The Kyoto Economic Review 75(1): 63–90 (June 2006)

On Factorial HMMs for Time Series in Finance

Mohamed Saidane[1] and Christian Lavergne[2]

[1] *I3M, Department of Mathematics, University of Montpellier II.*
E-mail: saidane@math.univ-montp2.fr
[2] *I3M, Department of Mathematics, University of Montpellier II.*
E-mail: Christian.Lavergne@math.univ-montp2.fr

In this article we propose a generalization of the linear factor model, that combines hidden Markov chain Models (HMM) with latent factor models. The HMM generates a piece-wise constant state evolution process and the observations are produced from the state vectors by a factor analysis observation process. This new switching specification provides an alternative, compact, model to handle intra-frame correlation in financial data. Furthermore, it allows variable dimension subspaces to be explored. For maximum likelihood estimation we have proposed an iterative approach based on the Expectation-Maximisation (EM) algorithm. Extensive Monte Carlo simulations and preliminary experiments obtained with a foreign exchange rate data set show promising results, especially for segmentation and tracking tasks.

Keywords: factor analysis, HMM, EM algorithm, finance
JEL Classification Numbers: C51, C52, C61, C63

1. Introduction

The factor Model, also called Index Model, is one of the basic models in finance to analyze and describe the return generation process and the risk/reward relationships of a large number of assets. It has been used extensively in finance for measuring co-movement in and forecasting financial time series. The motivation underlying these models is that there are a few common factors that drive fluctuations in large cross sections of financial and economic time series. Applications of factor model include portfolio construction and sensitivity analysis. Besides, theories, such as Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT), are built upon factor models. The Asset Pricing Model developed by Sharpe (1964), Lintner (1965) and Ross (1976) derived the Arbitrage Pricing Theory which characterizes the expected return on a security as an approximate linear function of the risk premiums on systematic factors in the economy (see Connor and Korajczyk, 1995 for a review).

In asset pricing, since the inception of the arbitrage pricing theory by Ross (1976), there is a growing evidence that high returns are driven by a multifactor model rather than the one factor capital asset pricing model[1)]. For example, Fama and French (1996) argues that the market anomalies[2)] largely disappear in a three-factor model. Connor and Korajczyk (1993) finds evidence for one to six latent factors in the cross-section of stock returns. The appeal and premise of factor analysis extends to a vast literature on forecasting macroeconomic activity and understanding the dynamics of international business cycles. Gregory et al. (1997)[3)] finds that the world common factor is statistically and quantitatively significant. Their results suggest the existence of a single common cycle influencing the countries' business cycles. Stock and Watson (1993) adresses the performance of inflation and industrial production forecasts using a small number of indexes extracted from a large number of predictors. The resulting forecasts outperformed many leading prediction procedures. The term structure literature also benefited from the parsimony of factor representation. Knez et al. (1994) and Dai and Singleton (1999) characterizes the term structure of interest rate as driven by three unobserved factors called "level", "slop" and "curvature". Ang and Piazzesi (2003) estimates a three factors model for the term structure with "mean reverting", "persistent" and "less persistent" as latent factors.

The main assumption of these models is that the relationships between variables has not changed over time, but in recent years, empirical works have shown that the assumption of structural stability is invalid for many financial and economic data sets (e.g., Hamilton, 1990). Many economic time series occasionally exhibit dramatic breaks in their behavior, associated with events such as financial crises (Jeanne and Masson, 2000; Cerra and Saxena, 2005; Hamilton, 2005) or abrupt changes in government policy (Hamilton, 1988; Sims and Zha, 2004; Davig, 2004). Of particular interest to economists is the apparent tendency of many economic variables to behave quite differently during economic downturns, when underutilization of factors of production rather than their long-run tendency to grow governs economic dynamics (Hamilton, 1989; Chauvet and Hamilton, 2005). Abrupt changes are also a prevalent feature of financial data (Ang and Bekaert, 2002; Garcia et al., 2003; Dai et al., 2003).

In this paper, a natural generalization of the standard factor model to a multistate model is achieved by allowing for model transitions that are governed by a hidden Markov chain (HMM) on a set of possible models that describe the different states of volatility. This new specification can be regarded as a dynamic state-space generalization of a multiple component factor analysis system. In this framework, the k-dimensional state vectors are generated by a standard diagonal

---

[1)] CAPM uses only one factor (regressor), the risk premium of the market as a whole, to explain excess returns.

[2)] Patterns in average returns not explained by the one factor CAPM.

[3)] In this paper, the authors show that the world cycle is statistically significant and persistent in the growth rates of output, consumption and investment for the G7 countries.

covariance Gaussian HMM and the q-dimensional observation vectors are generated by a multiple noise component factor analysis observation process.

The proposed model is presented in Section 2, followed by a description of a maximum likelihood estimation procedure, which constitutes the major contribution of this paper, based on the Expectation-Maximisation (EM) principle, the empirical results in Section 4, and a summary in Section 5.

2. Switching Factor Models

Valuation models for assets are based upon the theory of economic behavior in the situation of uncertainty. Valuation models for most kinds of assets rely in almost all cases solely on the first two moments of the return series, that is the means, variances and covariances. It is therefore necessary to model these moments in order to apply the asset pricing models. If one considers q assets and does not impose any restrictions on the model, one has to estimate $\frac{1}{2}(q^2+3q)$ parameters, i.e. q expected returns, q variances and $\frac{1}{2}(q^2-q)$ covariances. Therefore, one tries to introduce a restrictive structure such that the number of parameters to be estimated is significantly reduced without lowering the explanatory power of the model too much.

It is well known that the return series of different assets are correlated with each other, i.e. the assets follow common influences on their returns. This can be used to reduce the number of parameters to be estimated. Various forms of factor models such as the CAPM and the APT are often used. The CAPM treats the correlation of individual assets with the market portfolio, i.e. the portfolio consisting of all stocks in the market with the weights according to the share of the assets in the whole market, as a measure for risk. The APT allows several factors to influence the return series of the assets.

In general, factor models postulate that the return of an asset is composed as the sum of an expected and an unexpected part. The unexpected part of the return is assumed to consist of a systematic portion which cannot be diversified and an unsystematic portion which is specific to the single asset.

Economic theory states that there are common influences such as macroeconomic data which drive the returns of different assets. These are known factors. The systematic unexpected part of the return y_{it} ($i \in \{1, ..., q\}$) is assumed to follow a factor structure. To take into account the possibility of regime switching in stock market returns, we propose a model that combines:

- an HMM structure in order to take into account different states of the world that can affect the evolution of the time series, and
- a linear factor model with constant regime parameters for excess returns.

