + \frac{v}{\left(Tb^{p-1}\right)^{\frac{1}{2}}}\right) \end{pmatrix}$. We will show that:

$$\left(Tb^{p+1}\right)\left(\widehat{\delta}\left(v\right) - \widehat{\delta}\left(0\right)\right) \Rightarrow \eta\left(v\right)$$

where $\eta\left(v\right) \sim N\left(\left(\begin{array}{c} \frac{\widetilde{\Delta}_1(z_k)k_{(1)}^{-}(0)v^2}{2} \\ \frac{\widetilde{\Delta}_2(z_k)k_{(1)}^{-}(0)v^2}{2} \end{array}\right), \left(\begin{array}{cc} \frac{v^2 s_1(z_k)\gamma_{(1)}}{f(z_k)} & \frac{v^2\gamma_{(1)}H(z_k)}{f(z_k)} \\ \frac{v^2\gamma_{(1)}H(z_k)}{f(z_k)} & \frac{v^2 s_2(z_k)\gamma_{(1)}}{f(z_k)} \end{array}\right)\right)$ with $s_j\left(z_k\right) = \sigma_{j+}^2\left(z_k\right) + \sigma_{j-}^2\left(z_k\right)$ for $j = 1, 2$ and

$$\begin{aligned} H\left(z_k\right) &= E\left[u_1^{+}\left(z_k\right)u_2^{+}\left(z_k\right)\right] + E\left[u_1^{-}\left(z_k\right)u_2^{-}\left(z_k\right)\right] + E\left[u_1^{+}\left(z_k\right)m_2^{+}\left(z_k\right)\right] \\ &\quad + E\left[u_1^{-}\left(z_k\right)m_2^{-}\left(z_k\right)\right] + E\left[u_2^{+}\left(z_k\right)m_1^{+}\left(z_k\right)\right] + E\left[u_2^{-}\left(z_k\right)m_1^{-}\left(z_k\right)\right]. \end{aligned}$$

Write $\widehat{\delta}_j\left(v\right) - \widehat{\delta}_j\left(0\right) = r_j\left(v\right) + o_p\left(Tb^{p+1}\right)$ and $\widehat{\delta}\left(v\right) - \widehat{\delta}\left(0\right) = r\left(v\right) = \begin{pmatrix} r_1\left(v\right) \\ r_2\left(v\right) \end{pmatrix}$ where $r_j\left(v\right) = \frac{1}{f(z_k)}\left\{\mathbb{P}_j^{+}\left(v\right) - \mathbb{P}_j^{-}\left(v\right) - m_j^{+}\left(z_k\right)\mathbb{F}_j^{+}\left(v\right) + m_j^{-}\left(z_k\right)\mathbb{F}_j^{-}\left(v\right)\right\}$, $\mathbb{P}_j^{\pm}\left(v\right) = \widehat{P}_j^{\pm}\left(z_k + \frac{v}{\left(Tb^{p-1}\right)^{\frac{1}{2}}}\right) - \widehat{P}_j^{\pm}\left(z_k\right)$ and $\mathbb{F}_j^{\pm}\left(v\right) = \widehat{f}_j^{\pm}\left(z_k + \frac{v}{\left(Tb^{p-1}\right)^{\frac{1}{2}}}\right) - \widehat{f}_j\left(z_k\right) = \mathbb{F}_i^{\pm}\left(v\right)$ for $\forall i, j$.

By using lemma 7.1 in Robinson (1983) $(Tb^{p+1})\, r_j(v)$ will have a limiting normal distribution for fixed v, while by using propositions 1-3 in Appendix B, we obtain $E[r_j(v)] = \frac{\widetilde{\Delta}_j(z_k)k^{-}_{(1)}(0)v^2}{2}$. The diagonal elements of the covariance matrix are obtained via propositions 4-8. To obtain the off-diagonal elements, we need to evaluate $cov\left[(Tb^{p+1})\, r_1(v_1), (Tb^{p+1})\, r_2(v_2)\right]$. Expanding, we obtain:

$$
\begin{aligned}
\lim_{T\to\infty} cov\left[\left(Tb^{p+1}\right) r_1(v_1), \left(Tb^{p+1}\right) r_2(v_2)\right] &= \lim_{T\to\infty} \frac{\left(Tb^{p+1}\right)^2}{f(z_k)^2} cov\{\mathbb{P}_1^{+}(v_1)\,\mathbb{P}_2^{+}(v_2) - \mathbb{P}_1^{+}(v_1)\,\mathbb{P}_2^{-}(v_2) \\
&\quad -\mathbb{P}_1^{-}(v_1)\,\mathbb{P}_2^{+}(v_2) + \mathbb{P}_1^{-}(v_1)\,\mathbb{P}_2^{-}(v_2) \\
&\quad -m_2^{+}(z_k)\,\mathbb{F}_2^{+}(v_2)\,\mathbb{P}_1^{+}(v_1) \\
&\quad +m_2^{-}(z_k)\,\mathbb{P}_1^{+}(v_1)\,\mathbb{F}_2^{-}(v_2) \\
&\quad +m_2^{+}(z_k)\,\mathbb{P}_1^{-}(v_1)\,\mathbb{F}_2^{+}(v_2) \\
&\quad -m_2^{+}(z_k)\,\mathbb{P}_1^{-}(v_1)\,\mathbb{F}_2^{-}(v_2) \\
&\quad -m_1^{+}(z_k)\,\mathbb{F}_1^{+}(v_1)\,\mathbb{P}_2^{+}(v_2) \\
&\quad +m_1^{+}(z_k)\,\mathbb{F}_1^{+}(v_1)\,\mathbb{P}_2^{-}(v_2) \\
&\quad +m_1^{+}(z_k)\,m_2^{+}(z_k)\,\mathbb{F}_1^{+}(v_1)\,\mathbb{F}_2^{+}(v_2) \\
&\quad +m_1^{-}(z_k)\,\mathbb{P}_2^{+}(v_2)\,\mathbb{F}_1^{-}(v_1) \\
&\quad -m_1^{-}(z_k)\,\mathbb{F}_1^{-}(v_1)\,\mathbb{P}_2^{-}(v_2) \\
&\quad -m_1^{+}(z_k)\,m_2^{-}(z_k)\,\mathbb{F}_1^{+}(v_1)\,\mathbb{F}_2^{-}(v_2) \\
&\quad -m_1^{-}(z_k)\,m_2^{+}(z_k)\,\mathbb{F}_2^{+}(v_2)\,\mathbb{F}_1^{-}(v_1) \\
&\quad +m_1^{-}(z_k)\,m_2^{-}(z_k)\,\mathbb{F}_1^{-}(v_1)\,\mathbb{F}_2^{-}(v_2)\}
\end{aligned}
$$

