LINEAR DYNAMICAL SYSTEMS:
A MACHINE LEARNING FRAMEWORK FOR FINANCIAL TIME SERIES ANALYSIS

KEMBEY GBARAYOR JR

Advisor: Professor Amy Greenwald
Department of Computer Science, Brown University, Providence, RI, USA

Introduction

Linear dynamical systems are a class of probabilistic models capable of capturing the temporal structure of Gaussian stochastic processes. This paper presents an application of the linear dynamical system paradigm and the associated machine learning algorithms to financial time series analysis. In particular, we develop an unsupervised learning framework to represent the evolution of observed market returns for an individual asset as a perturbed random walk controlled by a set of unknown parameters. In the first part of our study, the maximum likelihood model parameters are found numerically via the Kalman Filter EM algorithm. In the second part it is shown that, given a series of observed market returns, the fitted model can be used to estimate the associated series of unobserved mean returns. The efficacy of our model is tested in a trading simulation of six real financial assets.

1. Definitions and Terminology

The following definitions from probability theory provide the basis for our discussion of stochastic models and dynamical systems.

Definition 1.0.1. A random variable X is a function from the sample space Ω to the real line $\mathbb{R}$. Its cumulative distribution function (cdf) F is a non-decreasing function between zero and one, such that $\forall$ x $\in \mathbb{R}$

$$F(X) \doteq P(X \leq x) \tag{1.0.1}$$

Definition 1.0.2. A random variable is said to be discrete if it can take on at most a countable set of possible values with probability

$$P(X = x_i) \doteq p(x_i) \tag{1.0.2}$$

where $\sum_i p(x_i) = 1$.

Definition 1.0.3. A random variable is called continuous if there exists a nonnegative function f(x), called the density, such that $\forall$ B $\subseteq$ $\mathbb{R}$

$$P(X \in B) \doteq \int_B f(x)dx \tag{1.0.3}$$

where $\int_{\mathbb{R}} f(x)dx = 1$.

Definition 1.0.4. A random vector is a collection of random variables $\mathbf{X} = [X_1, X_2, ..., X_k]$ that maps the sample space Ω to $\mathbb{R}^+$. Its cdf is defined

$$F(x_1, x_2, ..., x_k) \doteq P(X \leq x_1, x_2, ..., x_k) \tag{1.0.4}$$

Definition 1.0.5. A real valued random variable X is said to follow a normal or Gaussian distribution, if its continuous probability density function is the Gaussian function

$$N = \frac{1}{(2\pi\sigma^2)^{1/2}}\ exp[-\frac{1}{2\sigma^2}(x-\bar{x})^2] \tag{1.0.5}$$

A Gaussian random variable is fully characterized by its mean $\bar{x}$ and variance σ, denoted, N($\bar{x}$, σ).

Definition 1.0.6. The multivariate Gaussian distribution is the generalization of a one dimensional Gaussian distribution $\mathbf{X} = [X_1, X_2, ...X_D]$ such that every linear combination $\mathbf{X} = a_1X_1 + a_2X_2 +, ..., +a_DX_D$ is normally distributed. The multivariate Gaussian distribution takes form

$$N = \frac{1}{(2\pi)^{D/2}}\ \frac{1}{|\Sigma|^{1/2}}\ exp[-\frac{1}{2}(\vec{x}-[\bar{x}])^T\Sigma^{-1}(\vec{x}-[\bar{x}])] \tag{1.0.6}$$

where $[\bar{x}]$ is a D-dimensional mean vector, Σ is a DXD covariance matrix, and $|\Sigma|$ denotes the determinant of Σ. A multivariate Gaussian random variable is fully described by its parameters, the collection of means $[\bar{x}]$ and covariance Σ, denoted N($[\bar{x}]$,Σ).

Definition 1.0.7. In general a stochastic process is defined as a collection of random variables $\mathbf{X} = [X_t : t \in T]$ on some probability space $(\Omega, \mathbb{P})$ where X_t is an X valued random variable and t is the time index of the process. An event ω in the sample space Ω is referred to as a sample path denoted $[\mathbf{X}(\omega\ ) = \mathrm{X}_t(\omega\ ) : \mathrm{t} \in T]$ The possible paths of the process $\mathbf{X}$ is known as the state space denoted ς.

Definition 1.0.8. For a stochastic process $\mathbf{X}$ if T is a grid then $\mathbf{X}$ is referred to as a discrete time process. If T $\in\ \mathbb{R}_{++}$ $\mathbf{X}$ is said to be a continuous time process.

2. Types of Stochastic Processes

2.1. **Random Walk.** A random walk is a simple stochastic process that models an individual walking on a straight line who at each point of time either takes one step to the right with probability p or one step to the left with probability 1 - p. Let $x_0 \in \mathbb{R}$ be the fixed starting point of the process.

Definition 2.1.1. A stochastic process $\mathbf{S} = [S_t]$ is referred to as a simple random walk if

$$S_t = x_0 + \sum_1^t X_t \tag{2.1.1}$$

where $X_t \in [-1, 1]$ and $X_t \sim P(X_t = 1) = p, P(X_t = -1) = 1 - p$

2.2. **Markov Processes.** A Markov process, is a stochastic process such that the conditional distribution for any future state X_{n+1} given the past states $X_0, X_1, ..., X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state. Let P denote the matrix of one step transition probabilities P_{ij}, so that $P_{ij} \leq 0$ and $\sum_j P_{ij} = 1$.

Definition 2.2.1. A stochastic process $\mathbf{X} = [X_0, X_1, ..., X_n]$ is said to be a Markov Chain with transition probability matrix P if

$$P(X_{n+1} = j|X_n = i_n, X_{n-1} = i_{n-1}, ..., X_0 = i_0) \tag{2.2.1}$$

$$= P(X_{n+1} = j|X_n = i_n) \tag{2.2.2}$$

$$= P_{ij} \tag{2.2.3}$$

It follows that a Markov chain is completely defined by its transition probability matrix and the initial distribution of X_0. A one-dimensional random walk can be looked at as a Markov process whose state space ς, is given by the integers $i = \pm 1, 2...$, and for some number $0 \leq p \leq 1, P(Y_k = +1) = p, P(Y_k = -1) = 1 - p$.

