An estimation technique for Time Indexed Gaussian Mixture Models

Moeti M. Ncube* and Anuj Srivastava*

Abstract

Assume that there is data $x_1, x_2, .., x_n$ coming from densities $N(\mu_1, \sigma_1)$ with probability α and $N(\mu_2, \sigma_2)$ with probability $1 - \alpha$ and our goal is to estimate the parameters $\theta = [\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha]$. The estimation of this parametric mixture model is a well known problem in statistics and can be solved using the well known Expecation-Maximization procedure for mixture models.

We now further generalize this problem to the case in which the observations $x_1, x_2, .., x_n$ are now coming from time-indexed densities $N(\mu(t)_1, \sigma(t)_1)$ with probability α and $N(\mu(t)_2, \sigma(t)_2)$ with probability $1-\alpha$. Again, we want to estimate the parameters $\theta = [\mu(t)_1, \sigma(t)_1, \mu(t)_2, \sigma(t)_2, \alpha]$. In this paper, we develop a novel Time Indexed Expectation Maximization (TSEM) procedure that can be used to estimate a general class of time indexed mixture models. In particular, we highlight models that incorporate unknown mixture densities in state equations and nonlinear observations. We use this procedure to solve the time-indexed Gaussian mixture problem and successfully apply this algorithm to the estimation of electricity price spikes.

1 Introduction

Electricity price movements is one example of a highly non-linear time series that typically features both jumps and spikes. Since the widespread deregulations of the electricity markets in the early 90s, the market activity has increased exponentially. This trend has made the evaluations of supply contracts a topic of serious concern. Electricity markets are known to have much more erratic behavior than typical commodities. The inability to store the commodity for a period of short supply leads to an inelastic demand as well as a steep supply function. Because of this, electricity

*Department of Statistics, Florida State University, Tallahassee, FL 32309

prices tend to have several jumps or spikes in a time series that occur at seemingly random times, and often revert back to some steady level quickly. Unhedged exposure of such dramatic price movements can have drastic consequences. In the summer of 1998, wholesale power prices in the Midwest of the US spike to \$7000 per MWh, while the normal price range is typically between \$30 to \$60 MWh; such price exposure can wipe out a portfolio not properly hedged. A good model for power would attempt to capture these dynamics by assuming a time series model that is a mixture between states, two or more.

In this paper, we will develop a simple time series model that can be used to model several different types of times series; in particular for this paper, we use it to model electricity prices with price spikes. We will develop a novel Time Indexed Expectation Maximization (TSEM) procedure that can be used to estimate the parameters of this model. We successfully apply this algorithm to the estimation of electricity price spikes.

2 Nonlinear Models With Mixture Densities

Several models such as [2], [5], [6], [3] and [4], have been developed to describe the evolution of the asset over time. Typically, models that describe commodity price movements have several underlying elements in common. First, a model specification will attempt to identify one or more sources of randomness believed to influence the commodity price. One of the most common features observed in commodities is mean-reversion. When the spot price is low, the supply of the product tends to decrease, driving the price upward. The same phenomenon is observed when the price is very high. Therefore, at least one mean reverting factor is identified in the model description. Other stochastic factors may also be modeled as random walks.

These models are often developed under the assumptions that the commodity is tradable and storable. In fact, this is a feature of many assets such as crude oil, gold, and natural gas. However, some commodities like electricity do not abide by these characteristics, as it is a non-storable commodity and cannot be treated as a security. Since the widespread deregulations of the electricity market's in the early 90s, the market activity has increased exponentially. This trend has made the evaluations of supply contracts a topic of serious concern. Electricity markets are known to have much more erratic behavior than typical commodities. The inability to store the commodity for a period of short supply leads to an inelastic demand as well as a steep supply function. Because of this, electricity prices tend to have several jumps or spikes in a time series that occur at seemingly

