Estimation procedure for a hidden Markov chain model with applications to finance, climate data and earthquake analysis

Ionuţ Florescu and Forrest Levin
Dept. of Mathematical Sciences, Stevens Institute of Technology,
Hoboken NJ 07030, USA
ifloresc@stevens.edu
flevin@stevens.edu

Abstract

In this work we present a methodology for estimating the variability of a signal modeled as a continuous time stochastic process observable only at discrete times with variability modeled by a continuous time Markov chain. The methodology estimates the parameters of the hidden Markov chain. The methodology is new however the major contribution of this work comes in the realm of applications. The methodology is applied to three domains: Finance (high frequency data), Climate studies (temperature data) and Geophysics (earthquake data). In all the applications the methodology provides insight into features of the signal which are hard to detect otherwise.

Keywords: stochastic volatility; hidden Markov chain; financial crisis; climate change; earthquake analysis.

1 Introduction

Suppose a stochastic signal evolving continuously in time may only be observed at discrete times. The most important feature of the observed signal is a change in

variability regimen which happens at random times. The purpose of the present work is to present a new methodology to completely characterize the behavior of this type of variability of the signal.

We then apply the methodology developed here to three areas of application. In all cases we find hidden features of the original signal process which are not obvious from just studying the signal. We believe that showing the modeling power of even such a simple model of variability provides incentive into further studying more complex stochastic volatility models and their applicability to other areas.

The structure of the paper is as follows. The remainder of this section gives an overview of the model and existing literature. Section 2 describes the main methodology employed in this paper. Section 3 is the most important part of the paper. In this section we present the results and conclusions obtained when using the estimating methodology with real data from three different areas: Finance (section 3.1), Climatology (3.2) and Geophysics (3.3). Section 4 concludes the article. The Appendix A at the end of the paper presents theoretical convergence results as well as empirical estimation results for generated signal with known variability.

1.1 The original motivation

The methodology developed in this article may be applied to any number of areas of science and technology. However, the original motivation of the work comes from the world of Finance. Today, it is well accepted in the financial literature that the Black-Scholes model (Black and Scholes (1973)) is not complex enough to capture the dynamics of data sampled with high frequency (for example see Mariani et al. (2009)).

Various other, more general models, have been proposed among them: stochastic volatility models Shephard (2005), jump diffusion models Merton (1992) and general Lévy models Cont and Tankov (2003).

In this article we present a model where the variability of the signal is driven by a continuous time Markov chain. The methodology estimates the parameters of this chain by calibrating to real signals. In the financial literature the model we analyze here is sometimes called a regime switching model, or a stochastic volatility model with volatility driven by a hidden Markov chain.

1.2 The model and the problem

We assume we observe a 1-dimensional signal S at discrete moments in time $S_{t_1}, \ldots, S_{t_T}$. We assume as given a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. On this space, we are given a complete filtration $\mathscr{F} = \{\mathscr{F}_t\}_t$ (see Protter, 2005, page 3). The continuous time evolution of the signal S is described by the strong solution of the following stochastic differential equation:

$$S_t = S_0 + \int_0^t r S_u du + \int_0^t S_u Y_u dW_u, \tag{1}$$

where r is known, W_t is a standard Brownian motion with respect to the filtration $\mathscr{F}$ defined on the original probability space, and Y_t is modeled as a continuous time Markov Chain. We use the classical definition of a continuous Markov chain. The chain has finite state space $\{a_1, \ldots, a_p\}$, transition probability matrix $\Lambda = (\lambda_{ij})$ and

the transition times given by exponential random variables with parameter λ_i, i.e.

$$\mathbf{P}(Y_s = i, \forall s \in (t, t+u] | Y_t = i) = e^{-\lambda_i u}.$$

The times spent in different states are independent, exponentially distributed with rate λ_i dependent on the particular state i; λ_{ij} is the departing intensity rate from i to j. Both processes S and Y are adapted to the filtration $\mathscr{F}$. We note that the process Y is not directly observable, this process will be referred as the "hidden factor" or the "hidden Markov process".

The problem we have solving in this work is the estimation of the number of states p, the values of the states $\{a_1, \ldots, a_p\}$, as well as the transition rates λ_i and the transition probability matrix Λ. Most importantly, since we wish to apply this model to fields other than Finance we do not have any other input than a history of the process S. No derivative data (options, futures etc) is available to us.

1.3 A brief historical note

The history of analyzing the model presented in (1) comes from the Finance literature. The problem of parameter estimation was studied before, most notably by Cvitanić, Liptser, and Rozovskii (2006) and Cvitanić, Rozovskii, and Zaliapin (2006). The first article describes the estimation theory while the second implements the theory and presents numerical results. The current work distinguishes itself from the previous citations. While the model used is similar (our Markov chain is defined differently) the estimation methodology we propose is different. Specifically, the papers cited

above use a sequential Bayes methodology designed for the particular model under consideration. This allows the authors to obtain specific convergence rates for their algorithm. However, the methodology cannot be generalized beyond the model under consideration. Furthermore, the analysis devoted to the number of states and the specific values of the states of the hidden Markov chain is somewhat ad-hoc and is performed by visual inspection of the results given by the Bayesian filter.

A second approach which we regard related to our particular methodology is provided by Genon-Catalot, Jeantheau, and Larédo (2000). Culminating a series of three different papers dedicated to parameter estimation for Stochastic Volatility models, in the final article the authors use a filtering methodology similar with our approach to a more general model (the process Y_t is a stationary continuous time process). However, in contrast with our estimation method the authors use a method of moments estimation for the parameters which would be cumbersome to apply in our particular context of the hidden Markov chain.

We also mention earlier work on estimating parameters of stochastic volatility models using discrete data, even though these papers are not directly using the model presented in this work they were a source of inspiration. Nielsen and Vestergaard (2000), Sørensen (2003), Bladt and Sørensen (2007), Aït-Sahalia and Kimmel (2007) are but a few of these important references.

2 The methodology

The theoretical foundation of the estimating methodology is extending fundamental work in Del Moral, Jacod and Protter Del Moral et al. (2001). We use a particle filtering methodology to estimate the empirical distribution at time t of the volatility process $\{Y_t\}_t$. Then we use the theoretical distribution form to find optimal parameter values which allows the best match with the empirical estimated distribution. We note that due to the specific form of the process S_t in (1) we may consider the process $X_t = \log S_t$. This process has a simplified dynamic as an application of the Itô's lemma shows:

$$X_t = X_0 + \int_0^t \left(r - \frac{Y_u^2}{2} \right) du + \int_0^t Y_u dW_u. \tag{2}$$

The Y_t process remains unmodified under this transformation and the observations are now $\{x_{t_1}, \ldots, x_{t_T}\}$ with $x_{t_i} = \log s_{t_i}$.

2.1 Obtaining filtered empirical distributions at $t_1, \ldots, t_T$

Suppose that we are looking at the process evolution over the interval $[t_{i-1}, t_i)$. We are given the two endpoints $x_{t_{i-1}}, x_{t_i}$, and a previous estimate of the distribution of the volatility process $Y_{t_{i-1}}$, denoted Φ_{i-1}. We want to calculate an estimate of the distribution of volatility process at t_i. To this end we need the following basic elements:

- A weighting function $\varphi(\cdot)$. This would be called a kernel function in other areas of statistics (any nonnegative function with L^1 norm equal to 1), but for

practical consideration it has to have most of the mass centered around zero. The kernel does not need to be symmetric, though in practical applications it always is (we do not know if we should penalize more under or overestimates). In our applications we typically use one of the triangular, Epanechnikov or Gaussian kernels. The choice has little effect on the estimated distribution.

- A number n that represents the number of intermediate paths for the filter. This number is typically large.

- A number m which represents the number of intermediate points within the interval $[t_{i-1}, t_i)$. For theoretical convergence reasons we need to have $m \geq \sqrt[3]{n}$.

With these basic elements the filtering methodology consists in two steps: an evolution (mutation) step and a selection step.

