Appears in CIFEr '98, March 29-31, New York City

MIXED MEMORY MARKOV MODELS FOR TIME SERIES ANALYSIS

Constantine P. Papageorgiou

Center for Biological and Computational Learning
and
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
cpapa@ai.mit.edu

Abstract

This paper presents a method for analyzing coupled time series using Markov models in a domain where the state space is immense. To make the parameter estimation tractable, the large state space is represented as the Cartesian product of smaller state spaces, a paradigm known as factorial Markov models. The transition matrix for this model is represented as a mixture of the transition matrices of the underlying dynamical processes. This formulation is know as *mixed memory Markov models*. Using this framework, we analyze the daily exchange rates for five currencies – British pound, Canadian dollar, Deutsch mark, Japanese yen, and Swiss franc – as measured against the U.S. dollar.

1 Introduction

Markov models have achieved a prominent position in the analysis and recognition of time series in several domains, most notably, speech recognition [3] and natural language processing. Their application in the domain of financial time series has been limited, at least partially due to the problem of parameter estimation; for a Markov model with n possible states that uses the past k states to predict the future, the number of parameters that need to be estimated are n^{k+1}.

This paper presents a framework for analyzing coupled time series using Markov models that reduces the burden of estimating the (usually large number of) parameters involved in all but the most trivial of models. For this, we appeal to work in factorial Markov models [2] and mixed memory Markov models [4].

2 Markov models

Markov models offer a stochastic interpretation of time series; the next event has a probabilistic dependency on the past k events. The most trivial Markov model is a Markov chain, a simple integer time process composed of a set of n states. The current state is temporally linked to k states in the past via a set of n^k transition probabilities. We denote the state of the system by i_t and the possible states as $i_t \in \{1, 2, \ldots, n\}$. Let $a(i_t|i_{t-1}, i_{t-2}, \ldots, i_{t-k})$ denote the transition probability for a model of order k; this transition probability is the probability of seeing state i_t given the previously observed sequence of states $i_{t-1} \leftarrow i_{t-2} \leftarrow \ldots \leftarrow i_{t-k}$ extending backwards in time.

Using a Markov model for coupled processes (see Figure 1), where a time series has several components, is also straightforward at the conceptual level at least. The most obvious way of modeling the coupling in this componential series is by defining a state space where each state is a Cartesian product of the components; for a three component series, the state space becomes $I_t \in \{\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}\}$. This combinatorial explosion in the size of the state space leads to large increases in the number of parameters that need to be estimated in the transition matrix.

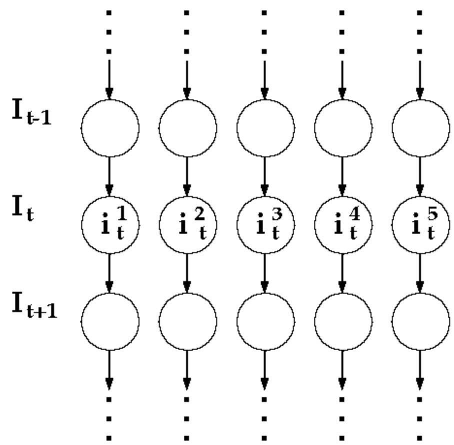

Figure 1: A coupled time series; I_t denotes the state at time t and i_t^μ denotes the μth component of I_t.

Mixed memory Markov models [4] overcome this problem by representing the transition matrix as a convex combination of the elementary transition matrices of each underlying component; the description presented here closely follows that presented in [4]. Let I_t denote the tth element of a coupled time series and i_t^μ denote the μth component of I_t. If the time series has k components and each component can have n values, the resulting state space has n^k elements; in a traditional first order Markov model, this would mean a transition matrix with $O(n^{2k})$ parameters to estimate. For even small values of n and k, this is usually not tractable; the problem is further exacerbated when we are typically dealing with small data sets. The new *mixed memory* transition matrix has the following form, for a first order model:

$$P(I_t|I_{t-1}) = \prod_{\nu=1}^{k} \sum_{\mu=1}^{k} \psi^\nu(\mu) a^{\nu\mu}(i_t^\nu | i_{t-1}^\mu) \tag{1}$$