2.1. Introduction and Notation

Let $\mathbf{y}_t$ denote the q-vector of excess asset returns and $\mathbf{f}_t$ denote the k-vector of latent factor shocks in period t. In order to achieve reduction of the number of

parameters the number of factors k should be much smaller than the number of assets q. The general structure of our switching factor model is given by:

$$S_t \sim P(S_t = j|S_{t-1} = i)$$

$$\text{for } t = 1, ..., n \text{ and } i, j = 1, ..., m$$

$$\mathbf{y}_t = \mathbf{X}_{s_t}\mathbf{f}_t + \varepsilon_t \quad \text{with} \quad \begin{cases} \varepsilon_t \sim \mathcal{N}(\theta_{s_t}, \mathbf{\Psi}_{s_t}) \\ \mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_{s_t}) \end{cases}$$

where $S_t \sim P(S_t = j|S_{t-1} = i)$ is an homogenous hidden Markov chain indicating the state or the regime at the date t. The HMM state transition probabilities from state i to state j are represented by p_{ij}. In an unspecified state $S_t = j$ $(j = 1, ..., m)$, $\mathbf{0}$ and $\mathbf{H}_j$ are, respectively, the $(k\times 1)$ mean vectors and $(k\times k)$ diagonal and definite-positive covariance matrices of the latent common factors $\mathbf{f}_t$; θ_j and $\mathbf{\Psi}_j$ are, respectively, the $(q\times 1)$ mean vectors and $(q\times q)$ diagonal and definite-positive covariance matrices of the $(q \times 1)$ vectors of idiosyncratic noises ε_t; $\mathbf{X}_j$ are the $(q \times k)$ factor loadings matrices, with $q \geq k$ and $rank(\mathbf{X}_j) = k \ \forall \ j$. Here we suppose that the common and specific (idiosyncratic) factors are uncorrelated. We suppose also that $\mathbf{f}_t$ and $\varepsilon_{t'}$ are mutually independent for all t, t'.

Different ways of identifying the factors are discussed in the literature. Some approaches use prespecified factors which base upon macroeconomic data such as the inflation rate. Other lines of research build factors which are linear combinations of the time series considered with prespecified weights derived from economic theory or by using principal component analysis. This paper presents a solution which is based upon the EM algorithm and model selection criteria.

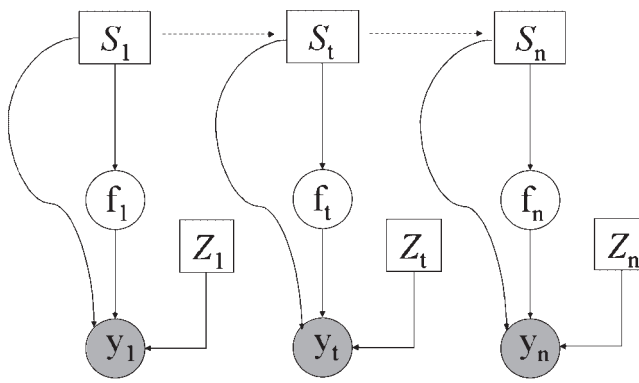

Figure 1 Dynamic Bayesian network representing a Factorial HMM. $\mathbf{Z}_t$'s are eventual exogenous variables that can be introduced in the model as explanatory variables.

A dynamic bayesian network describing a Factorial HMM is shown in Figure 1. The square nodes represent discrete random variables such as the HMM state $\{S_t\}$. Continuous random variables such as the state vectors, $\mathbf{f}_t$, are represented by round nodes. Shaded nodes depict observable variables, $\mathbf{y}_t$, leaving all the other Factorial HMM's variables hidden. A conditional independence assumption is made between variables that are not connected by directed arcs. The state conditional independence assumption between the output densities of a standard HMM is also used in a Factorial HMMs.

So, our generative model can be expressed by the two following Gaussian distributions:

$$\begin{aligned} p(\mathbf{f}_t|S_t = j) &= \mathcal{N}\left(\mathbf{0}, \mathbf{H}_j\right) \\ p(\mathbf{y}_t|\mathbf{f}_t, S_t = j) &= \mathcal{N}\left(\theta_j + \mathbf{X}_j\mathbf{f}_t, \mathbf{\Psi}_j\right) \end{aligned} \tag{1}$$

The likelihood of an observation $\mathbf{y}_t$ given the state $S_t = j$ can be obtained by integrating the state vector $\mathbf{f}_t$ out of the product of the above Gaussians. The resulting likelihood is also a Gaussian and can be written as: $b_j(\mathbf{y}_t) = p(\mathbf{y}_t|S_t = j) = \mathcal{N}\left(\theta_j, \mathbf{\Sigma}_j\right)$, where $\mathbf{\Sigma}_j = \mathbf{X}_j\mathbf{H}_j\mathbf{X}'_j + \mathbf{\Psi}_j$.

2.2. Identification

As it is well known in the literature, the k-factor model must be further constrained to define a unique model free from identification problems. A first constraint is that $\mathbf{X}_t$ be of full rank k, $\forall\ t$ to avoid identification problems arising through invariance of the model under location shifts of the factor loading matrix (e.g., Geweke and Singleton, 1980). Second, we must further constrain the factor loading matrix to avoid overparametrization - simply ensuring that the number of free parameters at time t in the factor representation does not exceed the $q(q+1)/2$ parameters in an unrestricted $\mathbf{\Sigma}_t$. Finally, we need to ensure invariance under invertible linear transformations of the factor vectors (Press, 1985, chapter 10). On this latter issue, our work follows Geweke and Zhou (1996), among others, in adopting the "hierarchical" structural constraint in which the loadings matrix has the form:

$$\mathbf{X}_j = \begin{pmatrix} x_{11j} & 0 & 0 & \dots & 0 \\ x_{21j} & x_{22j} & 0 & \dots & 0 \\ x_{31j} & x_{32j} & x_{33j} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{k1j} & x_{k2j} & x_{k3j} & \dots & x_{kkj} \\ x_{k+1,1j} & x_{k+1,2j} & x_{k+1,3j} & \dots & x_{k+1,kj} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{q1j} & x_{q2j} & x_{q3j} & \dots & x_{qkj} \end{pmatrix}$$

where $x_{i,ij} > 0$ for $i = 1, ..., k$; $j = 1, ..., m$ and $x_{i,lj} = 0$ for $i < l$, $i, l = 1, ..., k$. This form immediately ensures that $\mathbf{X}_j$ is of full rank k.

3. Maximum Likelihood Estimation

The empirical premise and tractability of factor models raised a great interest in the estimation and statistical properties of the estimated structure. In the latent factors literature, the common sources of systematic variation are not observable. These common factors are estimated from sample covariance matrices using statistical techniques like factor analysis and principal components. The maximum likelihood solution was first given by Lawley (1940) and was further developed in two papers by him (Lawley, 1942, 1943). A more condensed derivation of the method appears in a book by Lawley and Maxwell (1963). Rao (1955) related the maximum likelihood method to canonical correlation analysis, and Bargmann (1957) related it to the problem of testing partial independence in multivariate statistical analysis. The estimates obtained by Rao and Bargmann, though derived from principles other than the maximum likelihood principle, satisfy Lawley's likelihood equations, thus constituting another set of maximum likelihood estimates. If identification conditions are imposed, these estimates all become identical.

All the above-mentioned derivations show that the maximum likelihood estimates are determined as the solution of two matrix equations. These equations cannot be solved algebraically; instead some iterative procedure has to be used, such as the procedures proposed by Lawley (1942), Rao (1955), and Bargmann (1957). In carrying out these iterative procedures, certain difficulties have been involved.