Evolution step In this step of the algorithm we create n paths from t_{i-1} to t_i using the following process evolution. Take $\Delta_{i,m} = \frac{t_i - t_{i-1}}{m}$, and draw a value Y_{i-1} from the previously estimated distribution Φ_{i-1}. Each evolution path j is then constructed:

$$\begin{cases} \tilde{X}^j_{t_{i-1}+k\Delta_{i,m}} = \tilde{X}^j_{t_{i-1}+(k-1)\Delta_{i,m}} + & \left(r - \frac{(\tilde{Y}^j_{t_{i-1}+(k-1)\Delta_{i,m}})^2}{2}\right)\Delta t \\ & \quad + \tilde{Y}_{t_{i-1}+(k-1)\Delta_{i,m}} \Delta W^j_k \\ \tilde{Y}^j_{t_{i-1}+k\Delta_{i,m}} = \tilde{Y}^j_{t_{i-1}+(k-1)\Delta_{i,m}} + & \beta \Delta Z^j_k, \quad k \in \{1, \ldots, m\} \end{cases} \tag{3}$$

where ΔW^j_k and ΔZ^j_k represent the increments of independent Brownian motions (i.e., they are i.i.d. $N(0, \Delta^2_{i,m})$). β is a parameter of the algorithm. At the end of each evolution path we keep the pairs $(\tilde{X}^j_{t_i}, \tilde{Y}^j_{t_i})$, $j = 1, \ldots, n$.

Selection step The idea of this step is to give greater weights to the "better" values obtained. Specifically, we assign a weight to each final value $\tilde{Y}_{t_i}^j$ equal to:

$$w_j = C(n)\sqrt[3]{n}\ \ \varphi\left(|\tilde{X}_{t_i}^j - x_{t_i}|\sqrt[3]{n}\right),$$

where $C(n)$ is a normalizing constant which makes the weights sum to 1.

Our approximating distribution at the end of step i is the discrete distribution $(\tilde{Y}_{t_i}^j, w_j)_{j\in\{1,\ldots,n\}}$. We denote this distribution Φ_i. One may view this part of the scheme as an importance sampling technique.

2.2 Obtaining the parameters of the Markov Chain

Once we have the filtered distributions $\Phi_1, \ldots, \Phi_T$ obtained in the previous subsection we calculate the means for each of them $\bar{Y}_1, \ldots, \bar{Y}_T$ and we take these as realizations of the process Y_t. The next step is straightforward. The values $\bar{Y}_1, \ldots, \bar{Y}_T$ represent realizations from a mixture of Gaussian distributions with means precisely $a_1, \ldots, a_p$. Thus, we may use one of the established methods to estimate the nature of the means; for example the Expectation Maximization (EM) algorithm. However, the EM algorithm requires knowledge of the number of distributions present and it is slow so we prefer a simpler method. We follow the Minimum Error Thresholding method of Kittler and Illingworth as improved by Cho, Haralick, and Yi (1989). This method assumes that the best separating values for the mixture in the distribution are at $T_1, \ldots, T_{p-1}$. Given this assumption the method then calculates the likelihood of seeing the values $\bar{Y}_1, \ldots, \bar{Y}_T$. Thus, the method creates a real valued $p-1$ dimensional

function that needs to be maximized. The maximization may be performed using any optimization method. Under the hypothesis of multivariate normal distribution for Y the function to be maximized has a global maximum. The result of the maximization is the set of optimal threshold values $\hat{T}_1, \ldots, \hat{T}_{p-1}$. Once these are obtained the algorithm uses a truncated normal methodology to obtain estimates of the means for each distribution in the mixture. These are our estimates for $a_1, \ldots, a_p$. The value of p is chosen such that it corresponds to the largest Maximum likelihood.

Finally, to obtain the rates λ_i and the transition probability matrix Λ we use the estimated $\hat{a}_1, \ldots, \hat{a}_p$ and the thresholds $\hat{T}_1, \ldots, \hat{T}_{p-1}$. We go back to the estimated sequence $\bar{Y}_1, \ldots, \bar{Y}_T$ and we link them by grouping the values that are within the estimated thresholds $\hat{T}_1, \ldots, \hat{T}_{p-1}$ and assigning the corresponding a_i value. This allows to identify the time periods when the Markov chain stays constant as well as the destination of subsequent jumps.

We present convergence results both theoretical and empirical in the Appendix A. We feel that these results, while demonstrating the algorithm provides correct results, will detract from the principal merit of the paper which is the application to real data in the next section.

3 Results obtained applying the model to real data

3.1 Part I: Financial applications

Volatility has important applications in financial situations. Volatility provides a way to measure the risk associated with a financial asset. This has been known for

a long time in Mathematical Finance and indeed estimating volatility has a special place in the area either using historical data (realized volatility, integrated volatility, heteroskedastic models Tsay (2005); Shephard (2005)) or interpolating values computed using financial derivatives written on the stock (implied volatility, local volatility surfaces etc.). However, despite the long history, in practice there are irreconcilable differences between the values obtained using historical values and on the spot derivative values. Traditionally, the difference is explained using the time paradigm – implied volatility reflect present risk values while the historical volatility represents risks from the past and including the present. Yet, the current paper offers an alternative explanation. There are shifts in volatility, these shifts are captured using the methodology presented in this article. The estimated node values generally shift from traditional implied volatility levels to historical volatility levels during a normal day. This would explain why the estimation methods do not agree on their estimated values.

The financial data we chose to illustrate the method is from March 2008. Bear Stearns began the week on Monday, March 10, 2008 with a market stock price of about \$70/share. The 85 year old firm, the 5th largest Wall Street securities firm at the time, had never recorded a quarterly loss until the fourth quarter of the previous year (2007). They were anticipating a profit for the first quarter report for 2008. At the end of the week (Friday March 14, 2008) the share price had dropped to \$30 and over the weekend Bear Stearns, JPMorgan and the Federal Reserve Bank of New York arranged for J.P. Morgan to Purchase Bear Stearns for \$2/share to avoid a bankruptcy filing [source for general information: Cohan (2009)]

How did an apparently very healthy firm get in this situation? We do not know and perhaps will never know without analyzing the internal balance of the company. However, this is a perfect setting for testing our model. By getting closer to the conclusion of that week the market should become more and more volatile and consequently the volatility should perform a shift in magnitude. Will we be able to capture this shift (provided it exists)? Will we be able to determine an exact moment when the crisis had started? Were other companies affected by the collapse? Will we be able to determine all this from the analysis of stock data alone? All these are questions that we hoped to answer when we started the analysis of the minute data.

3.1.1 The events as they happened

At the time of the events mentioned, there had already been a large bankruptcy related to companies involved in mortgage instruments: the buy out of Countrywide Financial, "the nation's largest mortgage lender" by Bank of America early in January 2008 (Landon and Sorkin, 2008; Ross Sorkin, 2008; Dash, 2008, – NY Times articles). Many investment banks and other corporations either large or small, and including government agencies were in the volatile sub prime market. From the market facts and the media attention both the corporate sector and the investing public were aware of the sub-prime mortgage loan crisis.

The people in charge of the business at Bear Stearns (including those working at the time for the company) knew there was trouble with funding, overnight loans and liquidity in general from the start of the week of the collapse [source: undisclosed former employee]. Throughout the investment industry there were warning signals,

in particular home mortgage defaults [Geithner (2008)]

As late as Tuesday March 11 the stock market was heading up after the Federal Reserve opened up its lending policies to investment companies in addition to just commercial banks (which was the previous practice). On the evening of Thursday March 13, Bear Stearns executives were meeting with SEC regulators and Tim Geithner (president of NY Fed). The president of the Federal Reserves Bank of NY expressed the opinion that Bear Stearns would have to file for bankruptcy (source: U.S. Senate Committee on Banking, Housing and Urban Affairs (2008). We assume the market would react on Friday to this news so it is interesting to analyze the volatility behavior.

3.1.2 Data analysis and results

In Figure 1 we may observe the behavior of the Bear Stearns volatility during the week of March 10-March 14 and continuing with the following two days (trading did not stop after the weekend since the company still existed although under different management). The estimated states for the volatility of the Bear Stearns stock are 1.9, 5.4, and 6.8. All these values are outside the normal $[0, 1]$ annual volatility values generally characteristic for the stock market. Furthermore, clearly the volatility levels shift upward toward the end of the week starting as early as the middle of Tuesday and certainly after the Thursday meeting.

We recognized that the high volatility values may be due to the Monday (March 17) and Tuesday (March 18) following the buy-out announcement. During those days the stock traded at \$3-\$4 range even though the announcement was that shares will be

bought by JPMorgan at $2/share. The explanation for the high trading activity came on March 24, 2008 when a class action lawsuit was filed on behalf of shareholders, challenging the terms of JPMorgans recently announced acquisition of Bear Stearns. That same day, a settlement was reached that raised JPMorgan Chase's offer to $10 a share.