Using the propositions in Appendix B, we obtain, after cancellation,

$$
\begin{aligned}
\lim_{T\to\infty} cov\left[\left(Tb^{p+1}\right) r_1(v_1), \left(Tb^{p+1}\right) r_2(v_2)\right] &= \frac{\gamma_{(1)} v_1 v_2}{f(z_k)}\{E\left[u_1^{+}(z_k)\,u_2^{+}(z_k)\right] + E\left[u_1^{+}(z_k)\,m_2^{+}(z_k)\right] \\
&\quad +E\left[u_2^{+}(z_k)\,m_1^{+}(z_k)\right] + E\left[u_1^{-}(z_k)\,u_2^{-}(z_k)\right] \\
&\quad +E\left[u_1^{-}(z_k)\,m_2^{-}(z_k)\right] + E\left[u_2^{-}(z_k)\,m_1^{-}(z_k)\right]\} \\
&= \frac{\gamma_{(1)} v_1 v_2 H(z_k)}{f(z_k)}
\end{aligned}
$$

Application of the Cramer-Wold device completes the proof of the convergence of the finite-dimensional distributions. Since the two marginal probability measures are shown to be tight in Delgado and Hidalgo (1996), the joint probability measure is also tight (Davidson (1994), theorem 26.23) and thus $(Tb^{p+1})\left(\widehat{\delta}(v)-\widehat{\delta}(0)\right) \Rightarrow \eta(v)$ on $C(-\infty,\infty)$ by Whitt (1970).

Next, define $\mu_j^{\pm}(v) = m_j^{\pm}\left(z_k + \frac{v}{(Tb^{p-1})^{\frac{1}{2}}}\right)$ and rewrite $\widehat{\delta}_h(v)-\widehat{\delta}_h(0)$ as

$$\begin{aligned}
\widehat{\delta}_h(v)-\widehat{\delta}_h(0) &= \widehat{\delta}_2(v)-\widehat{\delta}_2(0)-\left[\widehat{\mu}_1^{+}(v)^2-\widehat{\mu}_1^{-}(v)^2\right]+\left[\widehat{\mu}_1^{+}(0)^2-\widehat{\mu}_1^{-}(0)^2\right] \\
&= \widehat{\delta}_2(v)-\widehat{\delta}_2(0)-\left[\widehat{\delta}_1(v)^2+2\widehat{\delta}_1(v)\widehat{\mu}_1^{-}(v)\right]+\left[\widehat{\delta}_1(0)^2+2\widehat{\delta}_1(0)\widehat{\mu}_1^{-}(0)\right] \\
&= \widehat{\delta}_2(v)-\widehat{\delta}_2(0)-\left[\widehat{\delta}_1(v)^2-\widehat{\delta}_1(0)^2\right]-2\widehat{\delta}_1(v)\left[\widehat{\mu}_1^{-}(v)-\mu_1^{-}(v)\right] \\
&\quad -2\widehat{\delta}_1(0)\left[\widehat{\mu}_1^{-}(0)-\mu_1^{-}(0)\right]-2\widehat{\delta}_1(v)\mu_1^{-}(v)+2\widehat{\delta}_1(0)\mu_1^{-}(0) \\
&= \widehat{\delta}_2(v)-\widehat{\delta}_2(0)-\left[\widehat{\delta}_1(v)^2-\widehat{\delta}_1(0)^2\right]-2\widehat{\delta}_1(v)\mu_1^{-}(v)+2\widehat{\delta}_1(0)\mu_1^{-}(0) \\
&\quad +o_p\left(Tb^{p+1}\right)
\end{aligned}$$

where the last line follows from the fact that $\widehat{\delta}_1(v)$ is $O_p(Tb^{p+1})$ as proved above and $\widehat{\mu}_1^{-}(v) \xrightarrow{p} \mu_1^{-}(v)$.

Adding and subtracting $2\widehat{\delta}_1(0)\mu_1^{-}(v)$ and linearizing the squared terms around $\delta_1(v)$ and $\delta_1(0)$ respectively and $2\widehat{\delta}_1(0)\mu_1^{-}(v)$ around $2\widehat{\delta}_1(0)\mu_1^{-}(0)$ gives the representation:

$$\widehat{\delta}_h(v)-\widehat{\delta}_h(0) = \widehat{\delta}_2(v)-\widehat{\delta}_2(0)-2\widetilde{\Delta}_1(z_k)\left[\widehat{\delta}_1(v)-\widehat{\delta}_1(0)\right]-2\left[\widehat{\delta}_1(v)-\widehat{\delta}_1(0)\right]m_1^{-}(z_k)+o_p\left(Tb^{p+1}\right)$$

Using the delta method around $v=0$ for the function $(Tb^{p+1})\,d(v)'\left(\widehat{\delta}(v)-\widehat{\delta}(0)\right)$ where $d(v) = \begin{pmatrix} -2\delta_1(v)-2\mu_1^{-}(v) \\ 1 \end{pmatrix}$ gives, as required,

$$\left(Tb^{p+1}\right)\left[\widehat{\delta}_h(v)-\widehat{\delta}_h(0)\right] \Rightarrow \eta_h(v)$$

where $\eta_h(v) \sim N\left(\Gamma_h v^2, \sum_h(v)\right)$ with $\Gamma_h = \frac{k_{(1)}^{-}(0)\left[\widetilde{\Delta}_2(z_k)-2\widetilde{\Delta}_1(z_k)m_1^{+}(z_k)\right]}{2}$ and $\sum_h(v)$ derived as:

$$\begin{aligned}
\sum\nolimits_h(v) &= d(0)'\,var\left(\left(Tb^{p+1}\right)\widehat{\delta}(v)-\widehat{\delta}(0)\right)d(0) \\
&= \begin{pmatrix} -2m_1^{+}(z_k) & 1 \end{pmatrix} var\left(\left(Tb^{p+1}\right)\left[\widehat{\delta}(v)-\widehat{\delta}(0)\right]\right)\begin{pmatrix} -2m_1^{+}(z_k) \\ 1 \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{4m_1^+(z_k)^2 v^2 s_1(z_k)\gamma_{(1)}}{f(z_k)} - \frac{4m_1^+(z_k) v^2 \gamma_{(1)} H(z_k)}{f(z_k)} + \frac{v^2 s_2(z_k)\gamma_{(1)}}{f(z_k)} \\
&= \frac{v^2 \gamma_{(1)} S(z_k)}{f(z_k)}
\end{aligned}
$$