A new computational method for the maximum likelihood solution in factor models was proposed by Rubin and Thayer (1982). In this paper the maximum likelihood factor analysis has been conceptualized as maximum likelihood estimation in a multivariate normal model with missing data (Dempster et al., 1977, section 4.7). Consequently, they demonstrated the existence of a corresponding EM algorithm to find maximum likelihood estimates. This algorithm is iterative, and each cycle, which consists of an E step followed by an M step, increases the likelihood of the parameters. The general theory of EM algorithms given in Dempster et al., (1977) proves not only that each iteration of EM increases the likelihood, even if starting from a point where the likelihood is not convex, but also that if an instance of the algorithm converges, it converges to a (local) maximum of the likelihood. Experience with EM algorithms suggests that, although the rate of convergence measured by number of steps can be slow, they reliably converge in a wide range of examples. Another advantage of EM algorithms, such as those for factor models, is that each iteration is simple to program and computationally inexpensive. Even for confirmatory factor analysis with correlations among factors to be estimated and a priori zeros in the factor loadings, each iteration of EM involves only simple matrix manipulations with the most difficult task being the inversion of a $k \times k$ symmetric index, where k is the number of factors. A final advantage of EM algorithms is that they climb the hill of likelihood on which the starting point is located without leaping over valleys in the likelihood; that is, there is a continuous path in the parameter space from the starting point to the stopping point along which the

likelihood monotonically increases.

3.1. The EM Algorithm

A discriminative training scheme such as minimum classification error (e.g., Saul and Rahim, 2000) may be used to optimize the parameters of our model. However, in the present work only maximum likelihood training is considered. We here explicitly define the E and M steps of the algorithm. Thereafter, we present the Forward-Backward algorithm for carrying out the computations.

For a sequence of observation vectors $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$, a sequence of continuous state vectors $\mathcal{F} = \{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n\}$ and a sequence of discrete HMM states $\mathcal{S} = \{S_1, S_2, ..., S_n\}$, the complete likelihood function can be written as:

$$p(\mathcal{Y}, \mathcal{F}, \mathcal{S}|\Theta) = p(S_1) \prod_{t=2}^{n} p(S_t|S_{t-1}) \prod_{t=1}^{n} p(\mathbf{f}_t|S_t; \Theta) p(\mathbf{y}_t|\mathbf{f}_t, S_t; \Theta) \tag{2}$$

where $p(S_1) = \pi_{s_1}$ is initial state probability, $p(S_t|S_{t-1}) = p_{s_{t-1}s_t}$ are the discrete state transition probabilities and $\Theta = \left\{\pi, p_{ij}, \theta_j, \mathbf{X}_j, \mathbf{H}_j, \mathbf{\Psi}_j\right\}$. The auxiliary function that will be maximized is given by:

$$Q(\Theta, \Theta^{(i)}) = \sum_{\forall \mathcal{S}} \int p(\mathcal{F}|\mathcal{Y}, \mathcal{S}, \Theta) p(\mathcal{S}|\mathcal{Y}, \Theta) \log p(\mathcal{Y}, \mathcal{F}, \mathcal{S}|\Theta^{(i)}) \, d\mathcal{F} \tag{3}$$

here the set of current model parameters is represented by $\Theta^{(i)}$. A set of parameters, $\widehat{\Theta}$, that maximize the auxiliary function is found during the maximization step: $\widehat{\Theta} = \arg\max_{\Theta} Q(\Theta, \widehat{\Theta})$. These parameters will be used as the set of old parameters in the following iteration, $\widehat{\Theta} \longrightarrow \Theta^{(i+1)}$.

The first term in equation (3) is the state vector distribution given the observation sequence and the discrete state sequence. For the M step, only the first and second-order statistics are required since the distributions are conditionally Gaussian given the state. Sufficient statistics for the second term can be obtained using the forward-backward algorithm described below.

3.2. Forward-Backward Algorithm

Using the conditional independence assumption in HMMs, the likelihood of being in discrete state j and the observations up to time instant t, $\alpha_j(t)$, is defined by the following recursion:

$$\begin{aligned} \alpha_j(t) &= p(S_t = j, \mathcal{Y}_{1:t}) = p(\mathbf{y}_t|S_t = j) p(S_t = j, \mathcal{Y}_{1:t-1}) \\ &= p(\mathbf{y}_t|S_t = j) \sum_{i=1}^{m} p(S_t = j, S_{t-1} = i, \mathcal{Y}_{1:t-1}) \\ &= b_j(\mathbf{y}_t) \sum_{i=1}^{m} p_{ij} \alpha_i(t-1) \end{aligned} \tag{4}$$

This forward variable is initialized as: $b_1(\mathbf{y}_1)$ if $j = 1$ and 0 otherwise. Using the same independence assumptions, the likelihood of the observations from t to n given being in state i at time instant $t-1$, $\beta_i(t-1)$, are defined by the following recursion:

$$\begin{aligned}\beta_i(t-1) &= p(\mathcal{Y}_{t:n}|S_{t-1}=i) = \sum_{j=1}^{m} p(S_t=j, \mathcal{Y}_{t:n}|S_{t-1}=i) \\ &= \sum_{j=1}^{m} p(S_t=j|S_{t-1}=i)p(\mathbf{y}_t|S_t=j)p(\mathcal{Y}_{t+1:n}|S_t=j) \\ &= \sum_{j=1}^{m} p_{ij}b_j(\mathbf{y}_t)\beta_j(t) \end{aligned} \tag{5}$$

This backward variable is initialized as $\beta_i(n) = 1$ for all $i \in [1, m]$. We define also the probability of being in state j at time t given the observation sequence that is needed in the parameter update formulae.

$$\begin{aligned}\gamma_j(t) = p(S_t=j|\mathcal{Y}) &= \frac{p(S_t=j, \mathcal{Y})}{p(\mathcal{Y})} \\ &= \frac{p(S_t=j, \mathcal{Y}_{1:t})p(\mathcal{Y}_{t+1:n}|S_t=j)}{p(\mathcal{Y})} \\ &= \frac{\alpha_j(t)\beta_j(t)}{\sum_{i=1}^{m}\alpha_i(t)\beta_i(t)} \end{aligned} \tag{6}$$

and the joint probability of being in state i at time instant $t-1$ and in state j at time instant t given the observation sequence which is needed in the transition parameter update formulae.

$$\begin{aligned}\xi_{ij}(t) &= p(S_{t-1}=i, S_t=j|\mathcal{Y}) \\ &= \frac{p(S_{t-1}=i, \mathcal{Y}_{1:t-1})p(S_t=j|S_{t-1}=i)p(\mathbf{y}_t|S_t=j)p(\mathcal{Y}_{t+1:n}|S_t=j)}{p(\mathcal{Y})} \\ &= \frac{\alpha_i(t-1)p_{ij}b_j(\mathbf{y}_t)\beta_j(t)}{\sum_{i=1}^{m}\alpha_i(t)\beta_i(t)} \end{aligned} \tag{7}$$

3.3. Continuous State Posterior Statistics

Given the current discrete state, $S_t = j$, and using the conditional independence assumptions made in the model, the joint likelihood of the current observation and continuous state vector is Gaussian

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{f}_t \end{pmatrix} | S_t = j \sim \mathcal{N}\left[\begin{pmatrix} \theta_j \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{X}_j\mathbf{H}_j\mathbf{X}_j' + \mathbf{\Psi}_j & \mathbf{X}_j\mathbf{H}_j \\ \mathbf{H}_j\mathbf{X}_j' & \mathbf{H}_j \end{pmatrix}\right] \tag{8}$$

The posterior distribution is also Gaussian and can be written as:

$$(\mathbf{f}_t|\mathbf{y}_t, S_t = j) \sim \mathcal{N}\left[\mathbf{K}_j(\mathbf{y}_t - \theta_j), \mathbf{H}_j - \mathbf{K}_j\mathbf{X}_j\mathbf{H}_j\right] \tag{9}$$

where $\mathbf{K}_j = \mathbf{H}_j\mathbf{X}_j'\left[\mathbf{X}_j\mathbf{H}_j\mathbf{X}_j' + \mathbf{\Psi}_j\right]^{-1}$. For parameter update formulae, the statistics $\widetilde{\mathbf{f}}_{jt} = \mathbf{K}_j\left(\mathbf{y}_t - \theta_j\right)$ and $\widetilde{\mathbf{R}}_j = \mathbf{H}_j - \mathbf{K}_j\mathbf{X}_j\mathbf{H}_j$ are also needed.