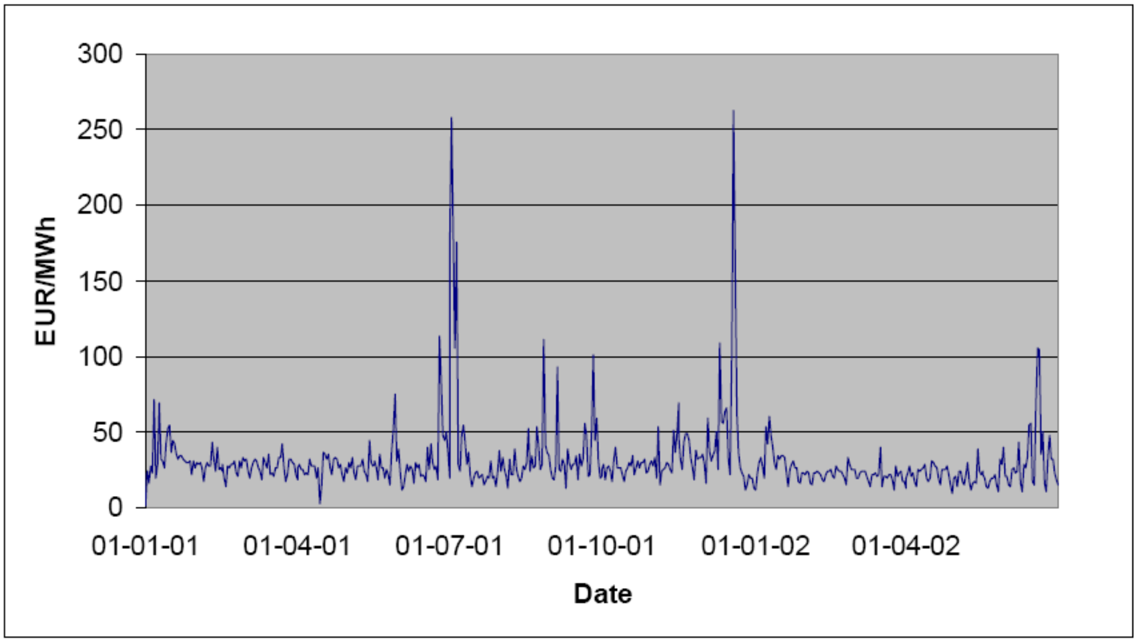

Figure 1: The Baseload APX prices in the period: January 2, 2001-June 30, 2002

random times, and often revert back to some steady level quickly. Unhedged exposure of such dramatic price movements can have drastic consequences. In the summer of 1998, wholesale power prices in the Midwest of the US spike to \$7000 per MWh, while the normal price range is typically between \$30 to \$60 MWh. Due to government regulation, price ceilings are often imposed on the suppliers of electricity so that they cannot pass extreme rises in prices on to customers. This phenomenon leads to the default of two power marketers in the east coast [1].

It is of vital importance that suppliers are able to hedge their exposure to risk through derivatives such as forward, futures, and/or swaps. These products require an understanding of the spot movements, which necessarily leads to more sophisticated models of commodities that can describe the special characteristics of this market. Specifically, these models can account for mean revertion, volatility changes, spikes, or jumps. In our model, we assume that, at any given time, there is a fixed supply S_t and demand D_t for a commodity, at each time t. The state variable X jointly represents the interactions of these stochastic factors by modeling the log ratio of the terms as follows: $x_t = log(D_t/S_t)$. We will call Equation (1) the supply-demand state process, where x_t is assumed to follow an AR(1) process that allows for sudden and drastic changes in the supply-demand ratios process that can occur with a probability λ, at each time t. The size of this jump is normally distributed with the mean and variance given in (3). We consider the following model specifications that can incorporate all of these features:

$$x_t = ax_{t-1} + n_t + J_t(g, hb^2, \lambda) \tag{1}$$

$$y_t = ce^{x_t} + v_t, \quad \text{where, } \mathrm{var}(n_t) = b^2,\ \mathrm{var}(v_t) = d^2 \text{ and,} \tag{2}$$

$$J_t(g, hb^2, \lambda) = \begin{cases} N(g, hb^2) & \text{with prob.}\lambda \\ 0 & \text{with prob.}1-\lambda \end{cases}. \tag{3}$$

The observations model, given in Equation (1), has a simple interpretation. When the supply of a commodity is exactly equal to the demand, the observed futures price y_t will be equal to the quantity c, representing the equilibrium level of the commodity, plus some random noise. If the supply is far greater than the demand, then the exponent of the exponential becomes negative, causing y_t to drop below the equilibrium level. Conversely, when the demand far out paces the supply, y_t will rise above the equilibrium level c.