2.3. **Gaussian Processes and White Noise.** A stochastic process $\mathbf{X} = [X_t : t \in \mathbb{R}_+]$ is called a Gaussian process, if $X_{t1}, X_{t2}, ..., X_{tn}$ has a multivariate normal distribution for all $t_1, t_2, ..., t_n$. It follows that a Gaussian process is fully described by its parameters of mean $[\bar{x}]$ and covariance Σ. If $X_{t1}, X_{t2}, ..., X_{tn}$ are serially uncorrelated normally distributed random variables with zero mean and constant variance the process is known as Gaussian white noise. In this case, $X_{t1}, X_{t2}, ..., X_{tn}$ are independent for all $t_1, t_2, ..., t_n$.

There are two particularly useful characteristics of Gaussian processes 1.Gaussian processes are stationary in the strict sense and 2. Any linear function of a jointly Gaussian process results in another Gaussian process. (Rasmussen and Williams 2006)

3. Linear Dynamical Systems

A linear dynamical system is a model of a stochastic process with latent variables in which the observed output Y_t and hidden state X_t are related by first order differential equations. The basic, generative model for the dynamical system can be written

$$X_{t+1} = AX_t + W_t \tag{3.0.1}$$

$$Y_t = CX_t + Z_t \tag{3.0.2}$$

$$X_t : m\text{x}1 \quad A : m\text{x}m \quad W_t : m\text{x}1$$
$$Y_t : n\text{x}1 \quad C : n\text{x}m \quad Z_t : n\text{x}1$$

where equation (3.0.1) is said to be the state equation and (3.0.2) is said to be the measurement equation (Welch and Bishop 2001). The latent process $\mathbf{X}$ is assumed to evolve according to simple first-order Markov dynamics with an associated state transition matrix A. The state of the process X_t is a vector valued continuous random variable. At each time step the system produces an output or observable measurement Y_t generated from the current state by a simple linear observation process described by the matrix C. Both the state evolution and the observation processes are corrupted by zero mean white Gaussian noise, W_t and Z_t, with respective covariance matrices denoted Q and R. Further, W_t and Z_t are assumed to be independent(Roweis and Ghahramani 1999).

It follows that $\mathbf{X}$ is a first order Gauss-Markov random process. By the Markov property $\mathbf{X}$ is fully characterized by the distribution π_1 of the initial state X_1. By the Gaussian property X_1 is fully characterized by its mean, which is the distribution π_1, and covariance V_1. As a result:

$$X_1 \sim N(\pi_1, V_1) \tag{3.0.3}$$

We can also formulate the following conditional probability distributions for the states and measurements:

$$P(X_{t+1}|X_t) \sim N(AX_t, Q) \tag{3.0.4}$$

$$P(Y_t|X_t) \sim N(CX_t, R) \tag{3.0.5}$$

4. Kalman Filter EM Algorithm

Technically, linear dynamical systems of the form outlined in (3.0.1) and (3.0.2) are called Kalman filter models. Our primary interest is in the learning or system identification problem associated with Kalman filter models: given an observed sequence of outputs $Y_1, ..., Y_t$ find parameters $\Psi = [A, C, Q, R, \pi_1, V_1]$ which maximize the likelihood, $P(\mathbf{X}, \mathbf{Y}|\Psi)$, of the observed data. To learn these parameters we utilize the Kalman Filter EM algorithm.

4.1. **Mathematical Theory.** Due to the Markov and gaussian properties of the Kalman Filter model, we can formulate the complete likelihood $P(\mathbf{X}, \mathbf{Y}|\Psi)$ of the observed and latent variables as follows:

$$P(\mathbf{X}, \mathbf{Y}|\Psi) = \prod_1^T P(Y_t|X_t) \prod_1^T P(X_t|X_{t-1}) \ P(X_1) \tag{4.1.1}$$

In matrix notation the complete log likelihood can be written as the sum of quadratic forms

$$L(\Psi) \doteq log \ P(\mathbf{X}, \mathbf{Y}|\Psi) = \tag{4.1.2}$$

$$\sum_1^T (\frac{1}{2}[Y_t - CX_t]' R^{-1} [Y_t - CX_t]) - \frac{T}{2} log|R|$$

$$- \sum_1^T (\frac{1}{2}[X_t - AX_t]' Q^{-1} [X_t - AX_t]) - \frac{T-1}{2} log|Q|$$

$$-(\frac{1}{2}[X_1 - \pi_1]' V_1^{-1} [X_1 - \pi_1]) - \frac{T-1}{2} log|V_1| - \frac{T(m+n)}{2} log 2\pi$$

Let $\Gamma(\mathbf{X})$ be any distribution over hidden variables. We construct the following equality for the log likelihood.

$$L(\Psi) = log \int_{\mathbf{X}} P(\mathbf{X}, \mathbf{Y}|\Psi) d\mathbf{X} = log \int_{\mathbf{X}} \Gamma(\mathbf{X}) \frac{P(\mathbf{X}, \mathbf{Y}|\Psi)}{\Gamma(\mathbf{X})} d\mathbf{X} \tag{4.1.3}$$

Then by Jensen's Inequality

$$\geq \int_{\mathbf{X}} \Gamma(\mathbf{X}) log \frac{P(\mathbf{X}, \mathbf{Y}|\Psi)}{\Gamma(\mathbf{X})} d\mathbf{X} \tag{4.1.4}$$

$$= \int_{\mathbf{X}} \Gamma(\mathbf{X}) log P(\mathbf{X}, \mathbf{Y}|\Psi) - \int_{\mathbf{X}} \Gamma(\mathbf{X}) log \Gamma(\mathbf{X}) d\mathbf{X} \tag{4.1.5}$$

$$= F(\Gamma, \Psi) \tag{4.1.6}$$

The Expectation Maximization (EM) algorithm alternates between maximizing F with respect to the distribution Γ and the parameters Ψ, respectively.