3.4. Parameter Update Formulae

The parameter optimization scheme based on the expectation maximization (EM) algorithm is presented in this section. All the sufficient statistics are evaluated using the parameters from the previous iteration and therefore writing $\Theta^{(i)}$ explicitly is omitted for clarity. This derivation assumes that the first discrete state is always the initial state and all states are emitting. It is easy to extend the derivation for use with explicit initial discrete state probabilities and to include non-emitting states.

3.4.1. Initial State Probability Update Formulae

Discarding terms independent of the discrete initial state probabilities from equation (3), the auxiliary function can be written as

$$Q(\Theta, \Theta^{(i)}) = \sum_{j=1}^{m} \gamma_j(1) \log(p(S_1))$$

Maximizing this function with respect to the discrete initial state probabilities, π_j, can be carried out using the Lagrange multiplier λ together with the sum to unity constraint $\sum_{j=1}^{m} \pi_j = 1$. It is equivalent to maximizing the following Lagrangian

$$g(\pi_j) = \sum_{i=1}^{m} \gamma_i(1) \log(\pi_j) + \lambda\left(1 - \sum_{i=1}^{m} \pi_i\right)$$

Differentiating $g(\pi_j)$ yields

$$\begin{cases} \frac{\partial g(\pi_j)}{\partial \pi_j} = \frac{\gamma_j(1)}{\pi_j} - \lambda \\ \frac{\partial g(\pi_j)}{\partial \lambda} = 1 - \sum_{i=1}^{m} \pi_i \end{cases}$$

Setting the derivative to zero together with the sum to unity constraint forms the following pair of equations and solving for π_j, the new discrete initial state probabilities can be written as

$$\widehat{\pi}_j = \frac{\gamma_j(1)}{\sum_{i=1}^{m} \gamma_i(1)} \tag{10}$$

3.4.2. Transition State Probability Update Formulae

Discarding terms independent of the discrete state transition probabilities from equation (3), the auxiliary function can be written as

$$Q(\Theta, \Theta^{(i)}) = \sum_{t=2}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \xi_{ij}(t-1) \log(p_{ij})$$

Maximizing this function with respect to the discrete state transition probabilities, p_{ij}, can be carried out using the Lagrange multiplier λ together with the sum to unity constraint $\sum_{j=1}^{m} p_{ij} = 1$. It is equivalent to maximizing

$$g(p_{ij}) = \lambda\left(1 - \sum_{j=1}^{m} p_{ij}\right) + \sum_{t=2}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} \xi_{ij}(t-1) \log(p_{ij})$$

Differentiating $g(p_{ij})$ yields

$$\frac{\partial g(p_{ij})}{\partial p_{ij}} = -\lambda + \sum_{t=2}^{n} \frac{\xi_{ij}(t-1)}{p_{ij}}$$

Setting the derivative to zero together with the sum to unity constraint forms the following pair of equations

$$\begin{cases} -\lambda + \sum_{t=2}^{n} \frac{\xi_{ij}(t-1)}{p_{ij}} = 0 \\ 1 - \sum_{j=1}^{m} p_{ij} = 0 \end{cases}$$

Solving for p_{ij}, the new discrete state transition probabilities can be written as

$$\widehat{p_{ij}} = \frac{\sum_{t=2}^{n} \xi_{ij}(t-1)}{\sum_{t=2}^{n} \gamma_i(t-1)} \tag{11}$$

3.4.3. Factor Loadings Update Formulae

The new factor loadings matrix, $\mathbf{X}_j$, has to be optimized row by row as in Shared factor analysis (e.g., Gopinath et al., 1988). The scheme adopted in this work closely follows the maximum likelihood linear regression transform matrix optimization (e.g., Gales, 1998). Let $\mathbf{x}_{jl}$ denote the *l-th* row vector of $\mathbf{X}_j$. Maximizing the Equation (3) is equivalent to maximizing

$$g(\mathbf{x}_{jl}) = -\frac{1}{2} \sum_{l=1}^{q} \left[\mathbf{x}_{ji} \mathbf{G}_{jl} \mathbf{x}'_{jl} - \mathbf{x}_{jl} \mathbf{k}_{jl}\right]$$

where the k by k matrices $\mathbf{G}_{jl}$ and k-dimensional column vectors $\mathbf{k}_{jl}$ are defined as follows

$$\begin{cases} \mathbf{G}_{jl} = \frac{1}{\psi_{jl}} \sum_{t=1}^{n} \gamma_j(t) \left[\widetilde{\mathbf{R}}_j + \widetilde{\mathbf{f}}_{jt} \widetilde{\mathbf{f}}'_{jt} \right] \\ \mathbf{k}_{jl} = \frac{1}{\psi_{jl}} \sum_{t=1}^{n} \gamma_j(t) (y_{tl} - \theta_{jl}) \widetilde{\mathbf{f}}_{jt} \end{cases}$$

where ψ_{jl} is the *l-th* diagonal element of the idiosyncratic covariance matrix $\mathbf{\Psi}_j$, y_{tl} and θ_{jl} are the *l-th* elements of the current observation and the idiosyncratic noise mean vectors, respectively. Differentiating $g(\mathbf{x}_{jl})$ yields

$$\frac{\partial g(\mathbf{x}_{jl})}{\partial \mathbf{x}_{jl}} = -\mathbf{G}_{jl} \mathbf{x}'_{jl} + \mathbf{k}_{jl}$$

Setting the derivative to zero and solving for $\mathbf{x}_{jl}$ results in the updated row vector of the factor loading matrix

$$\widehat{\mathbf{x}}_{jl} = \mathbf{k}'_{jl} \mathbf{G}_{jl}^{-1} \tag{12}$$

3.4.4. Observation Noise Mean Update Formulae

Differentiating the auxiliary function in equation (3) with respect to the observation noise mean vector, θ_j, yields

$$\frac{\partial Q(\Theta, \Theta^{(i)})}{\partial \theta_j} = \mathbf{\Psi}_j^{-1} \sum_{t=1}^{n} \gamma_j(t) \left(\mathbf{y}_t - \mathbf{X}_j \widetilde{\mathbf{f}}_{jt} - \theta_j \right)$$

Equating this to zero and solving for θ_j result in the updated observation noise mean vector

$$\widehat{\theta}_j = \frac{1}{\sum_{t=1}^{n} \gamma_j(t)} \sum_{t=1}^{n} \gamma_j(t) \left(\mathbf{y}_t - \mathbf{X}_j \widetilde{\mathbf{f}}_{jt} \right) \tag{13}$$