The parameters $a^{\nu\mu}(i'|i)$ are k^2 elementary $n \times n$ transition matrices and $\psi^\nu(\mu)$ can be thought of as mixture coefficients that measure the amount of correlation between the components; the fact that this is a convex combination imposes the constraint that $\sum_{\mu=1}^{k} \psi^\nu(\mu) = 1$. Using this mixed memory Markov model has reduced the number of parameters in the transition matrix to $O(k^2 n^2)$ from $O(n^{2k})$.

The actual values of the parameters $\psi^\nu(\mu)$ in a fully trained model can tell us important information about the correlational structure of the time series. If $\psi^\nu(\mu) = \delta(\mu, \nu)$, where δ denotes the Kronecker delta function, then there is no correlation between the νth component at time t and all of the other components at time $t-1$, i.e. they are independent. If $\psi^\nu(\nu) < 1$, then this means that other components at time $t-1$ influence the νth component at time t.

The parameters of this model are estimated by using an Expectation Maximization (EM) procedure [1], that iterates until the estimated values reach quiescence. We present just the EM update equations here:

$$\psi^\nu(\mu) \leftarrow \frac{\sum_t P(x_t^\nu = \mu | I)}{\sum_t \sum_{\mu'} P(x_t^\nu = \mu' | I)} \tag{2}$$

$$a^{\nu\mu}(i'|i) \leftarrow \frac{\sum_t P(x_t^\nu = \mu | I) \delta(i, i_{t-1}^\mu) \delta(i', i_t^\nu)}{\sum_t P(x_t^\nu = \mu | I) \delta(i, i_{t-1}^\mu)} \tag{3}$$

where x_t^ν is the hidden variable indicating which component of I_{t-1} determines the transition matrix for i_t^ν.

3 Experimental Results

We apply the mixed memory Markov model to an exchange rate time series previously reported on in the literature [5] with the following components, each measured against the U.S. dollar: British pound, Canadian dollar, Deutsch mark, Japanese yen, and Swiss franc. The series is from 06/01/73 to 05/21/87 and the separate components are shown in Figure 2. Each of the currencies' exchange rates are measured at the futures market close in Chicago, so the lead/lag relationships that the model uncovers will not be an artifact of timing differences.

The data is preprocessed such that the actual points we use are the log ratios between consecutive exchange rates and is further coarsely quantized to yield a dictionary of nine states for our model; $S = \{s_1, ..., s_9\}$. We then use the equations presented in Section 2 to estimate the model parameters. Having estimated the parameters, we take two complementary approaches to using this model: *analysis*, which looks at the model parameters to uncover structure in the time series, and *synthesis*, which uses the model parameters to predict future changes.

3.1 Analysis

The degree of dependency of a single currency on each of the currencies in the time series can be determined by looking at the mixture coefficients in the fully trained model; a large mixture coefficient indicates strong first order dependency and a small mixture coefficient points to very little dependency.

Table 1 shows the dependency of the currencies listed in each row on each of the currencies' previous change, indexed in the columns. The results point to a high degree of first order correlation between each of the currencies and the Canadian dollar exchange rate, evident by the large mixing factors under the Canadian dollar column. This result may be due to the fact that the U.S. and Canadian economies are strongly coupled. The diagonal elements of the table are also fairly strong, indicating non-trivial dependency of each currency on it's own prior change, as expected.

Certain higher order dependencies may indeed exist, but our model only uncovers first order correlations.

	British pound	Canadian dollar	Deutsch mark	Japanese yen	Swiss franc
British pound	0.35	0.58	0.04	0.01	0.02
Canadian dollar	0.07	0.91	0.01	0.00	0.00
Deutsch mark	0.46	0.26	0.20	0.00	0.08
Japanese yen	0.07	0.68	0.03	0.19	0.04
Swiss franc	0.36	0.17	0.24	0.01	0.21

Table 1: Mixture coefficients for the currencies, indicating dependency of currency in each row on the currencies in the columns.