Let $\widehat{v} = \arg\max \left[\widehat{\delta}_h(v)^2 - \widehat{\delta}_h(0)^2\right]$. This objective function can be rewritten as:

$$
\widehat{\delta}_h(v)^2 - \widehat{\delta}_h(0)^2 = \left[\widehat{\delta}_h(v) - \widehat{\delta}_h(0)\right]^2 + 2\left[\widehat{\delta}_h(v) - \widehat{\delta}_h(0)\right]\left[\widehat{\delta}_h(0) - \delta_h(0)\right] + 2\delta_h(0)\left[\widehat{\delta}_h(v) - \widehat{\delta}_h(0)\right]
$$

>From above, only the last term will matter asymptotically and $(Tb^{p+1})\left[\widehat{\delta}_h(v)^2 - \widehat{\delta}_h(0)^2\right] = 2\delta_h(0)(Tb^{p+1})\left[\widehat{\delta}_h(v) - \widehat{\delta}_h(0)\right] + o_p(1) \sim N\left(2\delta_h(0)\Gamma_h v^2, 4\delta_h(0)^2 \sum_h(v)\right)$.

The limiting distribution can be rewritten as $2\delta_h(0)\left[\Gamma_h v^2 + vU\right]$ where $U \sim N\left(0, \frac{\gamma_{(1)}S(z_k)}{f(z_k)}\right)$. This function reaches a maximum at $v^* = \frac{-U}{2\Gamma_h}$.

By the argmax theorem, $\widehat{v} \xrightarrow{d} v^*$. By construction, $\widehat{v} = (Tb^{p-1})^{\frac{1}{2}}(\widehat{z}_k - z_k)$ so that $(Tb^{p-1})^{\frac{1}{2}}(\widehat{z}_k - z_k) \xrightarrow{d} N\left(0, \frac{\gamma_{(1)}S(z_k)}{\left[\widetilde{\Delta}_2(z_k) - 2\widetilde{\Delta}_1(z_k)m_1^+(z_k)\right]^2\left[k_{(1)}^-(0)\right]^2 f(z_k)}\right)$. Finally, note that the term $\left[\widetilde{\Delta}_2(z_k) - 2\widetilde{\Delta}_1(z_k)m_1^+(z_k)\right]$ is a linear approximation of Δ_h so that the distribution of $(Tb^{p-1})^{\frac{1}{2}}(\widehat{z}_k - z_k)$ is as postulated by the first part of the theorem.

To prove the second part of the theorem, first note that because $\widehat{v} = O_p(Tb^{p-1})$, we have $(Tb^{p+1})^{\frac{1}{2}}\left[\widehat{\Delta}_h(\widehat{z}_k) - \widehat{\Delta}_h(z_k)\right] = o_p(1)$ so that we can therefore proceed as if the jump points were known.

The results in Delgado and Hidalgo (1996) and the Cramer-Wold device show that:

$$
\left(Tb^{p+1}\right)^{\frac{1}{2}}\begin{pmatrix} \widehat{\widetilde{\Delta}}_1(\widehat{z}_k) - \widetilde{\Delta}_1(z_k) \\ \widehat{\widetilde{\Delta}}_2(\widehat{z}_k) - \widetilde{\Delta}_2(z_k) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\gamma_{(0)}s_1(z_k)}{f(z_k)} & \frac{\gamma_{(0)}H(z_k)}{f(z_k)} \\ \frac{\gamma_{(0)}H(z_k)}{f(z_k)} & \frac{\gamma_{(0)}s_2(z_k)}{f(z_k)} \end{pmatrix}\right).
$$

We can now linearize $(Tb^{p+1})^{\frac{1}{2}}\left(\widehat{\Delta}_h(\widehat{z}_k) - \Delta_h(z_k)\right)$ as before and obtain:

$$
\begin{aligned}
\left(Tb^{p+1}\right)^{\frac{1}{2}}\left(\widehat{\Delta}_h(\widehat{z}_k) - \Delta_h(z_k)\right) &= \left(Tb^{p+1}\right)^{\frac{1}{2}}\left[\widehat{\widetilde{\Delta}}_2(\widehat{z}_k) - \widetilde{\Delta}_2(z_k)\right] \\
&\quad -2m_1^+(z_k)\left(Tb^{p+1}\right)^{\frac{1}{2}}\left[\widehat{\widetilde{\Delta}}_1(\widehat{z}_k) - \widetilde{\Delta}_1(z_k)\right] \\
&\quad +o_p\left(\left(Tb^{p+1}\right)^{\frac{1}{2}}\right)
\end{aligned}
$$

Using the delta method as above gives the desired result:

$$\left(Tb^{p+1}\right)^{\frac{1}{2}}\left(\widehat{\Delta}_h\left(\widehat{z}_k\right)-\Delta_h\left(z_k\right)\right)\xrightarrow{d}N\left(0,\frac{\gamma_{(0)}S\left(z_k\right)}{f\left(z_k\right)}\right).\ \blacksquare$$