3.4.5. Idiosyncratic Variances Update Formulae

Applying some matrix manipulations and discarding terms independent of the idiosyncratic noise covariance matrix, $\mathbf{\Psi}_j$, the auxiliary function in equation (3) may be rewritten as

$$\begin{aligned} Q(\Theta, \Theta^{(i)}) \quad = \quad & -\frac{1}{2} \sum_{t=1}^{n} \sum_{j=1}^{m} \gamma_j(t) \Bigg((\log |\mathbf{\Psi}_j| + tr \Bigg\{ \mathbf{\Psi}_j^{-1} (\mathbf{y}_t \mathbf{y}'_t - \begin{bmatrix} \mathbf{X}_j & \theta_j \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{f}}_{jt} \mathbf{y}'_t \\ \mathbf{y}'_t \end{bmatrix} \\ & - \begin{bmatrix} \mathbf{y}_t \widetilde{\mathbf{f}}'_{jt} & \mathbf{y}_t \end{bmatrix} \begin{bmatrix} \mathbf{X}'_j \\ \theta'_j \end{bmatrix} + \begin{bmatrix} \mathbf{X}_j & \theta_j \end{bmatrix} \begin{bmatrix} \mathbf{R}_j + \widetilde{\mathbf{f}}_{jt} \widetilde{\mathbf{f}}'_{jt} & \widetilde{\mathbf{f}}_{jt} \\ \widetilde{\mathbf{f}}'_{jt} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}'_j \\ \theta'_j \end{bmatrix} \Bigg\} \Bigg) \end{aligned}$$

To find the new idiosyncratic noise covariance matrix, the auxiliary function above is differentiated with respect to its inverse, $\mathbf{\Psi}_j^{-1}$, and equated to zero. Solving for $\mathbf{\Psi}_j$ and setting the off-diagonal elements to zeroes result in the updated idiosyncratic noise covariance matrix

$$\widehat{\mathbf{\Psi}}_j = \frac{\sum_{t=1}^{n} \gamma_j(t) diag\left[\left(\mathbf{y}_t - \mathbf{X}_j\widetilde{\mathbf{f}}_{jt} - \theta_j\right)\left(\mathbf{y}_t - \mathbf{X}_j\widetilde{\mathbf{f}}_{jt} - \theta_j\right)' + \mathbf{X}_j\widetilde{\mathbf{R}}_j\mathbf{X}_j'\right]}{\sum_{t=1}^{n} \gamma_j(t)} \quad (14)$$

3.4.6. State Noise Covariance Matrix Update Formulae

To find the new state noise covariance matrix, the auxiliary function (3) is differentiated with respect to its inverse, $\mathbf{H}_j^{-1}$, as follows

$$\frac{\partial Q(\Theta, \Theta^{(i)})}{\partial \mathbf{H}_j^{-1}} = \frac{1}{2}\sum_{t=1}^{n} \gamma_j(t)\left\{\mathbf{H}_j - \mathbb{E}\left[\mathbf{f}_t\mathbf{f}_t' | \mathcal{Y}, \Theta^{(i)}\right]\right\}$$

Equating this to zero, solving for $\mathbf{H}_j$ and setting the off-diagonal elements to zeroes results in the updated state noise covariance matrix

$$\widehat{\mathbf{H}}_j = \frac{1}{\sum_{t=1}^{n} \gamma_j(t)} diag\left\{\sum_{t=1}^{n} \gamma_j(t)\left[\widetilde{\mathbf{R}}_j + \widetilde{\mathbf{f}}_{jt}\widetilde{\mathbf{f}}_{jt}'\right]\right\} \quad (15)$$

3.5. The Viterbi Algorithm

In several applications of HMMs (as in speech recognition and molecular biology applications, for example), the hidden state variable is associated with a particular meaning (e.g., phonemes and words, for speech recognition). In our case, the hidden state informs as about the the transition from a regime to another as a reaction to some events. It is therefore useful, given an observed sequence $\mathcal{Y}$, to infer the most likely state sequence $\mathcal{S}$ corresponding to it. It is possible to define the optimal state sequence according to many criterion of optimality. For example, one can maximize the expected number of correct states (S_t, S_{t+1}), or triple of states (S_t, S_{t+1}, S_{t+2}), etc. The most widely used criterion is to find the best single state sequence (path), i.e., to maximize $P(\mathcal{S}|\mathcal{Y}, \Theta)$ which is equivalent to maximize $P(\mathcal{S}, \mathcal{Y}|\Theta)$. The Viterbi algorithm (Viterbi, 1967; Forney, 1973) finds the above maximum with a relatively efficient recursive solution. This technique is based on Bellman's dynamic programming methods (Bellman, 1957).

Our goal is to find the best sequence,

$$\begin{aligned}\mathcal{S}_{1:n}^* &= \arg\max_{S_{1:n}} p(\mathcal{S}_{1:n}|\mathcal{Y}_{1:n}) = \arg\max_{\mathcal{S}_{1:n}} \frac{p(\mathcal{S}_{1:n}, \mathcal{Y}_{1:n})}{p(\mathcal{Y}_{1:n})} \\ &= \arg\max_{\mathcal{S}_{1:n}} p(\mathcal{S}_{1:n}, \mathcal{Y}_{1:n})\end{aligned}$$

where $\mathcal{Y}_{1:\tau} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_\tau\}$ and $\mathcal{S}_{1:\tau} = \{S_1, S_2, ..., S_\tau\}$. Let us define the following quantity, which may be calculated recursively,

$$\begin{aligned} \delta_t(S_t) &= \arg\max_{S_{1:t-1}} p(\mathcal{S}_{1:t}, \mathcal{Y}_{1:t}) \\ \delta_j(t) &= \arg\max_{\mathcal{S}_{1:t-1}} p(\mathcal{S}_{1:t-1}, S_t = j, \mathcal{Y}_{1:t}) \end{aligned}$$

Note again that we maximize over the complete sequence $\mathcal{S}_{1:t-1}$. Clearly, we have that,

$$\begin{aligned} \delta_1 &= p(S_1, \mathbf{y}_1) = p(\mathbf{y}_1|S_1)p(S_1) \\ \delta_j(1) &= p(\mathbf{y}_1|S_1 = j)p(S_1 = j) \qquad \text{and} \qquad (16) \\ \max_{\mathcal{S}_{1:n}} p(\mathcal{S}_{1:n}, \mathcal{Y}_{1:n}) &= \max_{\mathcal{S}_{1:n}} \delta_{1:n} = \max_j \delta_j(n) \qquad (17) \end{aligned}$$

We will now derive the recursion for δ_t, using the Markov structure of the model.