$$\textbf{E step: } \Gamma_{k+1} = \underset{\Gamma}{\operatorname{argmax}} F(\Gamma, \Psi_k) \tag{4.1.7}$$

$$\textbf{M step: } \Psi_{k+1} = \underset{\Psi}{\operatorname{argmax}} F(\Gamma_{k+1}, \Psi) \tag{4.1.8}$$

It can be shown that, given a set of known parameters, the maximum in the E step results when Γ is exactly the conditional distribution of $\mathbf{X}$ denoted log $P(\mathbf{X}, \mathbf{Y}|\mathbf{Y})$: at which point the bound becomes an equality $F(\Gamma, \Psi) = L(\Psi)$. Since F= L at the beginning of each M step, and since the E step does not change Ψ, we are guaranteed not to decrease the likelihood after each combined EM step (Roweis and Ghahramani 1999). The E and M steps are alternated repeatedly until the difference

$$L(\Psi^{k+1}) - L(\Psi^k) \tag{4.1.9}$$

changes by an arbitrarily small amount ϵ. Given F is bounded from above by L, under the appropriate conditions the algorithm will converge to a global maximum, yielding the set of maximum likelihood parameters Ψ^* (McLachlan and Krishnan 2008).

4.2. **E Step.** The goal of the E Step is to compute the function Γ that maximizes F. All that is necessary is the specification of the complete data $\mathbf{X}$, and conditional density of $\mathbf{X}$ given the observed data $\mathbf{Y}$. As the choice of the complete data vector $\mathbf{X}$ is not unique, specification of the conditional density is chosen for computational convenience (McLachlan and Krishnan 2008). We follow the specification of Ghahramini and Hinton (1996), who present the E step inference algorithm to compute

$$\Gamma = E[log P(\mathbf{X}, \mathbf{Y}|\mathbf{Y})] \tag{4.2.1}$$

which depends on the following quantities

$$\hat{X}_t \equiv E[X_t|\mathbf{Y}] \tag{4.2.2}$$

$$P_t \equiv E[X_t X_t'|\mathbf{Y}] \tag{4.2.3}$$

$$P_{t,t-1} \equiv E[X_t X_{t-1}'|\mathbf{Y}] \tag{4.2.4}$$

Given $\mathbf{X}$ is a Gaussian process and the covariance matrices V_1, and Q are assumed to be known, the computational problem of inferring Γ amounts to finding the vector $[\hat{X}_t]$ of mean values for the process $\mathbf{X}$. The Kalman Filter inference algorithm is decomposed into a forward and backward recursion, called Kalman Filtering, and Kalman Smoothing, respectively . Let $X_t^\tau \doteq \mathrm{E}(\mathrm{X}_t|\mathbf{Y})$, $V_t^\tau \doteq \mathrm{Var}(\mathrm{X}_t|\mathbf{Y})$, and Ψ be an initialization of the parameters A, C, Q, R, π_1, V_1. In matrix notation the E step is specified as follows:

E-Step$(\mathbf{Y}, \Psi)$

Kalman Filter

$$X_0^1 = \pi_1 \tag{4.2.5}$$

$$V_0^1 = V_1 \tag{4.2.6}$$

$$compute\ the\ foward\ recursion$$

$$X_t^{t-1} = AX_{t-1}^{t-1} \tag{4.2.7}$$

$$V_t^{t-1} = AV_{t-1}^{t-1}A' + Q \tag{4.2.8}$$

$$K_t = V_t^{t-1}C'(CV_t^{t-1}C' + R)^{-1} \tag{4.2.9}$$

$$X_t^t = X_t^{t-1} + K_t(Y_t - CX_t^{t-1}) \tag{4.2.10}$$

$$V_t^t = V_t^{t-1} - K_tCV_t^{t-1} \tag{4.2.11}$$

$$Kalman\ Smoothing$$

$$V_{T,T-1}^T = (I - K_TC)AV_{T-1}^{T-1} \tag{4.2.12}$$

$$compute\ the\ backward\ recursion$$

$$J_{t-1} = V_{t-1}^{t-1}A'(V_t^{t-1})^{-1} \tag{4.2.13}$$

$$X_{t_{-1}}^T = X_{t-1}^{t-1} + J_{t-1}(X_t^T - AX_{t-1}^{t-1}) \tag{4.2.14}$$

$$V_{t-1}^T = V_{t-1}^{t-1} + J_{t_{-1}}(V_t^T - V_t^{t-1})J_{t-1}^{'} \tag{4.2.15}$$

$$V_{t_{-1},t_{-2}}^T = V_{t-1}^{t+1}J_{t-2}' + J_{t-1}(V_{t,t-1}^T - AV_{t-1}^{t-1})J_{t-2}^{'} \tag{4.2.16}$$

$$\hat{X}_t = X_t^T \tag{4.2.17}$$

$$P_t = V_t^T + X_t^T X_t^{T'} \tag{4.2.18}$$

$$P_{t,t-1} = V_{t,t-1}^T + X_t^T X_{t-1}^{T'} \tag{4.2.19}$$

$$\textbf{RETURN}([\hat{X}_t], [P_t], [P_{t,t-1}])$$

Given $\hat{X}_t$ solve for $L(\Psi)$ for this iteration of EM.