A.2 Proof of corollary 2

By definition,

$$\begin{aligned}H\left(z_k\right) &= E\left[u_1^+\left(z_k\right)u_2^+\left(z_k\right)\right]+E\left[u_1^-\left(z_k\right)u_2^-\left(z_k\right)\right]+E\left[u_1^+\left(z_k\right)m_2^+\left(z_k\right)\right]\\ &\quad +E\left[u_1^-\left(z_k\right)m_2^-\left(z_k\right)\right]+E\left[u_2^+\left(z_k\right)m_1^+\left(z_k\right)\right]+E\left[u_2^-\left(z_k\right)m_1^-\left(z_k\right)\right]\\ &= A_1+A_2+A_3+A_4+A_5+A_6\end{aligned}$$

Evaluating each term in turn, we obtain:

$$\begin{aligned}A_1 &= 2m_1^+\left(z_k\right)h^+\left(z_k\right)+h^+\left(z_k\right)^{\frac{3}{2}}E\left(\varepsilon_t^3\right)\\ A_2 &= 2m_1^-\left(z_k\right)h^-\left(z_k\right)+h^-\left(z_k\right)^{\frac{3}{2}}E\left(\varepsilon_t^3\right)\\ A_3 &= 0\\ A_4 &= 0\\ A_5 &= 0\\ A_6 &= 0\end{aligned}$$

Thus,

$$H\left(z_k\right)=2\left[m_1^+\left(z_k\right)h^+\left(z_k\right)+m_1^-\left(z_k\right)h^-\left(z_k\right)\right]+\left[h^+\left(z_k\right)^{\frac{3}{2}}+h^-\left(z_k\right)^{\frac{3}{2}}\right]E\left(\varepsilon_t^3\right)$$

while

$$\begin{aligned}\sigma_{1\pm}^2 &= E\left[u_1^{\pm}\left(z_k\right)^2\right]\\ &= E\left[h^{\pm}\left(z_k\right)\varepsilon_t^2\right]\\ &= h^{\pm}\left(z_k\right)\end{aligned}$$

and

$$\begin{aligned}\sigma_{2\pm}^2 &= E\left[u_2^{\pm}\left(z_k\right)^2\right]\\ &= E\left[4m_1^{\pm}\left(z_k\right)^2h^{\pm}\left(z_k\right)\varepsilon_t^2+4m_1^{\pm}\left(z_k\right)h^{\pm}\left(z_k\right)^{\frac{3}{2}}\varepsilon_t\left(\varepsilon_t^2-1\right)+h^{\pm}\left(z_k\right)^2\left(\varepsilon_t^2-1\right)^2\right]\\ &= 4m_1^{\pm}\left(z_k\right)^2h^{\pm}\left(z_k\right)+4m_1^{\pm}\left(z_k\right)h^{\pm}\left(z_k\right)^{\frac{3}{2}}E\left(\varepsilon_t^3\right)+h^{\pm}\left(z_k\right)^2E\left(\varepsilon_t^4-1\right)\end{aligned}$$

Putting all these pieces together using the fact that $m_1^+(z_k) = m_1^-(z_k) + \widetilde{\Delta}_1(z_k)$ lead to the result:

$$S(z_k) = \left[h^+(z_k)^2 + h^+(z_k)^2\right] E\left(\varepsilon_t^4 - 1\right) - 4\widetilde{\Delta}_1(z_k)\, h^-(z_k)^{\frac{3}{2}}\, E\left(\varepsilon_t^3\right) - 4\widetilde{\Delta}_1(z_k)^2\, h^-(z_k) \;\blacksquare$$

B Auxiliary results

With no loss of generality, we will assume that $v > 0$ in the following propositions. Letting $v < 0$ would change the results by not having terms involving $\widetilde{\Delta}_j(z_k)$. Let $\overline{g}(x) = f(x)g(x)$, $\mathbb{S}^{\pm}(z_k, v_j) = K^{\pm}\left(\frac{\widetilde{Z}_i - \left(z_k + v_j\left(Tb^{p-1}\right)^{-\frac{1}{2}}\right)}{b}\right) - K^{\pm}\left(\frac{\widetilde{Z}_i - z_k}{b}\right)$, $\mathbb{P}_j^{\pm}(v) = \widehat{P}_j^{\pm}\left(z_k + \frac{v}{(Tb^{p-1})^{\frac{1}{2}}}\right) - \widehat{P}_j^{\pm}(z_k)$ and $\mathbb{F}_j^{\pm}(v) = \widehat{f}_j^{\pm}\left(z_k + \frac{v}{(Tb^{p-1})^{\frac{1}{2}}}\right) - \widehat{f}_j(z)_k = \mathbb{F}_i^{\pm}$ for $\forall i, j$

Proposition 4 $E\left[\mathbb{P}_j^{+}(v)\right] = \left[\overline{g}_{(1)}(z_k) + \Delta_j(z_k) f_{(1)}(z_k)\right]\frac{v}{(Tb^{p-1})^{\frac{1}{2}}} + o_p(Tb^{p-1})$.

Proof. Apply proposition 1 of Delgado and Hidalgo (1996). ■

Proposition 5 $E\left[\mathbb{P}_j^{-}(v)\right] = \frac{\Delta_j(z_k)f(z_k)k_{(1)}^{-}(0)v^2}{2Tb^{p-1}} + \overline{g}_{(1)}(z_k)\frac{v}{(Tb^{p-1})^{\frac{1}{2}}} + o_p(Tb^{p-1})$.

Proof. Apply proposition 2 of Delgado and Hidalgo (1996). ■

Proposition 6 $E\left[\mathbb{F}_j^{\pm}(v)\right] = f_{(1)}(z_k)\frac{v}{(Tb^{p-1})^{\frac{1}{2}}} + o_p(Tb^{p-1})$.