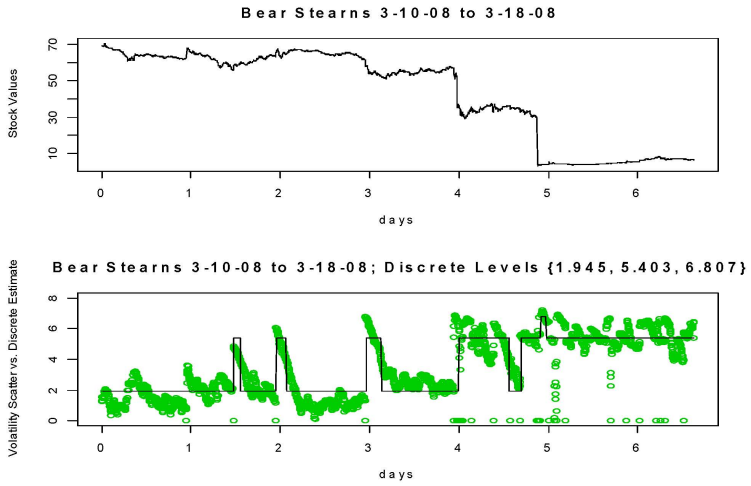


Figure 1: Bear Stearns Stock Price/Volatility March 10 to March 18, 2008

Recognizing the situation we had eliminated the days after the buy-out and have redone the analysis separately for each of the three weeks preceding March 14 (see Figures 2- 4 on page 14). From these pictures we see that the volatility levels (~ 0.59 and ~ 0.86) for the two weeks preceding the critical week are similar. We see that the levels during the final week are all higher than 1 (1.1, 2.4, and 5.2) so the whole week behavior is abnormal. The second week does exhibit prolonged periods of higher volatility so we decided to pull the final two weeks (March 3- March 14, 2008) together for a final analysis in Figure 5. In this figure we can finally see that abnormal levels

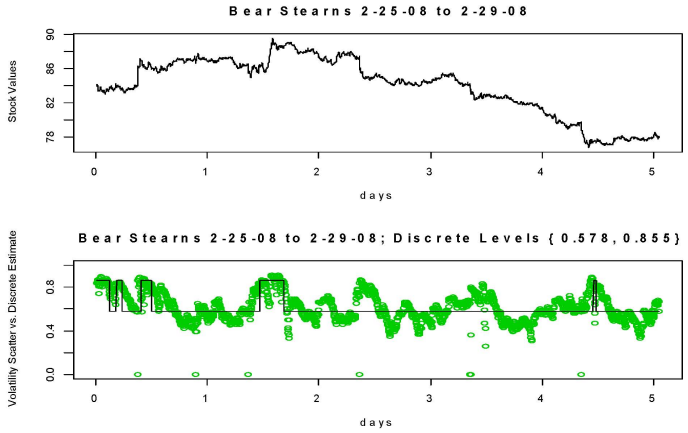


Figure 2: Bear Stearns Stock Price/Volatility two weeks before collapse

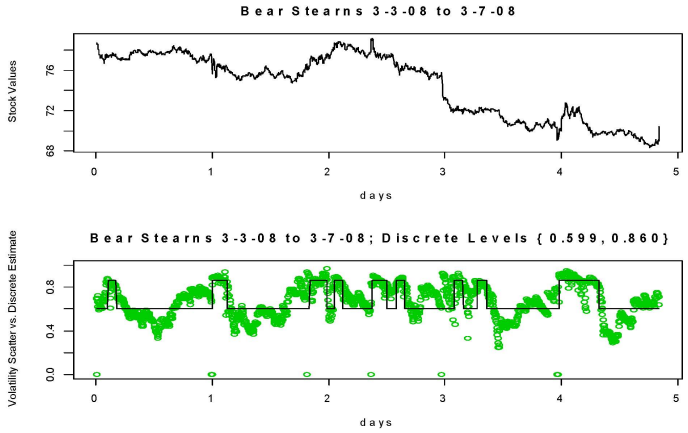


Figure 3: Bear Stearns Stock Price/Volatility one week before the collapse

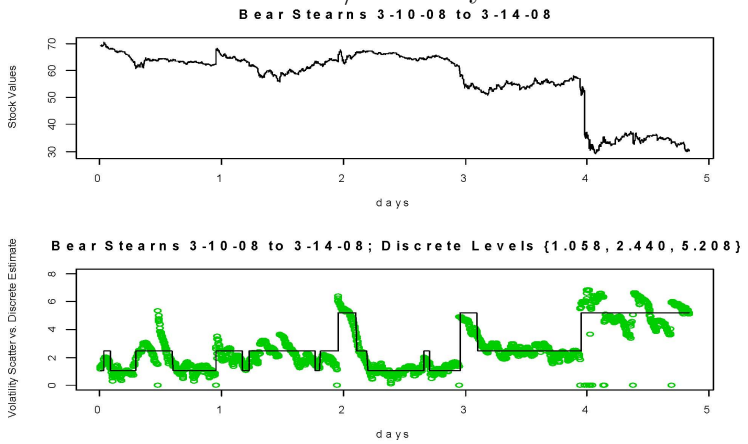


Figure 4: Bear Stearns Stock Price/Volatility week of the collapse

were touched in the beginning and end of day 4 (Friday March 7) and certainly during the entire next week (as we had observed from the earlier analysis). Furthermore, we do observe lowered values for volatility during the Wednesday March 12, this is the day before the SEC meeting with Bear Stearns executives and we suspect from this behavior that the bankruptcy announcement came as a surprise to the general market participants.

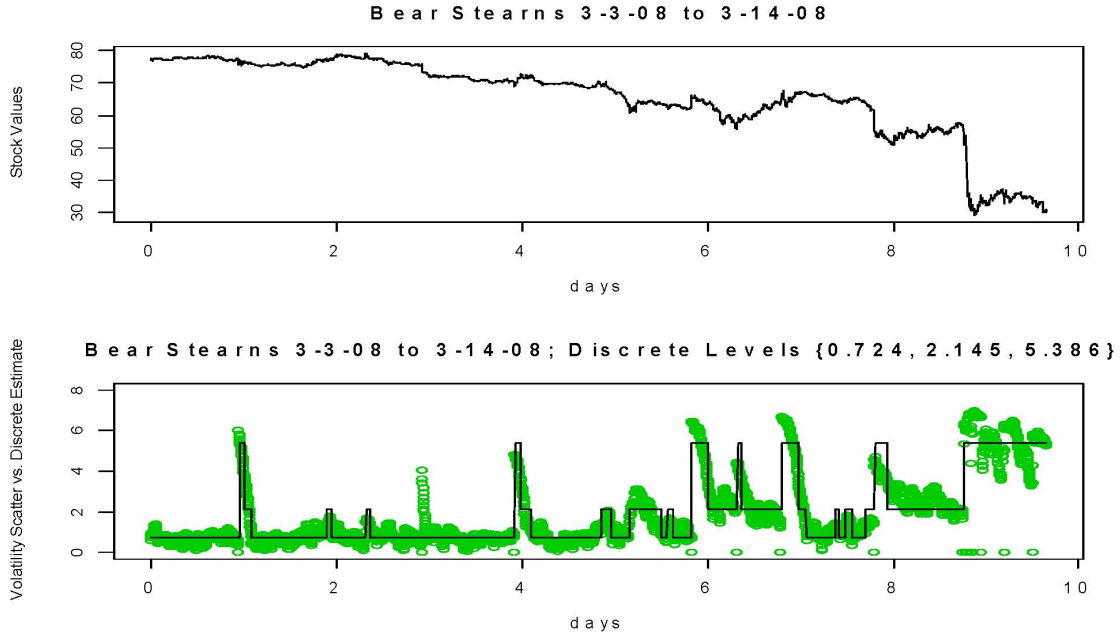


Figure 5: Bear Stearns Stock Price/Volatility during the last two weeks

We have performed a thorough investigation using data for several other equity from the investment sector (JPMorgan, Lehman Bros, Bank of America, Merrill Lynch) as well as non-investment sector (IBM, Chevron, Exxon, FedEx) during the week of the collapse March 10-March 14. For lack of space we present the plots and the detailed results for these stocks in the supplementary material Supplement B.

The results which were evident from this analysis are commented next.

The volatility levels estimated for each stock in the investment sector during the week are extremely high (though not at the levels encountered for Bear Stearns). Beyond the actual values, all of them exhibit a common behavior during the week. All stocks went to high levels at the beginning of the trading day on Thursday until about the middle of the day and then again on Friday. The corresponding stock values do not exhibit any obvious pattern from which this volatility behavior may be easily obtained.