3.2 Synthesis/Prediction

As well as determining the degree of dependency between different components of a coupled time series, we are also interested in using the parameters of a trained model to predict future changes in the series. Equivalently, we are interested in using the data we have seen up to time $t-1$ to synthesize the (unseen) data at time t.

With our mixed memory formulation, the probability that at time t component c is s_ℓ is:

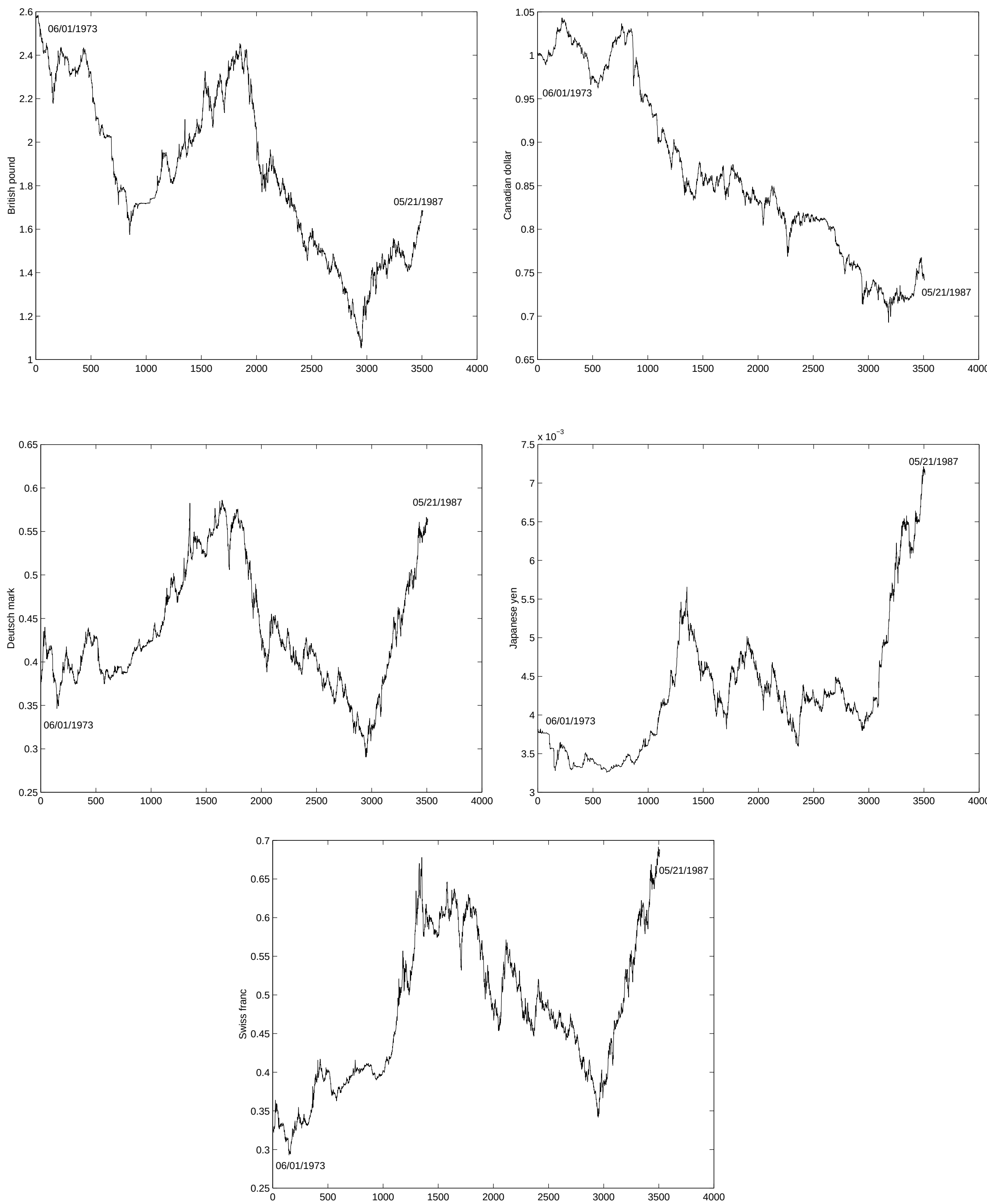

Figure 2: The exchange rate time series we analyze; from left to right, top to bottom, they are: British pound, Canadian dollar, Deutsch mark, Japanese yen, and Swiss franc.