Proof. Apply proposition 3 of Delgado and Hidalgo (1996). ■

Proposition 7 $cov\left[\mathbb{P}_1^{\pm}(v_1)\mathbb{P}_2^{\pm}(v_2)\right] = \left(Tb^{p+1}\right)^{-2} f(z_k)v_1v_2\gamma_{(1)}\{m_1^{\pm}(z_k)m_2^{\pm}(z_k) +$
$E\left[u_1^{\pm}(z_k)u_2^{\pm}(z_k)\right] + E\left[u_1^{\pm}(z_k)m_2^{\pm}(z_k)\right] + E\left[u_2^{\pm}(z_k)m_1^{\pm}(z_k)\right] +$
$o_p\left(\left(Tb^{p+1}\right)^{-2}\right)$

Proof. Expand the left-hand side as:

$$\begin{aligned}
cov\left[\mathbb{P}_1^{\pm}(v_1)\mathbb{P}_2^{\pm}(v_2)\right] &= cov\left[\widehat{P}_1^{\pm}(v) - \widehat{P}_1^{\pm}(0), \widehat{P}_2^{\pm}(v) - \widehat{P}_2^{\pm}(0)\right] \\
&= \frac{1}{\left(Tb^{p+1}\right)^2} cov\left[\sum_i \mathbb{S}^{\pm}(Z_i, v_1)y_i, \sum_j \mathbb{S}^{\pm}(Z_j, v_2)y_j^2\right] \\
&= \frac{2}{\left(Tb^{p+1}\right)^2}\sum_i\sum_{j>i} cov\left[\mathbb{S}^{\pm}(Z_i, v_1)y_i, \mathbb{S}^{\pm}(Z_j, v_2)y_j^2\right] \\
&\quad + \frac{1}{\left(Tb^{p+1}\right)^2}\sum_i cov\left[\mathbb{S}^{\pm}(Z_i, v_1)y_i, \mathbb{S}^{\pm}(Z_i, v_2)y_i^2\right] \\
&= \frac{2}{\left(Tb^{p+1}\right)^2}\sum_i\sum_{j>i} cov\left[\mathbb{S}^{\pm}(Z_i, v_1)m_1^{\pm}(Z_i), \mathbb{S}^{\pm}(Z_j, v_2)m_2^{\pm}(Z_j)\right]
\end{aligned}$$

$$
\begin{aligned}
& +\frac{2}{\left(Tb^{p+1}\right)^2}\sum_i\sum_{j>i} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)u_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_j,v_2\right)u_2^{\pm}\left(Z_j\right)\right] \\
& +\frac{2}{\left(Tb^{p+1}\right)^2}\sum_i\sum_{j>i} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)m_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_j,v_2\right)u_2^{\pm}\left(Z_j\right)\right] \\
& +\frac{2}{\left(Tb^{p+1}\right)^2}\sum_i\sum_{j>i} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)u_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_j,v_2\right)m_2^{\pm}\left(Z_j\right)\right] \\
& +\frac{T}{\left(Tb^{p+1}\right)^2} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)m_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_i,v_2\right)m_2^{\pm}\left(Z_i\right)\right] \\
& +\frac{T}{\left(Tb^{p+1}\right)^2} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)u_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_i,v_2\right)u_2^{\pm}\left(Z_i\right)\right] \\
& +\frac{T}{\left(Tb^{p+1}\right)^2} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)m_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_i,v_2\right)u_2^{\pm}\left(Z_i\right)\right] \\
& +\frac{T}{\left(Tb^{p+1}\right)^2} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)u_1^{\pm}\left(Z_i\right),\mathbb{S}^{\pm}\left(Z_i,v_2\right)m_2^{\pm}\left(_i\right)\right] \\
= \; & A_1+A_2+A_3+A_4+A_5+A_6+A_7+A_8
\end{aligned}
$$

Using lemma 5 in Delgado and Hidalgo (1996), A_1, A_6, A_7, and A_8 are all $o_p\left(\left(Tb^{p+1}\right)^2\right)$. On the other hand, using their lemma 7, we obtain:

$$
\begin{aligned}
A_2 &= \frac{1}{\left(Tb^{p+1}\right)^2} f\left(z_k\right)E\left[m_1^{\pm}\left(z_k\right)m_2^{\pm}\left(z_k\right)\right]v_1v_2\gamma_{(1)}+o_p\left(\left(Tb^{p+1}\right)^{-2}\right) \\
A_3 &= \frac{1}{\left(Tb^{p+1}\right)^2} f\left(z_k\right)E\left[u_1^{\pm}\left(z_k\right)u_2^{\pm}\left(z_k\right)\right]v_1v_2\gamma_{(1)}+o_p\left(\left(Tb^{p+1}\right)^{-2}\right) \\
A_4 &= \frac{1}{\left(Tb^{p+1}\right)^2} f\left(z_k\right)E\left[u_1^{\pm}\left(z_k\right)m_2^{\pm}\left(z_k\right)\right]v_1v_2\gamma_{(1)}+o_p\left(\left(Tb^{p+1}\right)^{-2}\right) \\
A_5 &= \frac{1}{\left(Tb^{p+1}\right)^2} f\left(z_k\right)E\left[m_1^{\pm}\left(z_k\right)u_2^{\pm}\left(z_k\right)\right]v_1v_2\gamma_{(1)}+o_p\left(\left(Tb^{p+1}\right)^{-2}\right)
\end{aligned}
$$

Combining these results completes the proof. ■

Proposition 8 $cov\left[\mathbb{P}_1^{\pm}\left(v_1\right)\mathbb{P}_2^{\mp}\left(v_2\right)\right]=o_p\left(\left(Tb^{p+1}\right)^{-2}\right)$

Proof. Expand the left-hand side as:

$$
\begin{aligned}
cov\left[\mathbb{P}_1^{\pm}\mathbb{P}_2^{\mp}\right] &= \frac{1}{\left(Tb^{p+1}\right)^2} cov\left[\sum_i \mathbb{S}^{\pm}\left(Z_i,v_1\right)y_i,\sum_j \mathbb{S}^{\mp}\left(Z_j,v_2\right)y_j^2\right] \\
&= \frac{2}{\left(Tb^{p+1}\right)^2}\sum_i\sum_{j>i} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)y_i,\mathbb{S}^{\mp}\left(Z_j,v_2\right)y_j^2\right] \\
&\quad +\frac{T}{\left(Tb^{p+1}\right)^2} cov\left[\mathbb{S}^{\pm}\left(Z_i,v_1\right)y_i,\mathbb{S}^{\mp}\left(Z_i,v_2\right)y_i^2\right]
\end{aligned}
$$

which is $o_p\left(\left(Tb^{p+1}\right)^{-2}\right)$ by using lemma 6 from Delgado and Hidalgo (1996) for the first term and their lemma 8 for the second term. ■