4.3. **M Step.** The M step re-estimates the parameters to be used in the E step. Each iteration of the M step computes the values Ψ that maximizes F, by 1. taking the respective partial derivatives $(\frac{\partial F}{\partial \pi_1}, \frac{\partial F}{\partial V_1}, \frac{\partial F}{\partial C}, \frac{\partial F}{\partial R}, \frac{\partial F}{\partial A}, \frac{\partial F}{\partial Q})$ 2. setting them to zero 3. then solving for the value of the respective parameter. In matrix notation the updated parameters are computed as follows:

$$\textbf{M-Step}([\hat{X}_t], [P_t], [P_{t,t-1}])$$

$$\pi_1^{new} = \hat{X}_1 \tag{4.3.1}$$

$$V_1^{new} = P_1 - \hat{X}_1 \hat{X}_1' \tag{4.3.2}$$

$$Re - estimateParameters$$

$$C^{new} = (\sum_1^T Y_t \hat{X}_t)(\sum_1^T P_t) \tag{4.3.3}$$

$$R^{new} = \frac{1}{T} \sum_1^T (Y_t Y_t' - C^{new} \hat{X}_t Y_t') \tag{4.3.4}$$

$$A^{new} = (\sum_2^T P_{t,t-1})(\sum_2^T P_{t-1})^{-1} \tag{4.3.5}$$

$$Q^{new} = \frac{1}{T-1}(\sum_2^T P_t - A^{new} \sum_2^T P_{t-1,t}) \tag{4.3.6}$$

$$\textbf{RETURN}(\pi_1^{new}, V_1^{new}, R^{new}, A^{new}, Q^{new})$$

This completes one iteration or cycle of the Kalman Filter EM algorithm.

5. The Kalman Filter Model applied to Financial Assets

The Kalman filter model as defined in equations (3.0.1) and (3.0.2)is the simplest state space model of a stochastic process and is often used in control theory to describe the imprecise measurement of a stochastic system whose dynamics are assumed to follow a random walk (Harvey 1989). We utilize the same model of a random walk plus noise in the financial setting to describe the relationship between the measured or observed market return and the mean return of an asset at any time t. Accordingly we introduce the following notation:

$$\mu_{t+1} = A\mu_t + W_t \tag{5.0.7}$$

$$Y_t = C\mu_t + V_t \tag{5.0.8}$$

Equation(5.0.7) specifies the random walk process of the mean return and equation(5.0.8) specifies the market return or the signal emitted from the underlying random walk plus noise. We can now use the EM algorithm to find the parameters of the perturbed random walk which maximize the likelihood of observed return data.

5.1. **Convergence Results.** We implement the EM algorithm for linear dynamical systems per Ghahramini and Hinton as outlined in 4.1. Given the values of the log likelihood for our data sets we choose a stopping condition of $\epsilon = .01$ or one hundred iterations (combined EM cycles) of the algorithm. 2D vectors are used to represent the states (the first dimension being the actual value, the second being the rate of change) while the observations are represented by scalars. Secondly, for the E step we choose a random initialization of the parameters. It can be shown that when the distribution in question is assumed to be a simple Gaussian process the initialization is arbitrary (McLachlan and Krishnan 2008).

We test the model on two Gaussian processes with known parameters. Specifically we generate one hundred corrupted observations of Gaussian white noise ($X_t \sim N(0, 2)$) and white noise ($X_t \sim N(4, 2)$), (the difference being that the respective means are not zero). Figure 1. and Figure 2. present the average convergence results for 20 runs of the EM algorithm for each asset - first in terms of the value of the log likelihood per cycle of the EM algorithm then in terms of the change in log likelihood per cycle.

Figure 1. Convergence Results for Gaussian White Noise

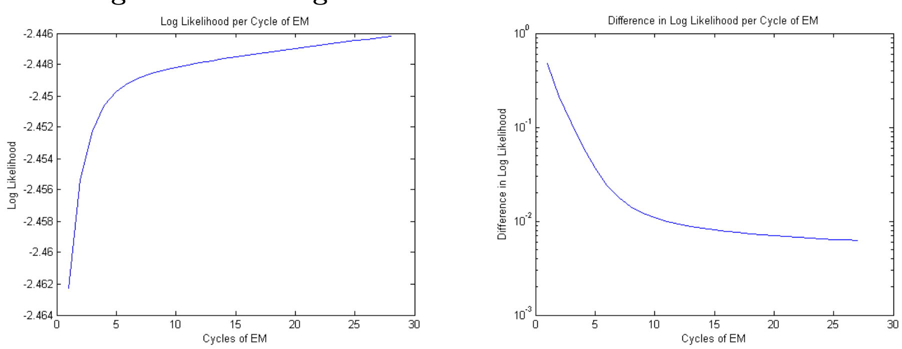
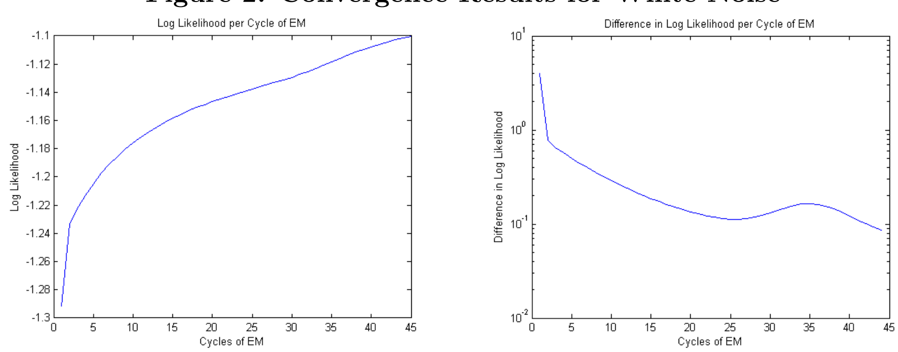


Figure 2. Convergence Results for White Noise



The dummy examples showed that the Kalman filter model is a rich model for true Gaussian processes. The EM algorithm computed values very close to the known parameters (see Appendix). Convergence was also fast for reasonable values of epsilon, specifically $\epsilon = .01$.