$$P(i_t^c = s_\ell | I_{t-1}) = \sum_{\mu=1}^{k} \psi^c(\mu) a^{c\mu}(s_\ell | i_{t-1}^{\mu}) \tag{4}$$

So, to predict the direction of change of component c at time t, we choose the state s_{pred} that maximizes the above equation:

$$s_{pred} = \arg\max_{s_\ell \in S} P(i_t^c = s_\ell | I_{t-1}) \tag{5}$$

$$= \arg\max_{s_\ell \in S} \sum_{\mu=1}^{k} \psi^c(\mu) a^{c\mu}(s_\ell | i_{t-1}^{\mu}) \tag{6}$$

The model predicts the next direction of change for a single currency based on a mixture of the current directions of change of all the currencies. We estimate the model parameters over a window of 500 days to predict the direction of change for each of the next 500 days. Results are presented in Table 2 and indicate that there is little first order predictability in this series. It remains to be seen whether or not a higher order model would perform better.

	percent of correct predictions of each type	
	Increases	*Decreases*
British pound	45.2% (114/252)	48.0% (119/248)
Canadian dollar	52.0 (129/248)	48.0 (121/252)
Deutsch mark	52.8 (123/233)	51.0 (126/247)
Japanese yen	27.3 (6/22)	44.1 (211/478)
Swiss franc	51.8 (127/245)	50.2 (128/255)

Table 2: Prediction results using the mixed memory Markov model.

4 Conclusion

Simple probabilistic models offer us a clear window into the dynamics and correlational structure of time series and are more readily understood than complex, highly nonlinear systems. In the analysis of coupled time series, the straightforward application of Markov models results in prohibitively large state spaces; the state space of the model developed in this paper would have $9^5 = 59,049$ states. We describe the notion of a factorial model where the state space is defined as the Cartesian product of smaller state spaces and show how using a transition matrix that is actually a mixture of smaller transition matrices results in a computationally tractable model. This formulation is called a mixed memory Markov model and the parameters of the model can be estimated efficiently using the Expectation-Maximization procedure.

Using a five component daily exchange rate series as a testbed for this model, we uncover some of the interdependencies between exchange rates by analyzing the mixture coefficients of the series. We also highlight the predictive capabilities of the model through an experiment that measures performance in determining direction of change; these results indicate that there is little first order predictably in this series and that a higher order model may need to be used.

In our model of the time series, we have assumed only first order dynamics; this is very likely a restrictive assumption and a higher order model may be more effective at capturing the true dynamics of the underlying series. Extending the experiments to use finer resolution data (minutes and ticks) is also a priority. Future work notwithstanding, this paper demonstrates that using a simple yet powerful probabilistic model can yield significant insight into financial time series.

References

[1] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B39:1–38, 1977.

[2] Z. Ghahramani and M.I. Jordan. Factorial hidden markov models. In D.S. Touretzky, M.C. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*. MIT Press, 1996.

[3] L.R. Rabiner and B.H. Juang. An introduction to hidden markov models. *IEEE Acoustics, Speech, and Signal Processing Magazine*, 3(1):4–16, January 1986.

[4] L.K. Saul and M.I. Jordan. Mixed memory markov models. In *Proceedings of the 1997 Conference on Artificial Intelligence and Statistics*, 1997.

[5] A.S. Weigend, B.A. Huberman, and D.E. Rumelhart. Predicting sunspots and exchange rates. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, pages 395–432, 1992.