Proposition 9 $cov\left[m_j^{\pm}(z_k)\,\mathbb{F}_j^{\pm}(v_1)\,\mathbb{P}_i^{\pm}(v_2)\right] = \left(Tb^{p+1}\right)^{\frac{1}{2}} f(z_k)\,v_1 v_2 \gamma_{(1)} m_2^{+}(z_k)\,m_1^{+}(z_k) + o_p\left(\left(Tb^{p+1}\right)^{-2}\right)$

Proof. Direct from Proposition 6 in Delgado and Hidalgo (1996). ■

Proposition 10 $cov\left[m_j^{\pm}(z_k)\,\mathbb{F}_j^{\pm}(v_1)\,\mathbb{P}_i^{\mp}(v_2)\right] = o_p\left(\left(Tb^{p+1}\right)^{-2}\right)$

Proof. Direct from Proposition 7 in Delgado and Hidalgo (1996). ■

Proposition 11 $cov\left[m_1^{\pm}(z_k)\,m_2^{\pm}(z_k)\,\mathbb{F}_1^{\pm}(v_1)\,\mathbb{F}_2^{\pm}(v_2)\right] = \frac{m_1^{\pm}(z_k) m_2^{\pm}(z_k) f(z_k) v_1 v_2 \gamma_{(1)}}{\left(Tb^{p+1}\right)^2} + o_p\left(\left(Tb^{p+1}\right)^{-2}\right)$

Proof. Direct from Proposition 5 in Delgado and Hidalgo (1996). ■

Table 1. Parameter values for simulation experiment

Values estimated from S&P 500 returns,

January 2nd, 1980-December 31st, 1996

	DGP 1	DGP 2
μ ($\times 100$)	0.0494	0.0463
Δ_0 ($\times 10^5$)	9.0109	8.1226
Δ_1 ($\times 10^5$)	0	-5.9173
τ_1	-	0.7046
Δ_2 ($\times 10^5$)	0	0.1829
τ_2	-	0.3626
α	0	0.0684
β	0	0.8886

Table 2. Frequency distribution of estimated number of jumps

$h_t = 9.0109 \times 10^{-5}$

True number of jumps is 0

$T = 1000,\ 500$ replications

	$p = 0$				$p = 1$				$p = 2$				$p = 3$			
	$c = 0.8$	1	1.2	cv	$c = 0.8$	1	1.2	cv	$c = 0.8$	1	1.2	cv	$c = 0.8$	1	1.2	cv
$\widehat{M} = 0$	100.0	100.0	100.0	100.0	11.2	39.2	61.8	69.2	0.0	2.6	31.6	32.6	0.0	0.6	16.0	17.4
1	0.0	0.0	0.0	0.0	20.8	23.2	20.4	23.0	1.6	20.4	28.0	38.0	0.0	9.4	32.2	34.8
2	0.0	0.0	0.0	0.0	28.0	23.0	14.8	6.8	16.2	55.4	37.0	25.8	25.0	72.4	50.4	46.4
3	0.0	0.0	0.0	0.0	23.0	12.6	3.0	1.0	66.8	21.6	3.4	3.6	70.4	17.6	1.4	1.4
4	0.0	0.0	0.0	0.0	15.6	2.0	0.0	0.0	15.4	0.0	0.0	0.0	4.6	0.0	0.0	0.0
≥ 5	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Incomplete	0.0	0.0	0.0	0.0	25.0	15.8	11.0	4.0	81.6	55.5	33.8	22.2	90.6	76.2	51.4	44.4
Mean	0.000	0.000	0.000	0.000	2.152	1.150	0.590	0.396	2.960	1.960	1.122	1.004	2.796	2.070	1.372	1.318

Table 3. Frequency distribution of estimated number of jumps

$$h_t = \omega_t(1-\alpha-\beta) + 0.0684(y_{t-1}-\mu)^2 + 0.8886h_{t-1}$$

$$\omega_t = \Delta_0 + \Delta_1 1\left(\tfrac{t}{T} \geq \tau_1\right) + \Delta_2 1\left(\tfrac{t}{T} \geq \tau_2\right)$$

True number of jumps is 2

$T = 1000,$ 500 replications

	$p=0$				$p=1$				$p=2$				$p=3$			
	$c=0.8$	1	1.2	*cv*	$c=0.8$	1	1.2	*cv*	$c=0.8$	1	1.2	*cv*	$c=0.8$	1	1.2	*cv*
$\widehat{M}=0$	53.0	46.6	27.6	28.6	0.8	1.2	0.2	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.2
1	13.0	23.8	45.8	43.4	8.4	13.6	20.8	19.0	2.0	12.2	20.4	21.6	1.6	10.2	33.8	32.4
2	5.4	6.4	12.8	12.0	13.2	30.8	52.2	53.8	16.6	53.8	64.0	66.6	33.8	74.6	63.0	62.6
3	6.2	14.8	13.4	8.2	32.2	45.2	26.8	25.8	58.4	33.6	15.6	11.2	59.6	15.2	3.2	4.8
4	10.2	8.2	0.4	4.8	40.4	9.2	0.0	1.0	23.0	0.4	0.0	0.2	5.0	0.0	0.0	0.0
≥ 5	12.2	0.2	0.0	3.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Incomplete	15.4	16.4	14.2	7.6	63.0	61.6	64.6	66.8	80.2	75.2	76.6	72.2	88.8	83.2	73.6	75.4
Mean	1.442	1.148	1.132	1.262	3.180	2.476	2.056	2.080	3.024	2.222	1.952	1.892	2.680	2.050	1.694	1.720

Table 4. Results from estimation of jumps in conditional variance

S&P 500 returns, January 2nd, 1980-December 31st, 1996

Bandwidth constant chosen by cross-validation

$p = 0$			1			2			3		
Date	Size ($\times 10^7$)	p-value	Date	Size ($\times 10^4$)	p-value	Date	Size ($\times 10^4$)	p-value	Date	Size ($\times 10^3$)	p-value
861218 (0.4099)	6.878	0.001	860707 (0.3829)	32.786	0.000	871127 (0.4652)	129.787	0.000	900803 (0.6229)	25.152	0.000
890307 (0.5399)	-0.452	0.013	900823 (0.6262)	-20.709	0.000	911115 (0.6988)	-0.038	0.012	830121 (0.1800)	-3.358	0.421
830617 (0.2038)	2.712	0.536	930216 (0.7720)	4.099	0.000	820819 (0.1549)	0.184	0.358			
			821122 (0.1703)	-0.066	0.184						
$c =$	1.2			0.9			1.2			1.1	