$$\begin{aligned} \delta_{t+1} &= \max_{\mathcal{S}_{1:t}} p(\mathcal{S}_{1:t+1}, \mathcal{Y}_{1:t+1}) \\ &= \max_{\mathcal{S}_{1:t}} \left[p(\mathbf{y}_{t+1}|S_{t+1}) p(S_{t+1}|S_t) p(\mathcal{S}_{1:t}, \mathcal{Y}_{1:t}) \right] \\ &= p(\mathbf{y}_{t+1}|S_{t+1}) \max_{S_t} \left[p(S_{t+1}|S_t) \max_{\mathcal{S}_{1:t-1}} [p(\mathcal{S}_{1:t}, \mathcal{Y}_{1:t})] \right] \\ &= p(\mathbf{y}_{t+1}|S_{t+1}) \max_{S_t} \left[p(S_{t+1}|S_t)\delta_t \right] \qquad \text{hence} \\ \delta_j(t+1) &= p(\mathbf{y}_{t+1}|S_{t+1} = j) \max_i \left[p_{ij}\ \delta_i(t) \right] \qquad (18) \end{aligned}$$

Thus, to find the maximum of $p(\mathcal{S}_{1:n}, \mathcal{Y}_{1:n})$ we initialize the recursion with (16), then compute $\delta_2, ..., \delta_n$ through the above recursions, and finally calculate the overall maximum by (17). Notice that the value of δ_t shrinks as t grows (we are multiplying probabilities). To avoid mathematical underflow, we may have to normalize δ_t at every iteration, for instance to unit length. Notice, that in the end we are only interested in the sequence that maximizes the overall probability and not the maximum probability itself and the renormalizing δ_t only effects the latter.

The above procedure will provide us with the maximum value, but not with the sequence that maximizes it. Therefore we define a quantity that stores the values of S_t which maximize the function $p(S_{t+1}|S_t)\delta_t(S_t)$ in (18) for all values of S_{t+1}, i.e.

$$\begin{aligned} F_{t+1}(S_{t+1}) &= \arg\max_{S_t} \left[p(S_{t+1}|S_t)\delta_t(S_t) \right] \\ F_j(t+1) &= \arg\max_i p_{ij}\delta_i(t) \quad \text{for} \quad t = 1, ..., n-1 \end{aligned}$$

It is now straightforward to find the sequence with highest probability by a process which is called backtracking,

$$\begin{aligned} S_n^* &= \arg\max_j \delta_j(n) \quad \text{and} \\ S_t^* &= F_{t+1}(S_{t+1}^*) \quad \text{for} \quad t = n-1, ..., 1 \end{aligned}$$

Table 1 Simulation parameters Θ_0

	θ	$\mathbf{X}$		$diag(\mathbf{\Psi})$
State 1	2.0000	1.0000	1.0000	1.0000
	2.0000	2.0000	1.0000	2.0000
	2.0000	3.0000	2.0000	1.0000
	2.0000	4.0000	2.0000	2.0000
	2.0000	5.0000	3.0000	2.0000
	2.0000	6.0000	3.0000	1.0000
State 2	1.0000	3.0000	2.0000	5.0000
	1.0000	3.0000	5.0000	5.0000
	1.0000	5.0000	3.0000	4.0000
	1.0000	5.0000	3.0000	4.0000
	1.0000	3.0000	2.0000	5.0000
	1.0000	3.0000	5.0000	5.0000
State 3	3.0000	4.0000	3.0000	2.0000
	2.0000	4.0000	2.0000	1.0000
	3.0000	1.0000	3.0000	2.0000
	2.0000	1.0000	2.0000	1.0000
	3.0000	4.0000	3.0000	1.0000
	2.0000	4.0000	2.0000	2.0000

4. Experimental Results

In this section, we study the performance of our proposed algorithm using synthetic and financial data. The example used for simulation experiments has $q = 6$ observable variables and two latent common factors. We consider the case of three states model with the initial state $S_1 = 1$ and a transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.95 & 0.05 & 0 \\ 0.05 & 0.90 & 0.05 \\ 0 & 0.05 & 0.95 \end{pmatrix}$$

The hypothesis of changing the state with the stock market's reaction to events implies that the properties of the considered hidden chain change over time: as time increases, the state index increases, decreases or stay the same ($S_t = i$), no transition are allowed to states whose indices are lower than $i - 1$ or grater than $i + 1$. The initial parameters for the EM algorithm, were obtained by randomly perturbing the true parameter values (given in Table 1) by up to 40% of their true value. The iterations of the EM algorithm stop when the relative change in the likelihood function between two subsequent iterations is smaller than a threshold value = 10^{-5}.

4.1. Accuracy and Stability of the Estimates

In this experiment we try to estimate the parameters of the model and to study the behavior of the estimates when the size of the sequence n increases. With this

Table 2 Averages and standard deviations (.) for the EM parameter estimates

	θ	$\mathbf{X}$	$diag(\mathbf{\Psi})$
State 1	2.0032 (0.0781)	0.9780 1.0110 (0.0507 0.0466)	0.9443 (0.0619)
	2.0104 (0.0868)	2.0009 1.0046 (0.0450 0.0349)	1.9392 (0.0649)
	2.0126 (0.0725)	2.9705 2.0061 (0.0498 0.0596)	0.9373 (0.0595)
	2.0095 (0.0940)	3.9987 1.9869 (0.0651 0.0559)	1.9660 (0.0635)
	2.0144 (0.1057)	4.9670 2.9665 (0.0594 0.0407)	1.9450 (0.0594)
	2.0139 (0.1063)	5.9751 2.9815 (0.0489 0.0335)	0.9720 (0.0763)
State 2	0.9601 (0.0635)	2.9805 1.9635 (0.0634 0.0536)	4.9367 (0.0753)
	0.9996 (0.0746)	2.9555 4.9800 (0.0736 0.0682)	4.9415 (0.0881)
	0.9629 (0.0786)	4.9716 2.9387 (0.0547 0.0592)	3.9248 (0.0707)
	0.9893 (0.0874)	4.9584 3.0013 (0.0679 0.0644)	3.9621 (0.0573)
	0.9964 (0.0938)	3.0028 1.9587 (0.0528 0.0512)	4.9838 (0.0540)
	0.9735 (0.0781)	2.9897 4.8978 (0.0596 0.0440)	4.9581 (0.0610)
State 3	2.9731 (0.0859)	4.0077 3.0250 (0.0551 0.0507)	1.9770 (0.0575)
	1.9889 (0.0932)	3.9870 2.0002 (0.0542 0.0664)	0.9848 (0.0524)
	2.9900 (0.0860)	1.0399 2.9845 (0.0452 0.0551)	2.0068 (0.0784)
	2.0075 (0.0796)	1.0106 2.0046 (0.0323 0.0480)	0.9437 (0.0635)
	2.9817 (0.0783)	3.9987 2.9994 (0.0661 0.0558)	0.9714 (0.0541)
	1.9755 (0.0903)	3.9878 1.9974 (0.0525 0.0470)	1.9652 (0.0682)

intention, we generated sequences of observations of sizes $n = 500$, 600, 700 and 800, with a hundred replications for each simulation. We have used the empirical Kullback-Leibler divergence $\widetilde{K}_n(\Theta_0, \widetilde{\Theta}_n)$ (e.g., Juang and Rabiner, 1985) to measure the distance of estimators from the true parameters.

$$K_n(\Theta_0, \Theta) \stackrel{def}{=} \frac{1}{n}\left\{\log \mathcal{L}(\mathbf{y}_1, ..., \mathbf{y}_n; \Theta_0) - \log\ \mathcal{L}(\mathbf{y}_1, ..., \mathbf{y}_n; \Theta)\right\}$$