We have defined several tuning parameters in this setup that specify this system: $\theta = [a, b, c, d, g, h, \lambda]$. The state Equation (1) is essentially a mixture indexed by the time t. We can specify this density as $f(x_t|x_{t-1}, l_t, \theta) = f_{l_t}(x_t|\mu_{l_t,t}, \sigma^2_{l_t,t})$, where $\mu_{l_t,t} = \begin{cases} ax_{t-1} + g & l_t = 1 \\ ax_{t-1} & l_t = 2 \end{cases}$ and $\sigma^2_{l_t,t} = \begin{cases} (1+h)b^2 & l_t = 1 \\ b^2 & l_t = 2 \end{cases}$.

2.1 Estimation of Time-Indexed Gaussian Mixture With Known State

We first demonstrate the flexibility of this model by generating several different time series processes with different values of the tuning parameter θ. Since our interest here is mainly in various stochastic descriptions that can be generated from the state variable, we set Y to be the identity function of X. We describe a few prominent processes that can be obtained from the model:

- Random Walk: A simple random walk process can be generated from this model by setting the values $a = 1$, $b > 0$ and $\lambda = 0$. An example of this process is illustrated in the top-left panel of Figure (2), generated using the values: $\theta = [1, .2, 0, 0, 0, 0, 0]$.

- AR(1) Process: A simple AR(1) process is obtained by setting the values $0 < a < 1$, $b > 0$ and $\lambda = 0$. An example is illustrated in the top-right panel of Figure (2), generated using the values: $\theta = [.5, .2, 0, 0, 0, 0, 0]$.

- Mean Reversion Process: A mean reverting process around a value v can be generated from this model by setting the values $0 < a < 1$, $b > 0$, $\lambda = 1$, $g = (1-a) * v$ and $h > 0$. An example is illustrated in the middle-left panel. In this case, we used $v = 4$ and $\theta = [.6, .2, 0, 0, (1 - .6) * 4, 1, 1]$.

- Stochastic Volatility: A discrete stochastic volatility process with two random states of volatility can be generated from this model by setting the values $0 < a < 1$, $b > 0$, $g = 0$, $\lambda = \lambda_0$ and $h > 0$. An example is shown in the middle-right panel. We generated this process using $\theta = [.6, .2, 0, 0, 0, 100, .5]$.

- Spikes: A stochastic process with random spikes occurrences can be generated from this model by setting the values $a \approx 0$,$b > 0$, $g > 0$, $0 < \lambda < 1$ and $h > 0$. The bottom-left panel shows an example generated for $\theta = [.01, .2, 0, 0, 5, 20, .1]$.

- AR(1) Jumps: A stochastic process with random jumps occurrences can be generated from this model by setting the values $a \approx 1$, $b > 0$, $g > 0$, $0 < \lambda < 1$ and $h > 0$. Finally, an example of this process is shown in the bottom-right panel, with the parameter $\theta = [.99, .2, 0, 0, 0, 200, .1]$.

In this model, we are given both the state and the observations X and Y and the goal is to derive an estimation procedure for θ. Since the state variable is a Gaussian mixture density, we introduced the label variable$L = [l_1, l_2, ...l_t]$, which identifies the density of x_t at time t as missing data. This problem is very similar to the well known Gaussian mixture problem, which can be solved using the E.M. algorithm [7]. The added complexity now is that the mean and variance of both normal densities is indexed by time. However, we can still solve this time-indexed Gaussian mixture using the E.M. algorithm.