When looking at the non-investment sector, we finally observe volatility levels which are characteristic of the respective stock. For example the estimated levels for IBM were 0.36, 0.63 and 0.78 (the typical implied volatility level for IBM stock calculated from option data using the Black Scholes model is historically around 0.4). Furthermore, IBM does exhibit increased levels of volatility as early as Thursday morning and further increasing during Friday. The oil companies stock analyzed (Chevron and Exxon Mobil) are very similar in both volatility values and behavior. They only show an increase in volatility starting Friday. This discrepancy in the volatility behavior between technology stock and others was explained to us by a trader who stated that in general technology stocks react much faster to the market movement than the commodities stock. Among the non-investment sector we also analyze FedEx volatility behavior during the week. This equity behavior exhibit slight to no reaction to the Bear Stearns collapse.

In conclusion, we believe that the model provided insight into market behavior during that week. First, we believe that the volatility levels for Bear Stearns equity

were abnormal starting with Friday March 7, 2008 and continued throughout the week of March 10-March 14 until the collapse of the investment firm. Signs were present in other investment equity as well but none as in the Bear Stearns case. We see higher volatility levels on Thursday morning (before SEC meets with the Bear Stearns officials) exhibited by the entire investment sector as well as the highly traded technology stock. Commodity sector was affected on Friday following the remarks by the Federal reserves bank of NY that a bankruptcy is imminent. The event did not seem to affect the entire market only the highly traded equity.

3.2 Part II: Physical data application. Temperature data

Global warming seems to be a "hot" topic these days so we wanted to see if we may adapt the model to study climate related problems. Temperature record is the most important signal for global warming. The problem we faced was finding temperature data sampled with high enough frequency (which is needed to detect shifts in variability). The highest frequency sampled data we found was hourly temperature data gathered in Central Park, New York City, New York, USA starting in 2000. *We realize that this data is obtained for a very large park located in a middle of arguably the largest metropolis on Earth and therefore our results are not to be extrapolated.* However, the results we found when estimating variability of this data were most interesting. We present two typical pictures of estimated volatility in Figure 6 on page 19 (the rest of images are in the supplementary material Supplement C).

It is remarkable that all the plots show evidence of a predominant variability value and the deviations from this value do not last very long. Accordingly, we conclude

that in general the hourly data show a large degree of consistency in variability of temperature. Consequently, Table 1 presents the value of the predominant variability each year. The last data available was from July 2010 and in order to have a consistent way of estimating we overlap some time period in the end to provide an entire year worth of data.

Table 1: Yearly predominant node volatility 2000-10 Central Park NY

Year	Predominant Volatility level
2000	0.27
2001	0.269
2002	0.271
2003	0.276
2004	0.276
2005	0.293
2006	0.43
2007	0.423
2008	0.425
08/08-07/09	0.424
08/09-07/10	0.411

When looking at these values one clear distinction stands out. For the years 2000 to 2005 the variability is remarkably close to 0.3. Beginning with 2006 there is a notable jump in variability (the rows are separated in the table for a better visibility) and the variability level stays remarkably close to 0.42. What exactly does this mean and what caused the shift in the temperature variability we do not know but we interpret this as some evidence of a certain change in the local climate in the NY City area.

We would have liked to compare the results obtained in this way with data from earlier years. Unfortunately, before 2000 the data was not gathered with high enough

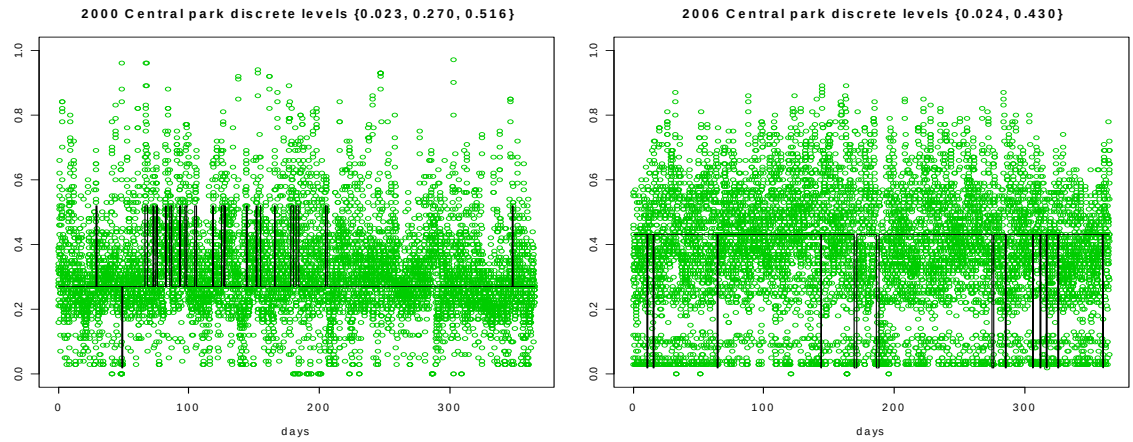


Figure 6: Hourly Temperature Data From 2000 and 2006

frequency. The only other temperature data we found was gathered at the same location (Central Park NYC) from 1976-1977 but this data was sampled every three hours and thus the numerical results obtained are not directly comparable. We may see the output in Figure 7 on page 20. This time the filter outputs only one volatility node as best fitting the data. The values for the two consecutive years are close too (0.49 vs. 0.48). However, as mentioned we may not compare this value with the results obtained for the 2000-2008 since the sampling frequency is different. Having said that, if the stationary distribution of the stochastic volatility process is Gaussian (as it seems to be) then a simple transformation should give an indication of the hourly volatility value.

$$\sigma_{3h} = \sigma_h * \sqrt{3}$$

With this we may guess the hourly values as: $0.48/\sqrt{3} = 0.277$ and $0.49/\sqrt{3} = 0.283$ which certainly look consistent with values we obtained from 2000 to 2005.

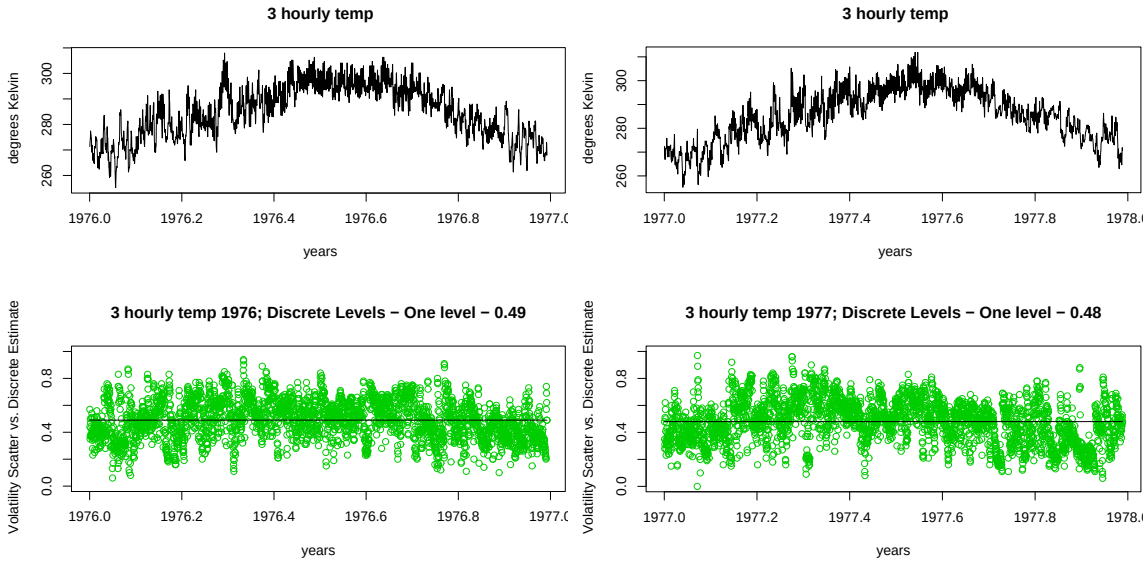


Figure 7: Three hour volatility estimates for 1976-1977

3.3 Part III: Analysis of seismometer readings during an Earthquake

For the final practical application of the model presented in this article we consider data collected for the Parkfield California earthquake of September 28, 2004: Magnitude 6.0 on the Richter scale. Studies in the 1980's had predicted an earthquake for 1993 in this area near the San Andreas fault. As a result a network of sensors was built and on December 20, 1994 a M=5.0 earthquake did occur. (CSMIP, 2004; Borcherdt et al., 2004, 2006, and references within). We owe the data analyzed herein to the buildup of these array sensors.

The seismographic readings were programmed to trigger for a P wave (primary wave) of magnitude $M > 3.0$. The P wave is a compression wave analogous to sound

waves. It is the fastest seismic of the seismic waves and consequently arrives first. The S waves (secondary waves) and the surface waves which are responsible for most of the destruction are slower. The Richter scale magnitude is a logarithmic (base 10) scale. This means that an $M = 5.0$ event has one hundred times the movement of a 3.0 event, a 6.0 event has one thousand times the movement of a 3.0 event. The signal used is the ground acceleration reading at 0.005 second intervals.