Fig. 1. Daily Returns on S&P 500 index
1980–1996

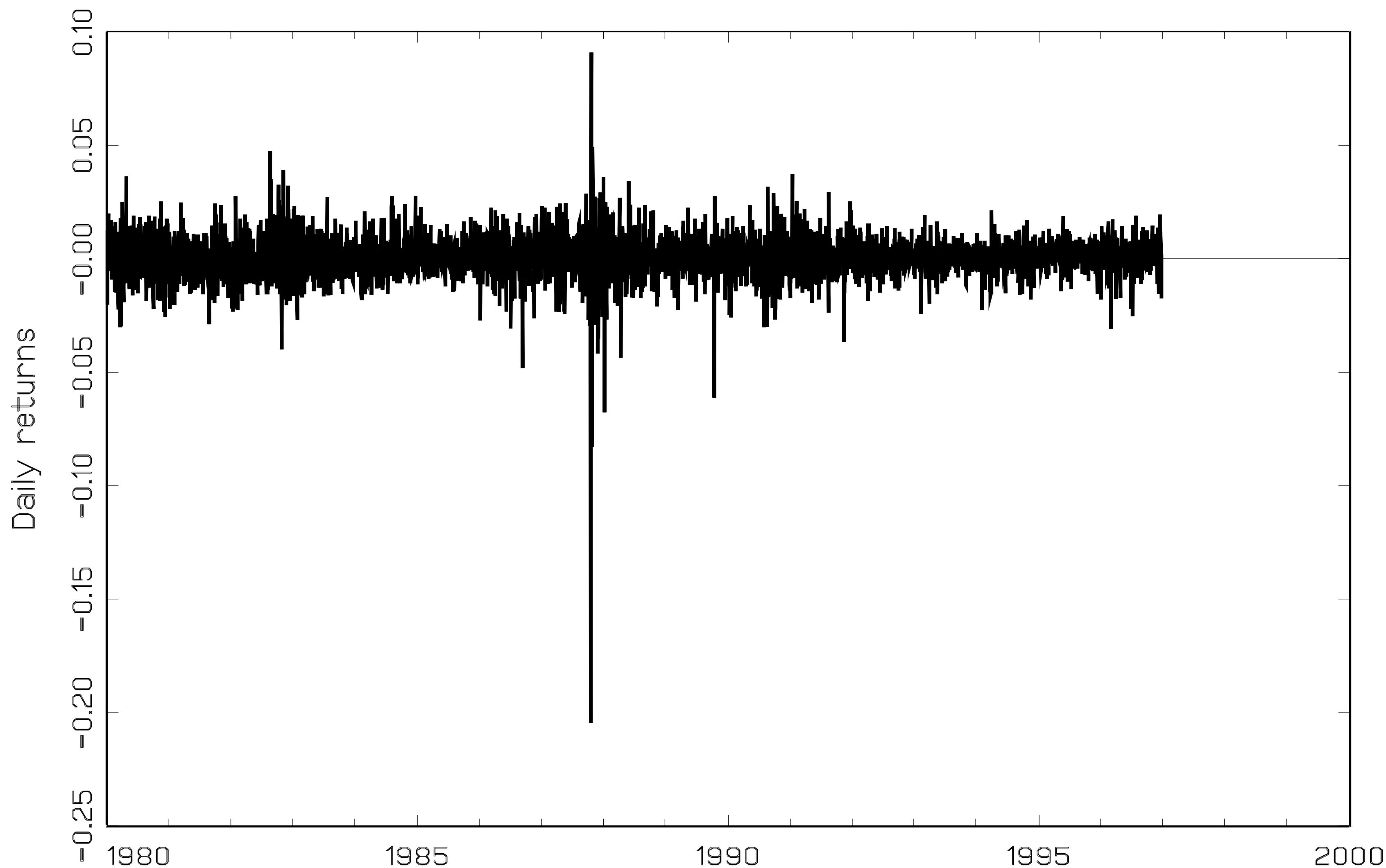
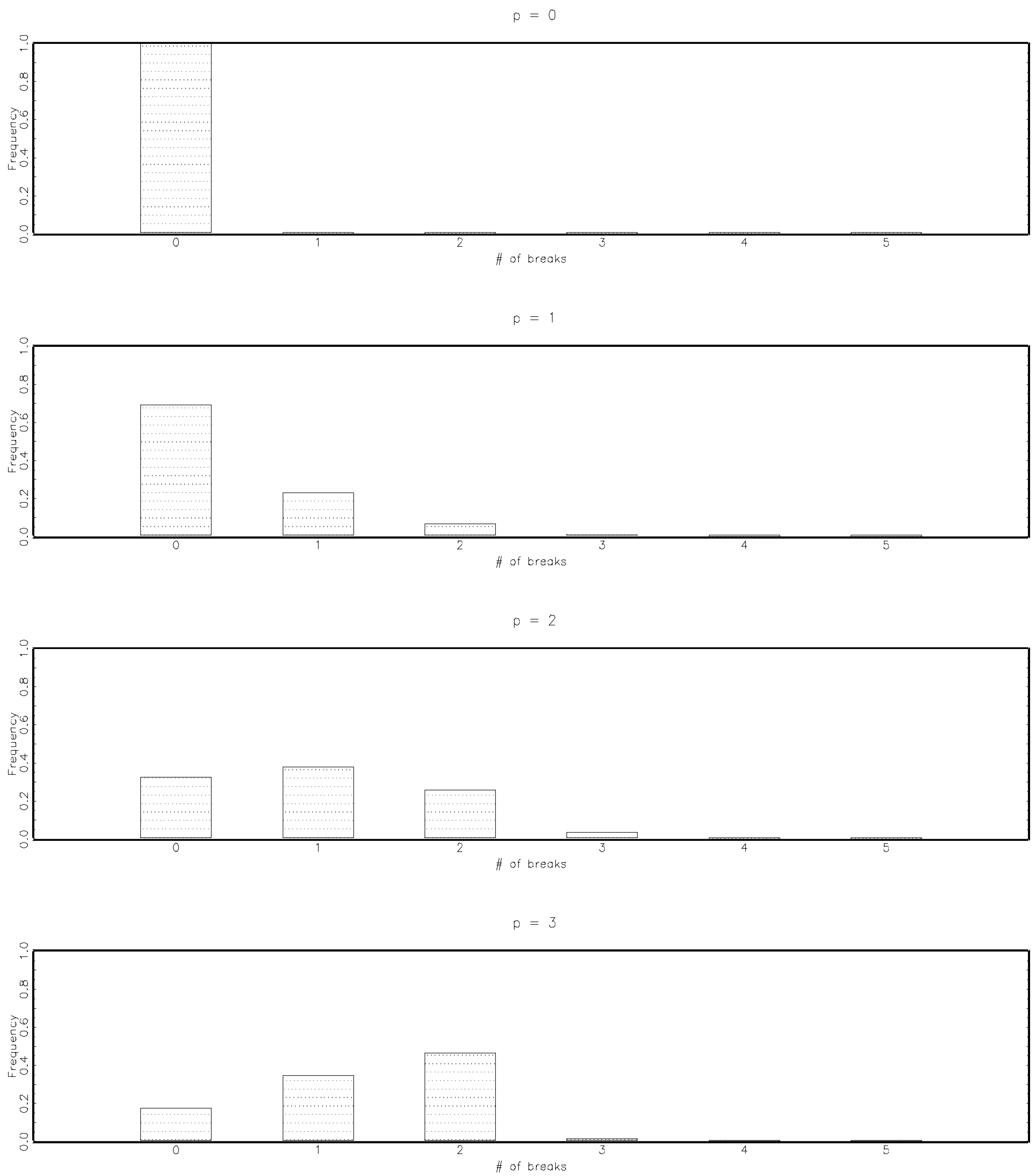