Expectation Step: Here we compute the expectation of the log-likelihood function with respect to distribution on $\{l_t\}$ conditioned on the current estimate θ_m and the given data X, Y: $Q(\theta|X, Y, \theta_m) = \sum_{t=1}^{T} E_{l_t}[logf(x_t, y_t, l_t)|X, Y, \theta_m]$. Expanding $E_{l_t}[logf(x_t, y_t, l_t)|Y, \theta_m]$ results in:

$$\begin{aligned} &= logf(y_t|x_t, \theta) + \sum_{l_t=1}^{2} logf(x_t|x_{t-1}, l_t, \theta)P(l_t|x_t, x_{t-1}, \theta_m) + \sum_{l_t=1}^{2} logf(l_t|\theta)P(l_t|x_t, x_{t-1}, \theta_m) \\ &= log(N(ce^{x_t}, d^2)) + log(N(ax_{t-1} + g, (1+h)b^2))p_{1,t}^m + log(N(ax_{t-1}, b^2))p_{2,t}^m \\ &\quad + \log(\lambda)p_{1,t}^m + \log(1-\lambda)p_{2,t}^m \ . \end{aligned} \tag{4}$$

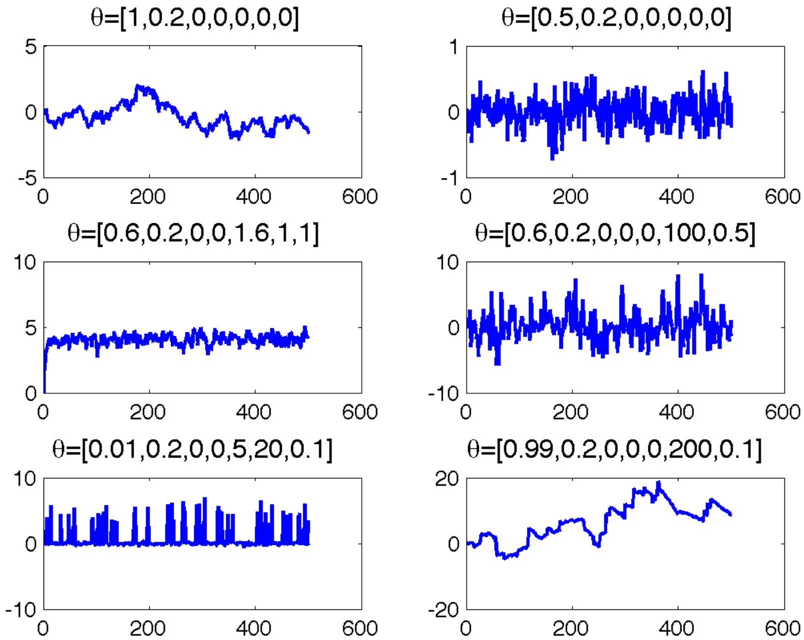

Figure 2: Sample paths from different processes. Top-left a random walk process, top right an AR(1) process, middle-left a Mean Reversion process, middle-right a discrete stochastic volatility process, bottom-left a stochastic process with spikes bottom-right a AR(1) process with jumps.

Here we have used $p^m_{l_t,t} = P(l_t|x_t, x_{t-1}, \theta_m)$ and $p^m_{2,t} = 1 - p^m_{1,t}$, and where:

$$
\begin{aligned}
p^m_{l_t,t} &= \frac{P(x_t, l_t|x_{t-1}, \theta_m)}{P(x_t|x_{t-1}, \theta_m} = \frac{P(x_t|l_t, x_{t-1}, \theta_m)P(l_t|\theta_m)}{\sum_{i=1}^2 P(x_t|i, x_{t-1}, \theta_m)P(i|\theta_m)} \\
&= \frac{N(\mu^m_{l_t,t}, (\sigma^2)^m_{l_t,t})\lambda_m}{N(\mu^m_{1,t}, (\sigma^2)^m_{1,t})\lambda_m + N(\mu^m_{2,t}, (\sigma^2)^m_{2,t})(1-\lambda_m)} \ .
\end{aligned} \tag{5}
$$

Maximization Step: We maximize $Q(\theta|Y, \theta_m)$ over each element in θ to obtain the following updates:

$$
\begin{aligned}
a_{m+1} &= \frac{\sum_{t=2}^T [\frac{(x_t - g_m)x_{t-1}}{b_m^2(h_m+1)} p^m_{1,t} + \frac{x_t x_{t-1}}{b_m^2} p^m_{2,t}]}{\sum_{t=2}^T [\frac{x_{t-1}^2}{b_m^2(h_m+1)} p^m_{1,t} + \frac{x_{t-1}^2}{b_m^2} p^m_{2,t}]}, \ b_{m+1} = (\frac{\sum_{t=2}^T [\frac{(x_t - \mu_{1,t})^2 p^m_{1,t}}{h_m+1} + (x_t - \mu_{2,t})^2 p^m_{2,t}]}{T-1})^{1/2} \\
c_{m+1} &= \frac{\sum_{t=1}^T y_t e^{x_t}}{\sum_{t=1}^T e^{2x_t}}, \ d_{m+1} = \frac{\sum_{t=2}^T [y_t^2 - 2c_{m+1} y_t e^{x_t} + c_{m+1}^2 e^{2x_t}]}{T}, \ g_{m+1} = \frac{\sum_{t=2}^T (x_t - a_{m+1}x_{t-1}) p^m_{1,t}}{\sum_{t=2}^T p^m_{1,t}}, \\
h_{m+1} &= (\frac{\sum_{t=2}^T (x_t - \mu_{1,t})^2 p^m_{1,t}}{b_{m+1}^2 \sum_{t=2}^T p^m_{1,t}})^{1/2} - 1, \ \lambda_{m+1} = \frac{\sum_{t=2}^T p^m_{1,t}}{T} \ .
\end{aligned} \tag{6}
$$

Using this algorithm we can identify the stochastic processes that were generated from the model. We illustrate this point by simulating $T = 1000$ observations from the six processes from Figure (2). Given the data and starting from an initial guess of $\theta_0 = [.5, .5, .5, .5, .5, .5, .5]$, we obtain an estimate of θ after $m = 100$ updates. From Figure (3), we can see that the algorithm is able to reduce the error from the initial θ in all six cases. So, here we have developed a novel routine for the identification of various stochastic processes.

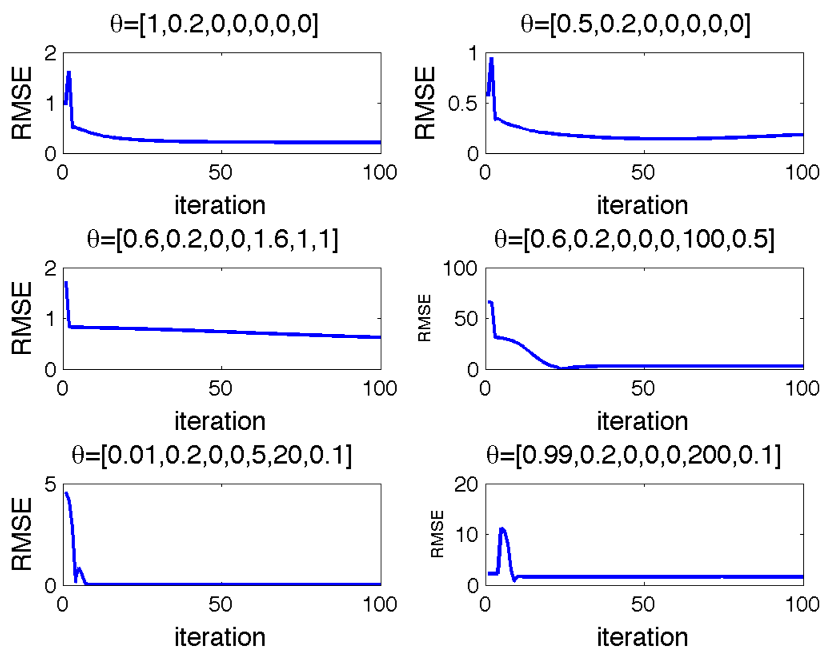

Figure 3: Evolutions of RMSE between the $\widehat{\theta}$, obtained using the EM algorithm, and θ for different processes. Top left is for a random walk process, top right is for an AR(1) process, middle left is for a mean reversion process, middle right is for a discrete stochastic volatility process, bottom left is for a stochastic process with spikes, and bottom right is for a AR(1) process with jumps.