3.4 Analysis of the earthquake signal. Beginning

In the analysis we use data for the same earthquake (the 2004 Parkfield earthquake) obtained from two different research stations: Red Hills and Donna Lee stations. The two stations are in a mountainous region about at the same elevation and the distance between the stations is about 23.8 miles (or 38.3 km, distance calculated using their coordinates). Figure 8 presents the raw ground acceleration signal at the two stations. It is evident from the two pictures that Donna Lee was much closer to the epicenter. If we use the threshold arrival time (the time at which the raw acceleration passes magnitude 3) as the time at which the initial P wave energy arrived at each location then the earthquake starts at Donna Lee at 17:27.849 and at Red Hills at 17:29.259.

We perform the variability estimates for a restricted region at the beginning of the signal and we plot the estimates obtained in Figure 9. While the Donna Lee signal presents a spike in variability at exactly the same time as the P wave is observed (29.85 sec), for the Red Hills station there is a spike in variability 20 hundreds of a seconds earlier than the official earthquake starts. The second time in the image corresponds to the variability increase and this is remarkably close to the official

P wave time (29.26 sec). The difference may be due to the interference in signal traveling through the ground.

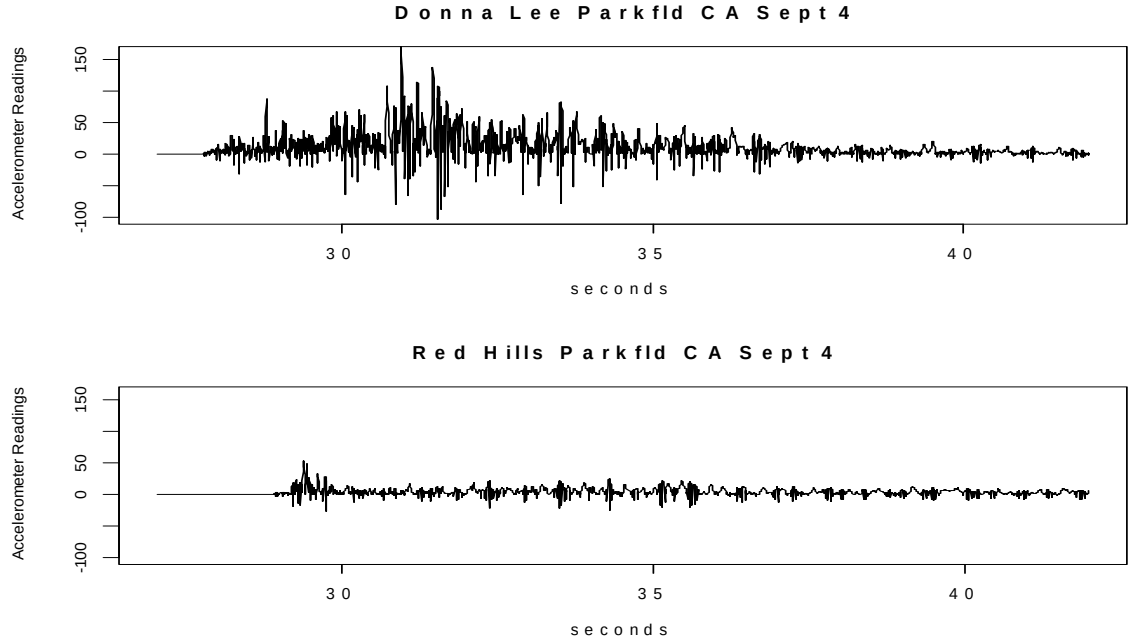


Figure 8: Raw acceleration signal at the two stations

One of the methods for locating the actual origin or epicenter of the earthquake is by using the arrival times of the P waves. A simple triangulation method may be used if one knows the speed of travel of the waves through the ground and such speeds are charted and known. However, there is a number of factors affecting this speed which are not charted. Ground temperature and humidity for instance are factors that change during the day and their influence is unknown. Furthermore, the more dispersed the wave (further from the epicenter) the greater the interference (and influence) of these factors. Note that the signal at Red Hills does not look anything like the signal at a location closer to the epicenter. For this reason we believe that

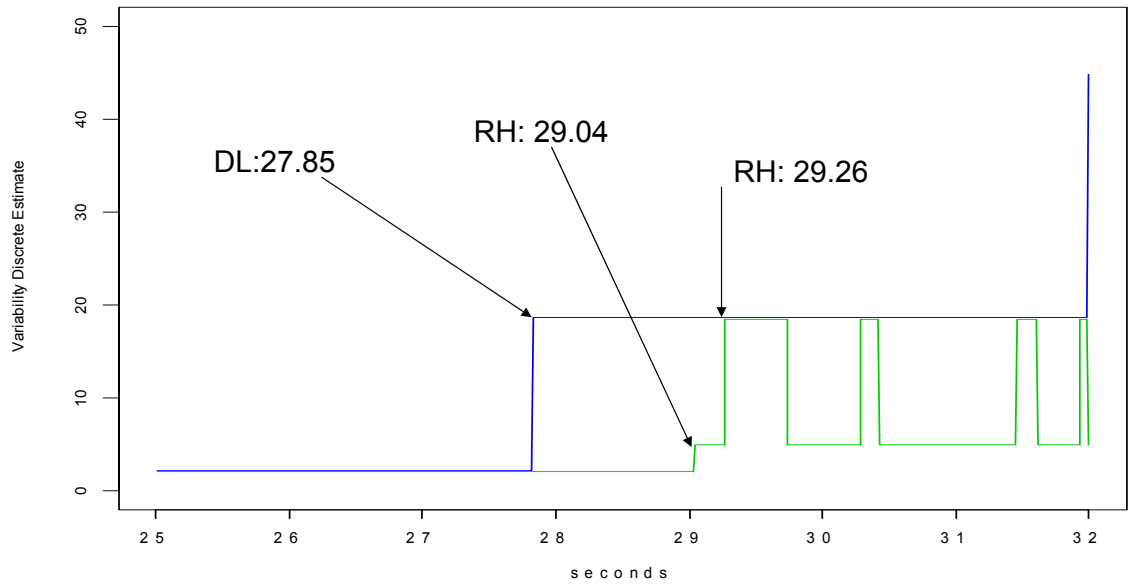


Figure 9: Variability comparison. Beginning of earthquake

the first time a shift in variance is detected may be an alternative way of providing triangulation results. The results obtained this way could be then compared with the more traditional methods either reinforcing or contradicting the results. In either case we believe that the methodology may be useful.

Calculating the focus or point of origin of the earthquake is very useful for the geophysical analysis and long term estimates of earthquake hazard. Even though short term prediction has not been successful some correlation between smaller earthquakes and detectable activity shortly before an earthquake has been observed. The timing and amplitude of seismographic signals is also used for fundamental research about the Earth's interior.

3.5 Analysis: During the earthquake

In Figure 10 we plot the the variability analysis obtained using the entire acceleration signal at the two stations. Figure 11 superimposes the two estimates on the same timeline for the ease of comparison.

We believe that the variability estimates may prove to be most useful for this period. First note that the shifts in variability are not evident from the accelerogram. People living through earthquakes will tell of the pulses that they feel during the earthquake. This shifts in variability may indicate exactly that. Furthermore, by combining the magnitude and the frequency of these intervals we may be able to measure the destructive power of an earthquake. This is done only approximately with the Richter scale. The Richter scale (adjusted logarithm base 10 of the highest perceived acceleration value at the epicenter) even though supposedly measuring the earthquake strength has in fact nothing to do with the actual destructive power. For example, the 2010 Haiti Earthquake an extremely devastating earthquake (230,000 mortal cases, 3 million people affected) was recorded as 7.0 on the Richter scale. The only other recorded earthquake at 7.0 is the West Java earthquake of 2009 - a very serious earthquake nonetheless one that cannot be compared with the severity of the Haiti earthquake (79 deaths displacing 210,000 people). We should also mention that Java is a much more densely populated region. The 2007 Tocopilla earthquake in Northern Chile recorded at 7.7 on the Richter scale caused two deaths and 150 minor injuries, 15,000 people were displaced. The economic damage of an earthquake is in fact measured by the Mercalli intensity scale an entirely subjective scale which rates the earthquake from I - Instrumental (not felt) to XII - Cataclysmic. But in fact