For each value of n, the estimation procedure was carried out a hundred times, and the distances $\widetilde{K}_n$ between each of the hundred estimators and the true parameter were evaluated on a new sequence, independent of the first hundred sequences. As an illustration Table 2 gives the average and standard deviation of the estimates with $n = 600$. In this case the estimated transition matrix $\widetilde{\mathbf{P}}$ is given by

$$\widetilde{\mathbf{P}} = \begin{bmatrix} 0.9472 & 0.0521 & 0.0007 \\ (0.0037) & (0.0046) & (0.0017) \\ 0.0654 & 0.8918 & 0.0428 \\ (0.0026) & (0.0050) & (0.0021) \\ 0.0012 & 0.0441 & 0.9547 \\ (0.0015) & (0.0039) & (0.0024) \end{bmatrix}$$

Values into brackets represent standard deviation of the estimates. The results indicate that the estimation method works well. The sample means are very close to the

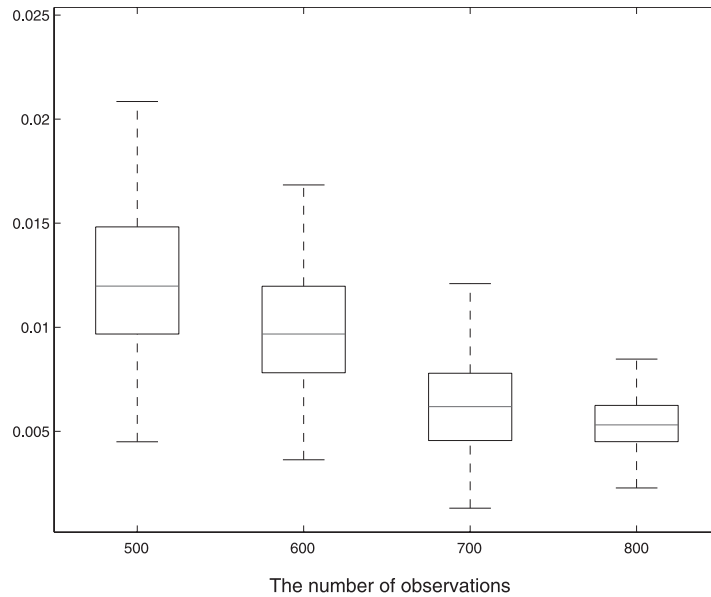

Figure 2 Box plots of $\widetilde{K}_n(\Theta_0, \widetilde{\Theta}_n)$.

true ones, and the standard deviations are small. Box plots of the sets of distances for the various values of n are presented under a unified scale in Figure 2. The plots clearly show a general decrease in average and spread of the distances with increasing n, which imply an increasing accuracy and stability of the estimators as n increases.

To investigate the asymptotic distribution of the estimates $\widetilde{\Theta}$, we have used the Shapiro-Francia (1972) statistic in order to test the univariate normality of each component of $\widetilde{\Theta}$. All the results show that, for $n \geq 400$, this test fails to reject the null hypothesis (the Θ_i are a random sample from $\mathcal{N}(\mu, \sigma)$, with μ and σ unknown) at the significance level $\alpha = 5\%$.

4.2. Model Selection

To find the number of latent factors and hidden states during a given time period, we use 2 selection criteria (AIC and BIC) in order to choose between $\mathcal{M}$ models. Thus for each selection criterion we, first, train various model configurations (obtained by varying the number of states and the number of factors from 1 to 3). Second, we use the output of EM to compute the values of the selection criterion for all configurations and we select the one that yields the lowest value. Using 1000 replications according to the true model (Table 1), the results reported in Table 3 show BIC to perform better than AIC. The AIC criterion generally fails in finding the exact numbers of factors and states. Whereas, the BIC criterion detect

Table 3 Model selection

Criterion	$k = 1$	$k = 2$	$k = 3$
		$m = 1$	
AIC	0	0	0
BIC	0	0	0
		$m = 2$	
	0	2	3.7
	0	2	2.1
		$m = 3$	
	1.3	23.8	**69.2**
	1.3	**94.6**	0

the exact numbers.

According to the literature, such result is not surprising. On one hand, despite the widespread use of the AIC, some believe that it is too liberal and tends to select overly complex models (Kass and Raftery, 1995). It has been pointed out that the AIC neglects the sampling variability of the estimated parameters. When the likelihood values for these parameters are not highly concentrated around their maximum value, this can lead to overly optimistic assessments. Furthermore, the AIC is not consistent. That is, as the number of observations grows very large, the probability that the AIC recovers a true low dimensional model does not approach unity. On the other hand, a comparison of BIC to AIC shows that the BIC penalty term is larger than the AIC penalty term when. The BIC assumes that the true generation model is in the set of candidate models, and it measures the degree of belief that a certain model is the true data-generating model. As we assume that the true model is in the candidate set and that it is relatively low dimensional, we favor BIC over AIC. Hence, only the BIC criterion will be used in finding the states and factors numbers.

The mean square error criterion given by $\widehat{e} = \frac{1}{n} \sum_{i=1}^{q} \sum_{t=1}^{n} \|y_{it} - \widehat{y}_{it}\|^2$ shows also that $k = 2$ and $m = 3$ is strongly favored (Fig. 3).

To illustrate the evolution of the model estimates obtained by the EM method, Figure 4 shows the HMM hidden states estimates at iteration 1, 3, 5 and 7. Each figure depicts the regime path process of the correct model. It can be concluded that a good segmentation is achieved after 7 iterations. Figure 5, reporting the estimated and simulated trajectories of the three hidden states, shows how the model is capable of accurately detecting all changes in the time series structure. Our results show also that the estimation errors based on the true model are not correlated (Fig. 6). Hence, all the correlation between the observed variables is fully explained by the common and specific factors.

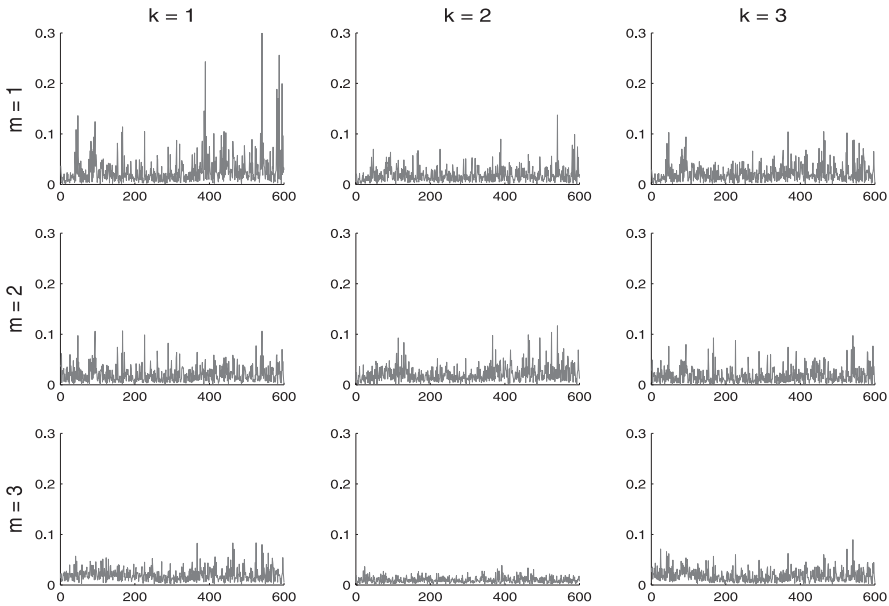

Figure 3 Computation of the estimation error for 9 different configurations.