Next the EM algorithm is utilized to find the unknown parameters of six financial assets which we assume follow a hidden random walk. We start with 43 years (1960 to 2003) of annual return data for six indices: the Standard and Poors 500

(SP 500), the one year US Treasury bill, the US Money Market Index, the Nasdaq, FTSE, and Nikkei, listed in the order of their annual volatility (standard deviation from the mean) Figures 3. through Figure 9. show the convergence results for these empirical data sets. Some of the final computed parameters are presented in the Appendix.

Figure 3. EM Convergence Results for SP 500

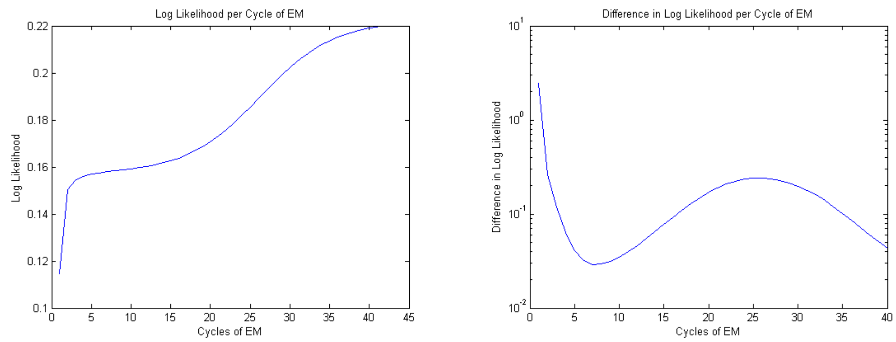


Figure 4. EM Convergence Results for US Treasury

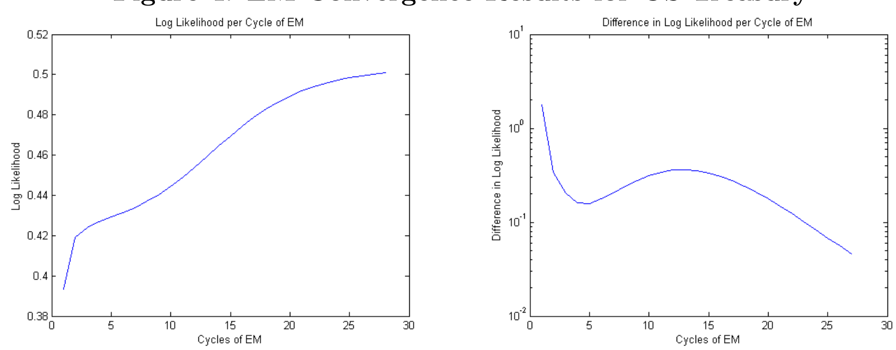


Figure 5. EM Convergence Results for US Money Market

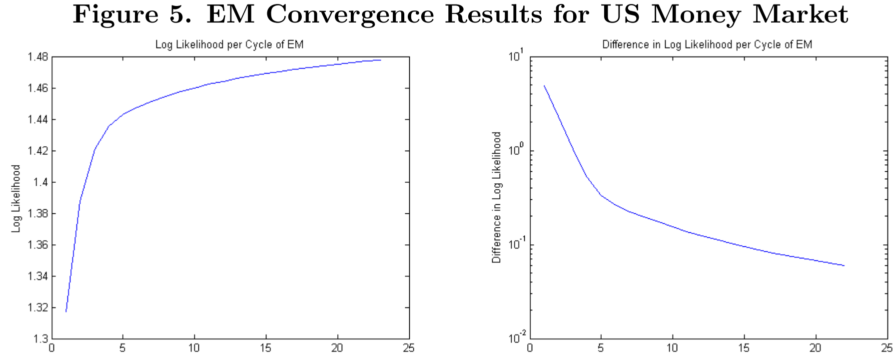


Figure 6. EM Convergence Results for Nasdaq

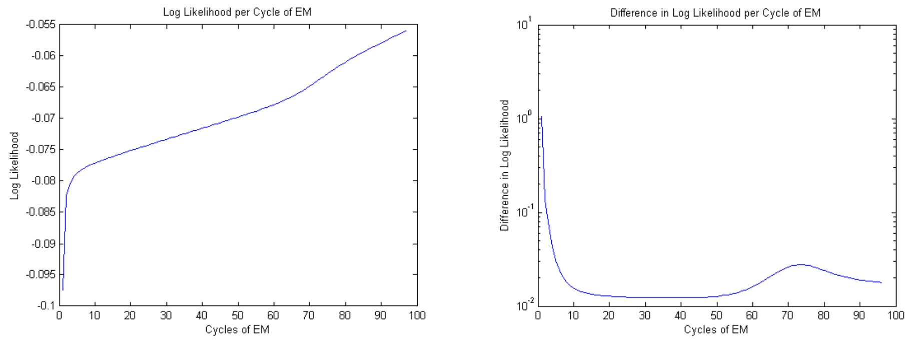


Figure 7. EM Convergence Results for FTSE

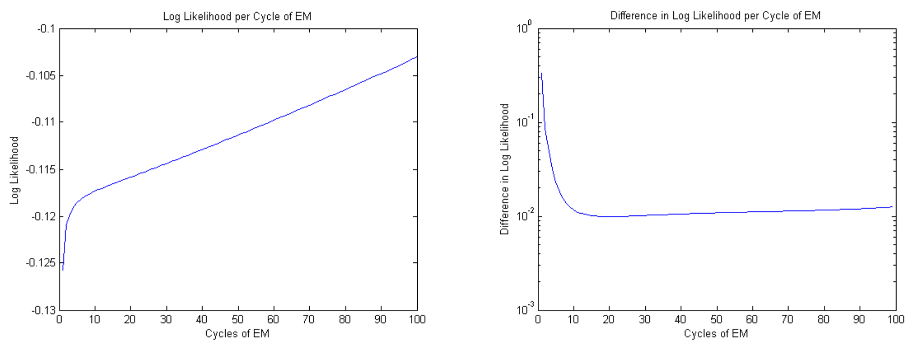
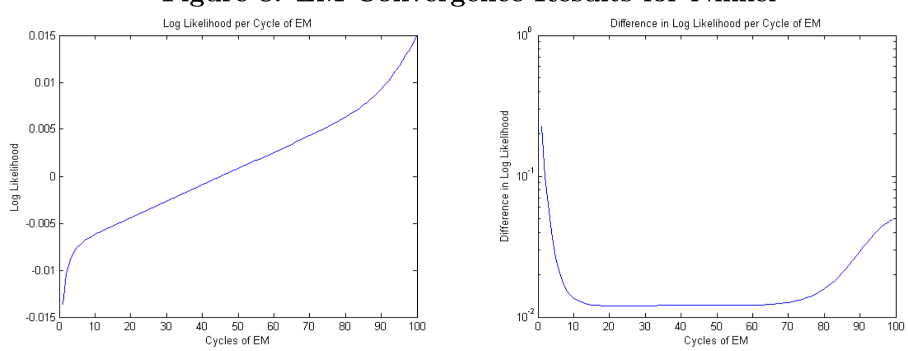