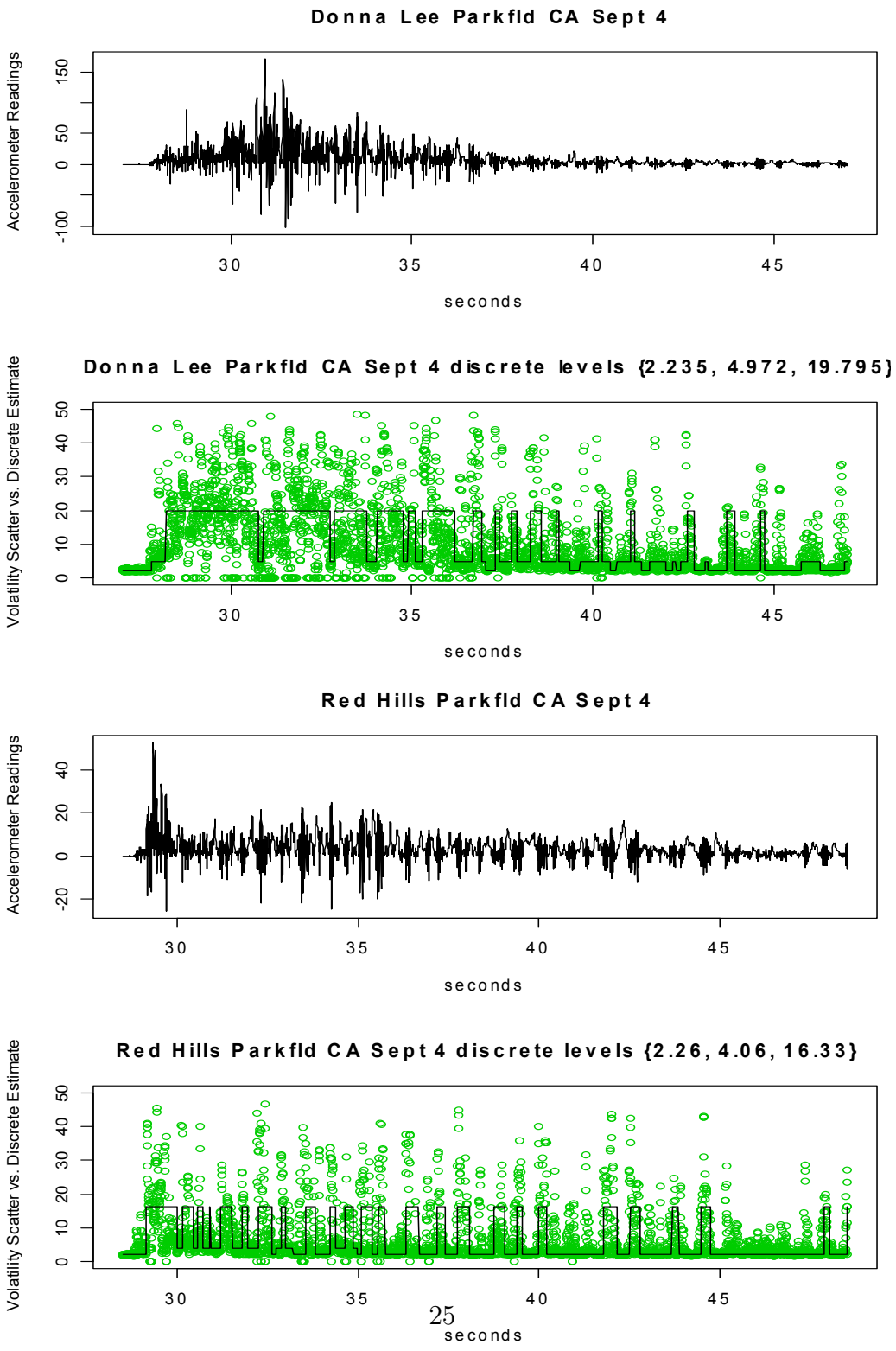


Figure 10: Parkfield CA, Donna Lee Peak Seismograph Readings

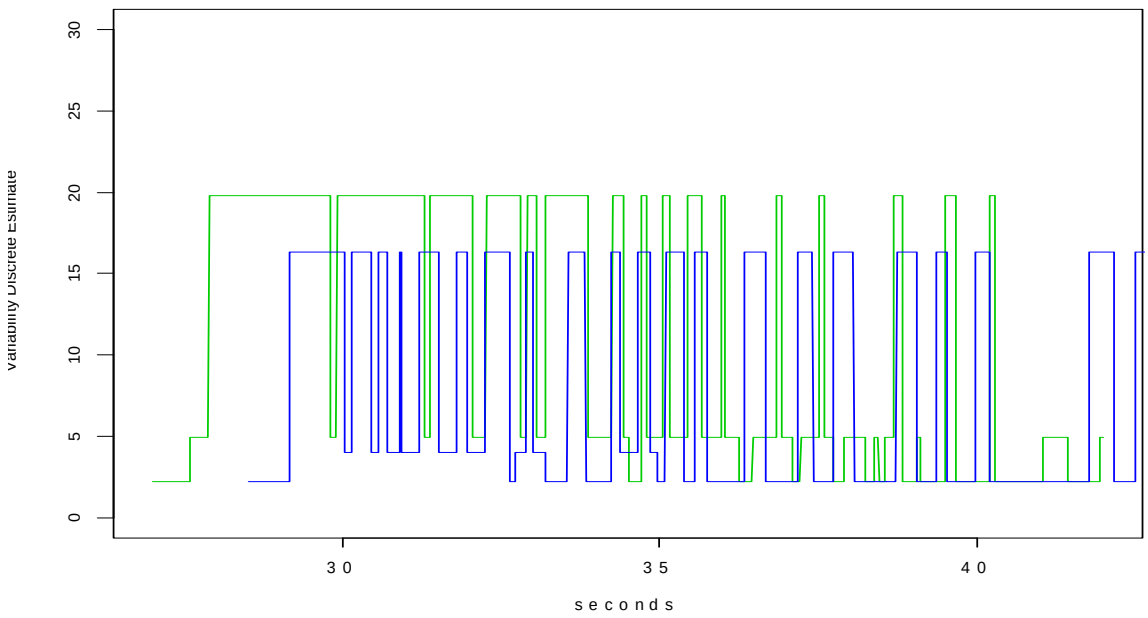


Figure 11: Parkfield CA earthquake. Comparison of the variability signals, Donna Lee variability in green and Red Hills variability in blue.

the earthquake is rated thus months after it takes place and by looking at the total devastation it produces. On the contrary looking at the estimated variability we may be able to provide a much more accurate measure for this quantity immediately after the earthquake takes place.

Furthermore, if we look at the variability estimates (Figure 11) and the ground signal itself (Fig. 8), it is pretty clear that this is the same earthquake even though one could not tell this from the signal itself. The variability patterns are parallel at the two locations. Given the relatively short distance between the two stations one would expect the signal at Red Hills station to have a similar destructive power. This may not be the case based on this preliminary study.

3.6 Analysis: End of the earthquake signal, aftershocks

Figure 10 presents the variability estimates at the two stations. Note that after about 10 seconds at both locations the predominant variability changes to a lower regime. The accelerometer signal show lower values for the ground acceleration, but from that signal alone there does not seem to be a clear end point of the earthquake.

By looking at the variability on this last region perhaps one could determine the probability of a future aftershock. An aftershock is a smaller earthquake that occurs after a previous large earthquake in the same area (the main shock). If an aftershock is larger than the main shock, the aftershock is redesignated as the main shock and the original main shock is redesignated as a foreshock. Aftershocks are mainly produced as the crust around the displaced fault plane adjusts to the effects of the main shock, and information about this perhaps may be gathered by looking at the variability

pattern at the end of the earthquake signal.

4 Conclusion

The work presents a methodology of estimating the unobservable variability of a signal. The method was used for three different areas and for each area provided insight into the working of the real life signal. We believe this is the main purpose of a model and that we have accomplished this purpose.

References

Aït-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics 83*, 413–452.

Black, F. and M. Scholes (1973). The valuation of options and corporate liability. *Journal of Political Economy 81*, 637–654.

Bladt, M. and M. Sørensen (2007). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Preprint.

Borcherdt, R. D., M. J. S. Johnston, G. Glassmoyer, and C. Dietel (2004). Recordings of the 2004 parkfield earthquake on the general earthquake observation sysem array. Technical report, U.S. Geological Survey. Open-file Report 2004-1376 October 11, 2004.

Borcherdt, R. D., M. J. S. Johnston, G. Glassmoyer, and C. Dietel (2006, June). Recordings of the 2004 parkfield earthquake on the general earthquake observation sysem array. *Bulletin of the Seismological Society of America 96*(4B), S73 – S89.