We would like to extend this model beyond the identity function. We set the observation function to be equal to $f(x) = ce^{x_t}$. This would give us a model in the form of the dynamic system, given in (2). When both X and Y are known, we can still use the same updating algorithm to estimate the parameters θ. To illustrate this, we simulate $T = 1000$ observations from Equations (1)-(2), assuming the following parameters: $\theta = [0.60, 0.20, 2.00, 0.20, 2.00, 30.00, 0.05]$. This process can be used to simulate the futures prices observations of electricity with a typical price range around 2, with occassional spikes. The first panel of Figure 4 gives the graphical representation of this process. We apply the EM estimation procedure to this model for over $m = 100$ updates and obtain very good estimates of θ. Starting from an initial guess of $\theta_0 = [.5, .5, .5, .5, .5, .5, .5]$, we obtain the following estimate of θ: $\widehat{\theta} = [0.59, 0.19, 2.00, 0.20, 1.82, 34.05, 0.05]$. Comparing this final update to the true $\theta = [0.60, 0.20, 2.00, 0.20, 2.00, 30.00, 0.05]$, given earlier, it is clear that we are able to estimate the parameter quite accurately. In the second panel of Figure 4, we see the graphical representation of the decreasing norm error between $\widehat{\theta}$ and θ at each iteration of the EM algorithm. The last panel shows the evolution of the log likelihood, which is increasing as expected.

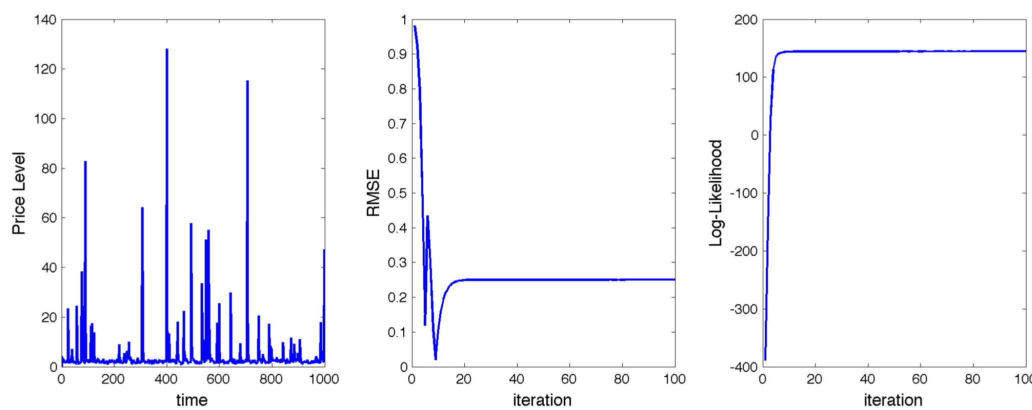

Figure 4: The first panel plots the futures price observations simulated from θ. The second panel (1,1) plots the RMSE between the $\widehat{\theta}$ and θ at each iteration of the EM algorithm. The third panel computes the log likelihood at each iteration of the EM algorithm

3 Conclusions

In this paper, we have developed a Time Indexed Expectation Maximization (TSEM) technique for the estimation of nonlinear dynamical systems. We develop this methodology specifically to test the application of it on datasets with extreme spikes. We performed several simulations to show the effectiveness of the parameter estimation on such datasets. Finally we successfully applied the algorithm to a electricity time series. Our results collectively suggest that this procedure has the potential to be a very powerful tool in the commodities literature.

Our main contribution in this paper is to demonstrate the benefit of implementing the TSEM procedure, particulary on datasets involving mixture densities. We have not explored whether or not the model used is the best choice to describe the data; howevere, our intent was not to implement the best model, but to improve the estimation of model specifications with similar features. The next step in this line of research will be to develop of robust model specification for commodities such as power; one that incorporates multiple states along with many of the nuances obtained within this market.

References

[1] S. Deng and S Oren. Electricity derivatives and risk management. *Science Direct*, 31:940–953, 2006.

[2] R. Gibson and E.S. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 45:959–976, 1990.

[3] K. Miltersen and E.S. Schwartz. Prcing of options on commodity futures with stochastic term structures of convenience yields and interest rates. *Journal of Financial and Quantitative Analysis*, 33:33–59, 1998.

[4] D. Ribeiro and S. Hodges. *A Two-Factor Model for Commodity Prices and Futures Valuation*. Basel Meetings Paper, 2004.

[5] E. Schwartz. The stochastic behaviour of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52:923–973, 1997.

[6] E.S. Schwartz and J. Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46:893–911, 2000.

[7] D. Smith and U. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley, first edition, 1985.