Figure 8. EM Convergence Results for Nikkei

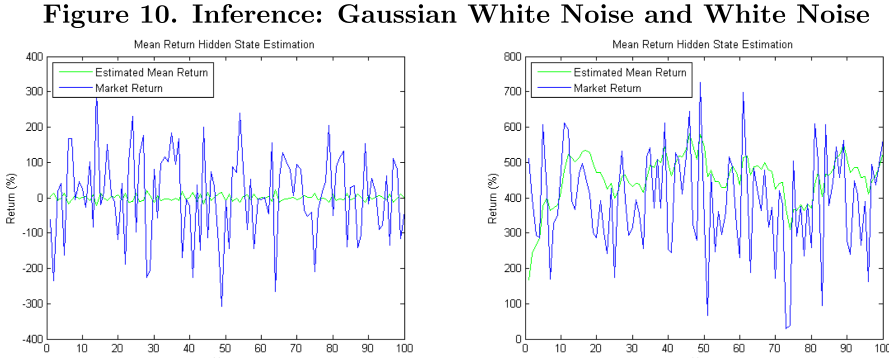


As was the case for the true Gaussian processes, convergence was fast for each financial data set. In few cases did the algorithm take one hundred iterations to terminate. In traditionally volatile markets like the Nasdaq and FTSE the values of the log likelihood were mostly negative, a promising result. Although not a rigorous fact, a rule of thumb is that while the log likelihood may be positive, it is usually a sign that the model does not fit the data or that something has gone wrong in terms of the E-Step initialization (Roweis and Gharamani 1999). In the case of the SP 500, Treasury markets, and Money markets, we computed strictly positive values for the log likelihood. Per the rule of thumb, it may be that these markets follow more complicated stochastic process other than a random walk or are deterministic. There is also a possibility that prior analysis of the distribution is required to choose the best initialization in the E step. Unlike in the simple Gaussian case, if the log likelihood has several local or global maxima and stationary points, convergence of the EM algorithm to either type of point depends on the choice of initialization (Wu 1983). In any case, the random walk model seems to work well in terms of computed values for the log likelihood and speed of convergence, for relatively noisier markets than the SP 500,Treasuries,and Money Market which the Nasdaq, Nikkei, FTSE are,as measured by historical volatility (FinFacts.com 2008). From here on we will refer to value markets as the equal weighted portfolio of the SP 500, Money Market and Treasuries. We will on the other hand refer to the more volatile markets, the equal weighted portfolio of the Nasdaq, Nikkei, and FTSE as the growth markets.

5.2. **Kalman Filter Inference.** The Kalman Filter paradigm is not only useful for the estimation of unknown model parameters but also useful in determining the most likely hidden states given a series of observations. With the maximum likelihood parameters calculated from the EM algorithm, we utilize the Kalman Filter inference algorithm, effectively the E step of the EM algorithm, to compute values of the most likely hidden states, or mean returns of the asset. We follow an analogous setup as in the convergence study, first testing the inference algorithm on the two corrupted Gaussian processes with known parameters, Gaussian white noise ($X_t \sim N(0, 2)$) and white noise ($X_t \sim N(4, 2)$); then finding the hidden states or mean returns of the real financial data sets.

Figure 10. shows the results of carrying out the inference procedure on the generated data sets with known parameters. It should be noted that for consistency, a mean of four for a Gaussian process corresponds to a return of four hundred percent.

Figure 10. Inference: Gaussian White Noise and White Noise



The Kalman Filter inference algorithm derived accurate estimates of the values of the hidden states of the dummy processes, ($X_t \sim N(0, 2)$) and ($X_t \sim N(4, 2)$), respectively. Thus if a process is truly Gaussian, the Kalman Filter framework is effective at finding the latent states of such a process given a series of observations.

6. Technical Analysis in Financial Markets

Financial theory is built upon the idea that all assets have some intrinsic although unobserved expected return. It is assumed that assets are mean reverting and that an asset which is dislocated (has a return higher or lower than its expected return) will eventually trend towards the unobserved mean. If one can infer the mean over a given time period one can in theory profit from actively trading the asset rather than a buy and hold strategy. Traditionally this mean is computed as the arithmetic mean of the asset returns over some time period. We hypothesize that because the arithmetic mean assumes that returns at each decision epoch are independent, and our model assumes returns have temporal structure (first order Markov dependence), we will better capture the mean level of the asset and subsequently attain higher profit in the trading simulation. For specificity we define the following:

Definition 6.0.1. A buy and hold strategy (BH) is one in which an investor simply buys an asset and does not sell it.

Definition 6.0.2. A LDS trading strategy (LDS)is one in which a trader at any decision epoch shorts the asset(sells the asset) if it its market return is above the mean inferred from the Kalman Filter inference algorithm and goes long the asset (buys the asset) if its market return is below the mean return derived from the the Kalman Filter inference algorithm.