Cho, S., R. Haralick, and S. Yi (1989). Improvement of Kittler and Illingworth's Minimum Error Thresholding. *Pattern Recognition 22*(5), 609–617.

Cohan, W. D. (2009). *House of Cards*. Thorndike Press.

Cont, R. and P. Tankov (2003). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC.

CSMIP (2004). Strong-motion data from the parkfield earthquake of 20 december 1994. Technical report, California Strong Motion Instrumentiation Program, Division of Mines, Department of Conservation. Open-file Report 2004-1376 October 11, 2004.

Cvitanić, J., R. Liptser, and B. Rozovskii (2006). A filtering approach to tracking volatility from prices observed at random times. *Annals of Applied Probability 16*(3), 1633–1652.

Cvitanić, J., B. Rozovskii, and I. Zaliapin (2006). Numerical estimation of volatility values from discretely observed diffusion data. *The Journal of Computational Finance 9*(4), 1–36.

Dash, M. (2008, September). Bank agrees to buy troubled loan giant for 4 billion. *NYT*. 9-25-08 archive resource www.nytimes.com.

DelMoral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.

Del Moral, P., J. Jacod, and P. Protter (2001). The monte-carlo method for filtering with discrete time observations. *Probability Theory and Related Fields 120*, 346–368.

Geithner, T. (2008). The current financial challenges: Policy and regulatory implications. New York. Council on Foreign Relations Corporate Conference 2008. March 6: Remarks by Timothy Geithner, President and CEO of the Federal Reserve Bank of New York.

Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli 6*(6), 1051–1079.

Landon and Sorkin (2008, March). J.p.morgan acts to buy ailing bear stearns at huge discount. *New York Times*. 3-15-08 archive resource www.nytimes.com.

Levin, F. (2010). *Monte Carlo estimation of stochastic volatility for stock values and potential applications to temperature and seismographic data*. Ph. D. thesis, Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey, USA.

Mariani, M., I. Florescu, M. B. Varela, and E. Ncheuguim (2009, April). Long correlations and levy models applied to the study of memory effects in high frequency (tick) data. *Physica A 388*(8), 1659–1664.

Merton, R. C. (1992). *Continuous-Time Finance (Macroeconomics and Finance)*. Wiley-Blackwell.

Nielsen, J. N. and M. Vestergaard (2000). Estimation in continuous-time stochastic volatility models using nonlinear filters. *International Journal of Theoretical and Applied Finance 3*(2), 279–308.

Protter, P. E. (2005). *Stochastic Integration and Diferential Equations* (second ed.). Springer.

Ross Sorkin, A. (2008, March). Jpmorgan pays 2 dollars a share for bear stearns. *New York Times*. 3-17-08 archive resource www.nytimes.com.

Shephard, N. (Ed.) (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

Sørensen, H. (2003). Simulated likelihood approximations for stochastic volatility models. *Scandinavian Journal of Statistics 30*(2), 257–276.

Tsay, R. (2005). *Analysis of Financial Time Series (2nd Ed.)*. John Wiley & Sons, Inc.

U.S. Senate Committee on Banking, Housing and Urban Affairs (2008). *Actions by the New York Fed in Response to Liquidity Pressures in Financial Markets*. U.S. Senate Committee on Banking, Housing and Urban Affairs. April 3: Testimony given by Timothy Geithner, President and CEO of the Federal Reserve Bank of New York.

A Theoretical results and empirical testing

In this section we provide convergence results of the estimation algorithm presented in the paper. The arguments presented are both: theoretical – by providing estimates for convergence rates, and empirical – by generating paths from a known process and comparing the estimates with the ground truth.

A.1 How does the particle filter work?

To start this section we note that the discrete process in (3) does not correspond to a discretization of the original continuous process in (2). Indeed, locally between any two times t_{i-1}, t_i the evolved process approximates the following continuous volatility process:

$$\begin{aligned} X_t &= X_{t_{i-1}} + \int_{t_{i-1}}^{t} \left(r - \frac{Y_u^2}{2} \right) du + \int_{t_{i-1}}^{t} Y_u dW_u \\ Y_t &= Y_{t_{i-1}} + \beta(Z_t - Z_{t_{i-1}}) \end{aligned} \tag{4}$$

The approximation in (3) of this auxiliary model is used to estimate the parameters of the Markov chain Y_t.

The convergence of the discretized filter Y_t in (3) to the continuous version from which it comes (equation (4)) is a well studied problem. Del Moral, Jacod, and Protter Del Moral et al. (2001); DelMoral (2004) develop bounds on the expected error of the filter and the filter expected value. In the cited work, the authors work with a given observed process $\{X_t \mid t \leq t_N\}$ a "noisy" measurement of $\{Y_t \mid t \leq t_N\}$

the un-observable process or in our case the auxiliary volatility process.

Specifically, the following general situation is presented where X_t, Y_t, a, a', b, and b' are q dimensional random vectors and valued functions.

$$\begin{cases} dX_t &= a(X_t, Y_t)dt + b(X_t, Y_t)dW_t \\ dY_t &= a'(Y_t)dt + b'(Y_t)dZ_t \end{cases}, \tag{5}$$

subject to the condition that the first and second derivatives of b and b' exist and are bounded. Furthermore, $\|b\|^2$ and $\|b'\|^2$ are not identically zero.

The distribution of the process being estimated $\sigma(Y_t)$ is denoted with $\Pi_N\sigma = \{\sigma_t \mid X_t, t \leq t_N)$. The output of the numerical filter at the N^{th} observation using n test paths is the estimated distribution of σ process, denoted with $\hat{\Pi}_N^n\sigma = \{\hat{\sigma}_t \mid X_i, i \leq N\}$. The work cited provides bounds for the expected error $\mathbf{E}(\hat{\Pi}_N^n\sigma - \pi_N\sigma)$ with a function of N, n and σ.

The weighting function used in the general filtering algorithm ϕ may be any arbitrary function integrating to 1 and the weights are assigned using:

$$\phi_n(v) = C(N)\phi(vn^{\frac{1}{2+q}}) \tag{6}$$

which is identical with what we use in our more specific algorithm ($q = 1$), while n is the number of generated test paths and N is the total number of observations.

The cited results prove two convergence results of the filter to the auxiliary model

as $n \to \infty$. Specifically, when $q = 1$:

$$\mathbf{E}\left[\hat{\Pi}_N^n \sigma - \pi_N \sigma\right] < \frac{C\sigma}{n^{\frac{1}{3}}} \tag{7}$$

$$\mathbf{P}\left(|\hat{\Pi}_N^n \sigma - \pi_N \sigma)| \geq \delta\right) \leq Ce^{-n^{\frac{2}{3}}\frac{\delta}{(C\sigma)^2}} \tag{8}$$

as $n \to \infty$ and $C = C(N)$ where n is the number of trial paths and N is the number of observation points.

In our case the convergence proof needs to be modified due to the presence of the auxiliary process. The idea in our case is to show that the auxiliary process converges to the discretization of the true process generated from the Markov chain volatility. The only theoretical case considered is when the weighting function is a Gaussian (we take advantage of the normal-normal conjugate distributions). In the empirical testing section (A.4) we look at the effects of many other weighting functions. They have little effect on the estimation.

We only state the relevant results in the appendix. For a more thorough analysis and the proof of the results we refer to the original thesis (Levin, 2010).

A.2 Theoretical results about convergence and parameter estimates

The objective of this section is to derive some simple and applicable convergence results for the hypothesized system (equation 2) and the filter (equation 3) used to analyze it.

A.2.1 First Pass: Initial Volatility Estimates

For a process X with volatility a Markov chain Y as used in this paper we use the following notations. For each volatility state i at time t when the estimator is obtained:

- σ_i is the "true" underlying volatility state
- N is the number of consecutive observations during one uninterrupted stay in the respective state, up until time t
- The total number of observations at the respective state i (throughout all the observed data available at time t) is denoted with Pts
- The number of test paths the filter generates is denoted with n
- $\tilde{\sigma}_i$ is volatility of the discrete state system.
- σ_{est} is the filter estimated volatility (which uses a continuous state space approximation)
- σ_{bound} is the theoretical point at which extreme test values for $\tilde{\sigma}_2$ are ignored or truncated. Its value is adapted to the filter performance and decreases as the filter converges.
- σ_ϕ is the standard deviation of the penalty function and is usually much smaller than σ_{bound} and $\tilde{\sigma}_i$

– σ_B is a derived factor in the theoretical description of the filter defined by:

$$\sigma_B^2 = \sigma_\phi^2 + \tilde{\sigma}_i^2 + \sigma_{bound}^2 \tag{9}$$

With these notations the following result relates the mean of the estimator obtained using the auxiliary process with the mean of the estimator we would have obtained using the real process at a certain point in time.