Fig 2. Estimated number of breaks
Number of breaks = 0
c chosen by CV



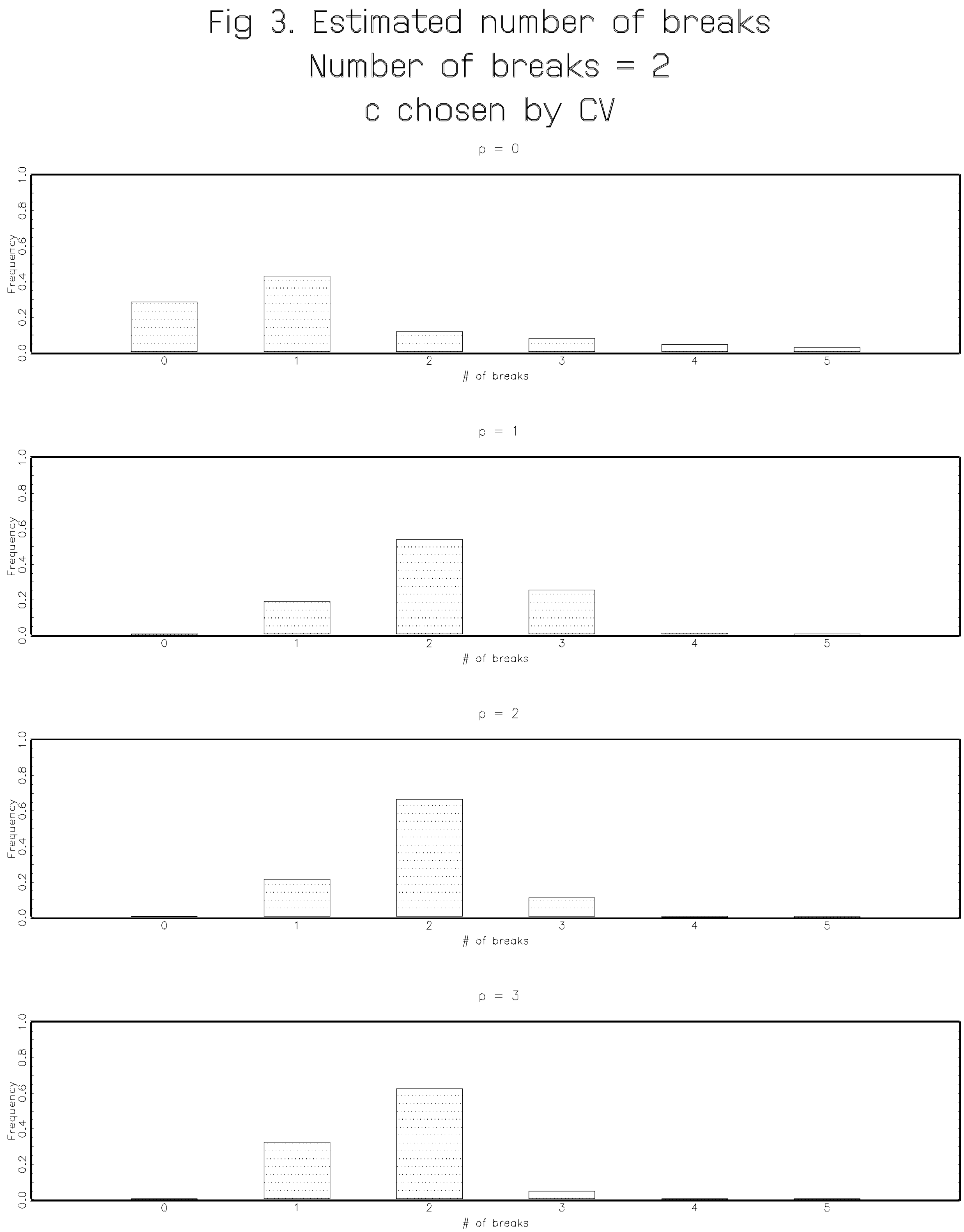


Fig 3. Estimated number of breaks
Number of breaks = 2
c chosen by CV
p = 0
p = 1
p = 2
p = 3
Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
# of breaks

Fig. 4. Distribution of estimates of jump locations
c chosen by CV