Definition 6.0.3. An Simple Arithmetic Mean trading strategy (STMA) is one in which a trader at any decision epoch goes short the asset(sells the asset) if it its market return is above the moving three year arithmetic mean return and takes a long position in the asset (buys the asset) if its market return is below the moving three year arithmetic mean return. The three year moving average is a popular metric in many quantitative macro trading strategies. Let y_t be the observed market return at time t. Then the simple three year moving average at time t is computed as follows:

$$STMA_t = \frac{y_{t-1} + y_{t-2} + y_{t-3}}{3} \tag{6.0.1}$$

6.1. **Inferring the Asset Mean.** As demonstrated with the gaussian dummy processes in the previous section, with the maximum likelihood model parameters computed fore each asset, we can apply the Kalman Filter inference algorithm outlined in the E step to yield the unobserved series of mean returns $[\hat{\mu_t}]$ associated with the observed market returns; that is $[\hat{\mu_1}, \hat{\mu_2}, ..., \hat{\mu_t}]$ given a series of observations $Y_1, Y_2, ...Y_t$. In Figure 11, Figure 12 and Figure 13., we present the results of carrying out the inference procedure to find $[\hat{\mu_1}, \hat{\mu_2}, ..., \hat{\mu_t}]$ for each of the financial data sets, along with the computed three year moving arithmetic average.

Figure 11. Inference: SP and US Treasury

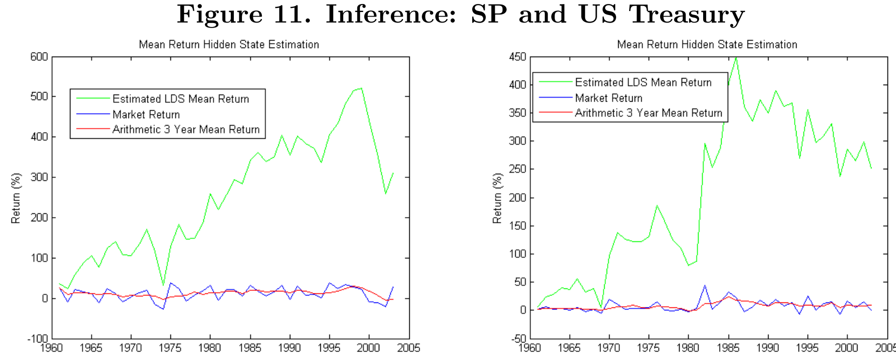


Figure 12. Inference: Money Market and Nasdaq

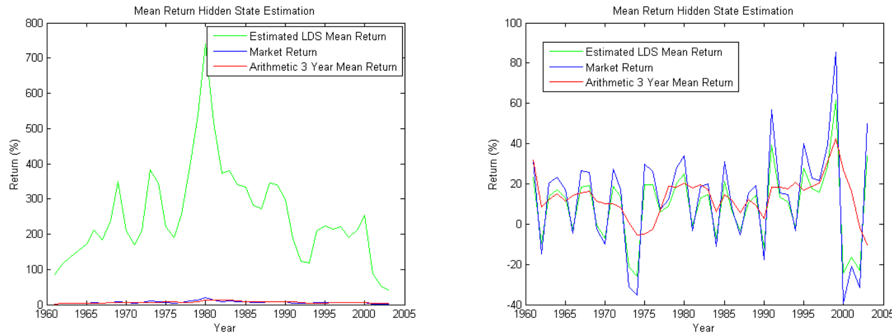
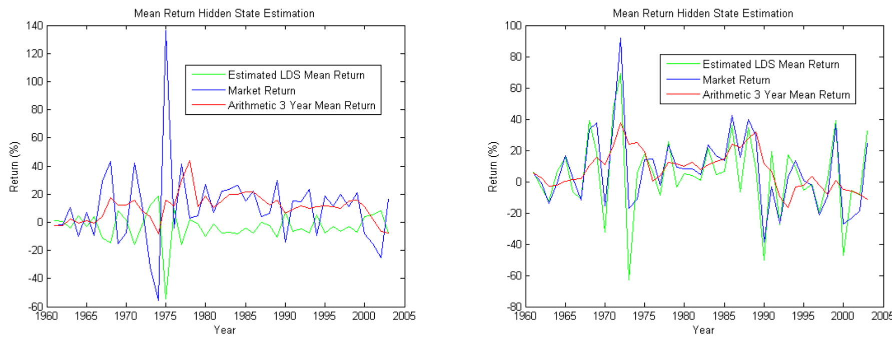


Figure 13. Inference: FTSE and Nikkei



6.2. **Trading Results.** In Table 1 we present the results for the trading simulation as described in section 6. The values in the table represent the return at year end 2003 for an investor who invests one dollar in 1963 in each strategy, takes profit/loss each year, and then invests another dollar in each strategy.

Table 1. Inference: Individual Asset Trading Simulation Results

Case	BH	STMA	LDS
SP	518.45	516.03	1.87
MM	337.56	369.42	-1.67
TR	271.89	7.04	-1.35
NSDQ	554.66	757.63	879.89
FTSE	455.30	961.17	820.79
Nikkei	320.03	543.76	590.94

In Table 2 we present the aggregate results for the trading simulation as described in section 6. We refer to two styles of equal weighted portfolios value and growth. The main distinction is historical volatility of the underlying assets. We compare how our model performs in each type of market. We then use these mean returns to simulate the three trading strategies defined in section 6.

Table 2. Aggregate Results: Equal Weighted Style Portfolios

Case	BH	STMA	LDS	LDS Edge over BH	LDS Edge over STMA
Market	2457.89	2290.48	3155.05	-706.09	-1209.039
Value	1127.90	-1.15	892.49	-1129.05	-893.64
Growth	1329.99	2291.62	2262.56	961.63	29.07

6.3. **Analysis Table 1 and 2.** Although the true mean returns, $[\mu_1, \mu_2, ..., \mu_t]$, in this framework are unknown there were some interesting observations that support our hypothesis that asset returns can be modeled as linear Gaussian systems. We focus our analysis on generalizations about the performance of the LDS model in value versus growth markets as summarized in Table 2. We again saw a discrepancy between the less volatile value markets (SP 500, Treasury, Money Market) and the more volatile growth markets (Nasdaq, FTSE, Nikkei). For the more volatile markets, those which in the previous section had negative likelihoods, the inference algorithm computed a smooth progression of estimated mean returns given the temporal structure of the data. In these markets, our strategy outperformed the buy and hold strategy and the arithmetic mean trading strategy by a significant margin, 960 percent and 29 percent respectively.