Theorem A.1. *For point volatility estimates after N points at a given state given* $\epsilon > 0$ *there exists C large enough so that:*

$$P\left(|\tilde{\sigma}_i - \sigma_{est}| > C\sqrt{\frac{\sigma_B}{Nn\Delta t}}\right) < \epsilon \tag{10}$$

The next theorem relates the nodes estimated using all the data history.

Theorem A.2. *For volatility node estimates after Pts total number observations at a given level and N observations per mean duration:*

$$P\left(|\tilde{\sigma}_i - \sigma_{est}| > C \times \sqrt{\frac{\sigma_B}{n \times Pts \times \Delta t} \quad \frac{\ln(N)}{N}}\right) < \epsilon \tag{11}$$

The next result relates the realized volatility of the discretized system to the parameter value.

Theorem A.3.

$$P\left[|\tilde{\sigma}_i - \sigma_i| > C \times \sqrt{\frac{2}{N}}\sigma_i\right] < \epsilon \tag{12}$$

For a two state system $\{a_1, a_2\}$ with transitions rates λ_1 from state a_1 to state a_2 and λ_2 the transition rate from state a_2 to state a_1 we want to estimate the node levels $\{a_1, a_2\}$ and the transition rates at each level. The node levels are determined from fitting a mixture of Gaussian distributions to the volatility estimates and using the means as node estimates. With enough points the convergence specified above allows us to distinguish and estimate nodes if the mean duration at the levels is long enough with enough transitions. From the above relation for point estimates (ref to equation 10) we see an increase in accuracy as the number of paths and the number of observations per duration at a given level increase. This holds even more strongly for the node estimates (refer to equation 11) and also increases as the total number of points at a given level increases

A.3 Markov Chain Parameter Estimates

The conditional transition probability from state i to state j is given by:

$$p_{i,j} = \frac{\lambda_{i,j}}{\sum_{j=1}^{n} \lambda_{i,j}} \tag{13}$$

This formula allows estimates for transition probabilities to be computed using the empirical transition rates. Confidence intervals for these transition rates for individual cases may be found using the binomial approximation (for two states) or a multinomial distribution or Pearson χ^2 test with more states.

In practice the larger the window or frame of points used the greater the percentage of correct estimates. One problem we encountered is the way in which isolated errors

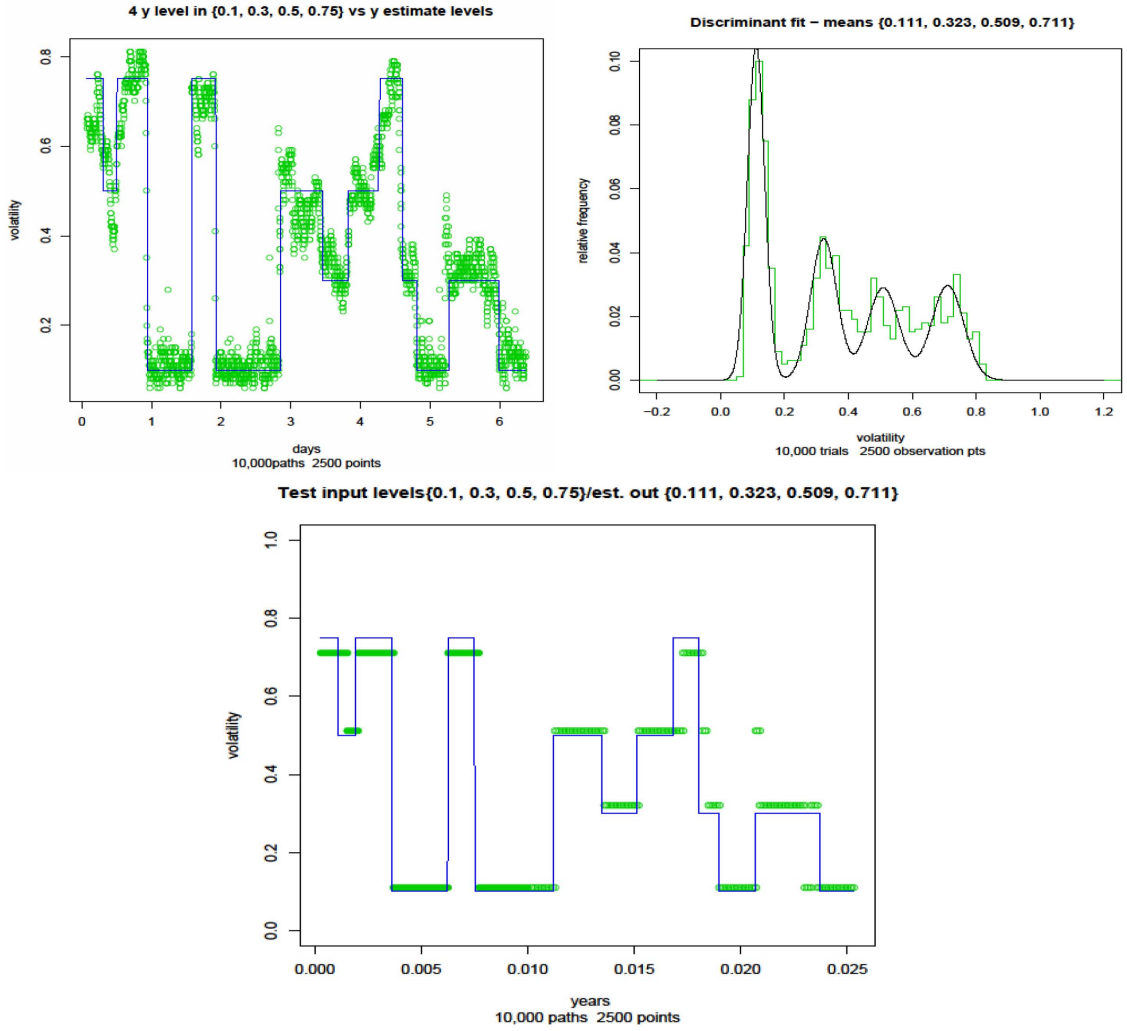
affect the estimates for mean duration at a certain state. If the algorithm erroneously detects a jump in volatility in the middle of a stay at a certain state this effectively halves the duration time at the state. With large enough data the problem goes away. However, this usually has rather absolute practical limitations. The quantitative relationship of probability of isolated errors to accuracy of mean durations estimates seems to be very complicated.

A.4 Empirical Testing

In this section we generate observations using a known Markov chain (ground truth). Then we perform the estimation procedure and we compare the estimates with the known values.

We performed simulation results for 2, 3 and 4 node Markov chains. We simulated $2,000$ observation points (which translates into about a week of minute financial data), 8 to 10 thousand simulation points per path and about 50 to 200 observations per state duration. In Fig. 12 we present the algorithm output for a 4 node estimation. The first image plots the real volatility (black line) and the estimated means (green points). The second image presents the histogram of the means (green) and the fitted gaussian mixture distribution on top. Finally the third image displays the real volatility (black line) and the estimated points restricted to the node values. In general the algorithm needs several observations at a new level before it detects a volatility shift. For the results obtaining extra test data, we refer the reader to the supplementary document Supplement A.

Figure 12: Test Data: 4 Nodes (0.1, 0.3, 0.5, 0.75)



A.5 A list of supplementary documents

Due to the page constraint of this document we were forced to leave out many results and images that we felt would have enhanced the work and demonstrated the extensive testing put into this work. Thus we have added three documents which are visible to the links provided as well as on the journal site.

Supplement A: Selected test results using 2, 3 and 4 nodes This pdf file presents test results using data generated from a process with known parameter values. The values we chose resemble the numerical values, frequencies and distances between nodes as observed in our applications. Some values were chosen due to their difficulty in estimation (nodes close to each other) to illustrate the power of the algorithm.

http://www.math.stevens.edu/∼ifloresc/SupplementA.pdf

Supplement B: Other stock volatility analysis This pdf file presents volatility runs obtained from minute stock data during the week of the Bear Stearns collapse. We use financial stocks as well as commodity stock to illustrate the differences in the volatility patterns.

http://www.math.stevens.edu/∼ifloresc/SupplementBStocks.pdf

Supplement C: Temperature variability results This pdf file presents volatility runs obtained from the climate data in Central Park New York from 2000 to 2010.

http://www.math.stevens.edu/∼ifloresc/SupplementCClimate.pdf