On the other hand, for the three assets, those which comprise the value markets, and had positive likelihoods (SP,Treasuries,Money Markets),the estimated returns derived from the LDS inference algorithm were characterized by significant disparities between the estimated LDS return and the market return, consistently in excess of three hundred percent and at times as great as seven hundred percent). Although the LDS model under-performs in the trading simulation in these markets, the model still provides valuable information. In fact, in these markets the model always signals to buy, in other words, the SP, Treasuries, and Money Market are systematically undervalued. In fact this is what we found, as the buy and hold strategies outperformed the other two strategies in these markets. From our results, the LDS mean when computed for growth markets is effective at providing excess return over both the buy and hold and simple three year average trading strategies.

6.4. **Statistical Significance of LDS Model.** For a given asset, the statistical significance of differences between the mean of a sample of observed market returns and a sample of LDS estimated returns, can be assessed using the p-value calculated as part of a t-test (Mackay 2003). In this case the metric for significance is a p-value of 0.05; that is, if the calculated p-value for the difference of means t-test is below 0.05 we reject the null hypothesis that the two means are from independent samples of the same population. Therefore for p-values above 0.05 we conclude the means are from independent samples from the same population, in our case, this means the LDS returns estimated for the asset are drawn from the same population as the observed market returns for that asset. We present these results in Table 3. For those assets for which we can conclude that the LDS mean and market return are from the same population, we further conclude that the LDS model is statistically significant and that trading around the LDS mean is a valid trading strategy for that asset.

Table 3. Inference: Statistical Significance of LDS Model

Case	p-value	Are LDS returns from a statistically different population than the Observed Market Returns?
SP	1.97239E-18	yes
MM	1.85905E-15	yes
TR	6.6273E-20	yes
NSDQ	0.533402074	no
FTSE	0.001884883	yes
Nikkei	0.795323065	no

Table 4 shows of the result of performing the trading simulation only in the markets in which our model is statistically significant, meaning the LDS model computes mean values from the same poulation as the observed market returns. We compare the results across the three trading strategies for an equal weighted portfolio of the Nasdaq and Nikkei (the previous growth portfolio without the FTSE).

Table 4. Inference: Trading Simulation for Statistically Significant Portfolio

	BH	STMA	LDS
Growth Ex FTSE	874.69	1301.39	1470.83
LDS Edge	596.14	169.44	NA

6.5. **Analysis Table 3 and 4.** We find that in the case of the Nikkei and Nasdaq we get p-values that indicate that the LDS returns computed are from the same population or stochastic process that generated the observed market returns. This is promising given these are the two markets in which our model outperformed the other two strategies. The result from our empirical study, is that we can attain higher profibility using the LDS mean in the markets where our model is statistically significant. In particular, given our data set, over the forty year span from 1963 to 2003 for every dollar put into our strategy one would earn 516 percent, and 169 percent, excess return over a buy and hold strategy, and arithmetic mean trading strategy, respectively.

7. Conclusion

This paper is an investigation of linear dynamical systems, a useful tool for the artificial intelligence practitioner, notably those interested in unsupervised learning, pattern recognition and time series analysis. In particular we use a Kalman filter framework as a model for financial time series which we assume have temporal covariance and spatial Gaussian structure. We ultimately showed that the linear dynamical systems framework provides an effective solution to two problems encountered in the examination of financial time series 1. estimating the parameters that control the stochastic behavior of market returns, and 2. inferring the true mean return given a noisy market return. Further, we showed that trading in financial markets with moderate levels of volatility (noise) using the LDS mean is a profitable trading strategy which outperforms both a buy and hold strategy as well as a arithmetic mean trading strategy. Future work may include time series analysis using nonlinear models such as extended Kalman filters or discrete state analogs to the LDS framework, namely hidden Markov models, both capable of parameter estimation and inference for more complex stochastic processes than random walks.

References

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* Springer Science and Business Media, LL., New York, NY, USA

Harvey, A. (1989). *Forecasting Structural Time Series Models.* Cambridge University Press., Melbourne, Australia

McLachlan, G., Krishnan, T. (2008). *The EM Algorithm and Extensions.* John Wiley Sons,Inc., Hoboken, NJ, USA

Ghahramani, Z., Hinton, G.(February 1996). Parameter Estimation for Linear Dynamical Systems. *Technical Report, CRG-TR-96-2.* Department of Computer Science, University of Toronto, Toronto, Canada

Rasmussen, C., Williams, C. (2006). *Gaussian Processes for Machine Learning.* MIT Press., Cambridge, MA, USA

Roweis, S., Ghahramani, Z. (1999). A Unifying Review of Linear Gaussian Models. *Neural Computation Volume 11 No. 2*

Shumway, R. , Stoffer, D. (1982). An Approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4):263-264.

Wang, Hui.(2008) Probabilistic Models Course Notes. Unpublished manuscript. Brown University, Providence, RI,USA.

Welch, G., Bishop, G. (2001). An Introduction to the Kalman Filter. *SIGGRAPH August 12-17.*, Los Angeles, CA,USA

Wu, Jeff C.F. (March 1983). On the Convergence Properties of the EM algorithm. *The Annals of Statistics Vol. 11, No. 1, 19-103.*

Mackay, D. (2003). *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, Cambridge, UK

Price data obtained from Finfacts.com. Retrieved April 2008. http://www.finfacts.com.