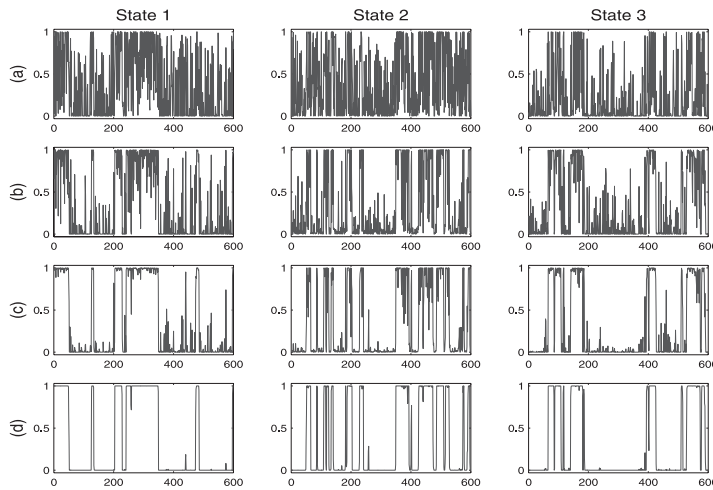

Figure 4 Evolution of the HMM state estimates using the true model: (a) iteration 1, (b) iteration 3, (c) iteration 5, (c) iteration 7.

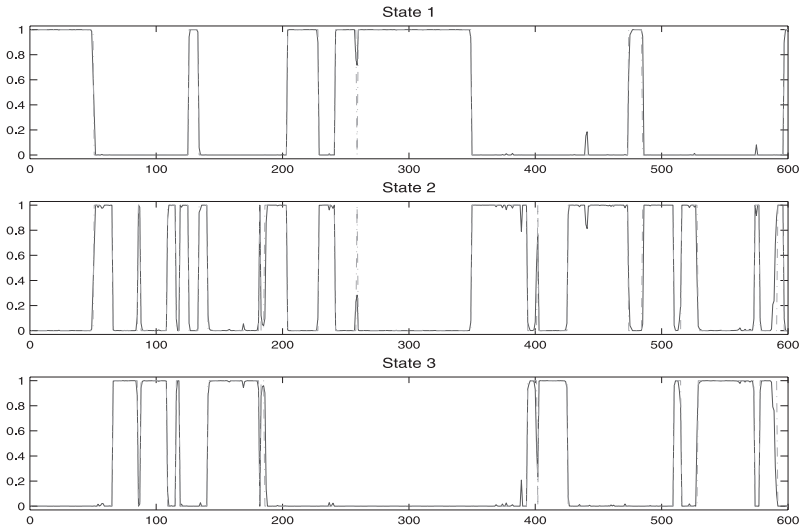

Figure 5 Posterior Probabilities of the three hidden states $\gamma_j(t)$ (solid line) and the simulated trajectories of the three hidden states (dashed line).

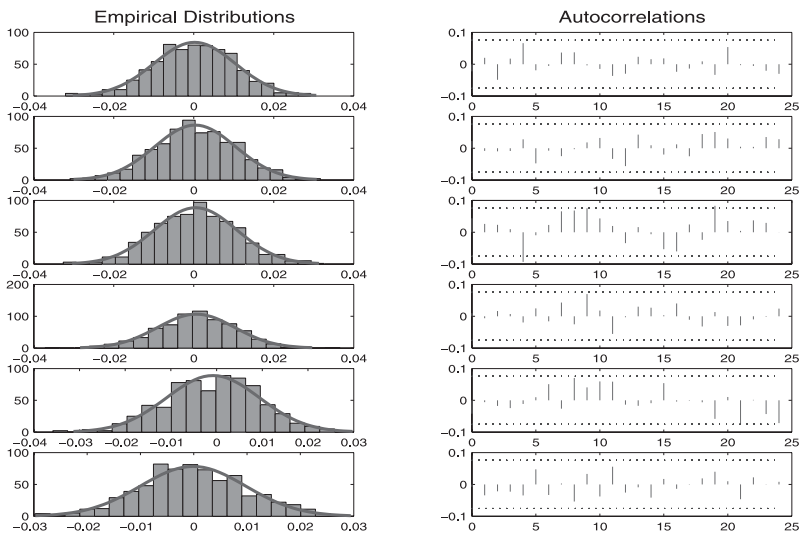

Figure 6 Empirical distributions of the estimation errors and their autocorrelation functions based on the true model.

Table 4 Estimation results for the chosen factor model with $m = 3$ and $k = 2$

	Currencies	θ	$\mathbf{X}$		$diag(\mathbf{\Psi})$
State 1	USD	0.0100	0.7709	0.0017	0.0011
	CAD	−0.0139	0.7279	0.0199	0.0805
	FRF	−0.0291	0.1451	0.5741	0.0162
	CHF	−0.0299	0.0464	0.5891	0.0805
	ITL	−0.0622	0.1259	0.4904	0.3072
	DEM	−0.0322	0.0841	0.5994	0.0008
	JPY	0.1050	0.4488	0.1644	0.3739
	HKD	0.0085	0.7660	−0.0012	0.0014
State 2		0.0004	1.0304	0.0132	0.0004
		−0.0156	0.9620	0.0195	0.0415
		0.0150	−0.0136	0.3990	0.0026
		0.0084	−0.0543	0.3902	0.0734
		0.0065	0.0048	0.3923	0.0127
		0.0163	−0.0589	0.4049	0.0041
		0.0165	0.5294	0.0054	0.1610
		0.0024	1.0225	0.0033	0.0032
State 3		0.0086	1.5156	−0.0086	0.0032
		−0.0104	1.5522	−0.0998	0.1475
		0.0113	0.3043	1.2107	0.0739
		−0.0295	0.4027	1.1271	0.3013
		−0.1367	0.2025	0.9753	1.4654
		0.0007	0.2829	1.2753	0.0037
		0.0327	1.3457	0.2620	0.2914
		−0.0020	1.5580	−0.0054	0.0278

4.3. An Illustrative Empirical Study

To illustrate the application of our method to learn and analyze co-movements amongst financial data, we have considered several exchange rate returns during the financial crisis that the European exchange markets has faced in fall 1992[4)]. The time series considered here are the daily returns of closing spot prices relative to the British pounds in price notations of the United States Dollar (USD), Canadian Dollar (CAD), French Franc (FRF), Swiss Franc (CHF), Italian Lira (ITL), German Marks (DEM), Japanese Yen (JPY), and Hong Kong Dollar (HKD) from 03/05/1991 to 07/05/1993 (600 observations). We have especially chosen this dataset to test the ability of this new specification to accurately reproduce observed patterns and to identify the major dimensions of change over the period from 1991 to 1993, including the impact of the crisis on the common and specific variances and the correlation structure of the different series. However, we can say that the empirical part of this paper is an illustration of our methods on real data, and not a very detailed empirical analysis studying the dynamic behavior of the exchange rates: our objective here is to investigate the reliability of the model and congruency

[4)] PACIFIC EXCHANGE RATE SERVICE, Sauder School of Business, http://fx.sauder.ubc.ca/.

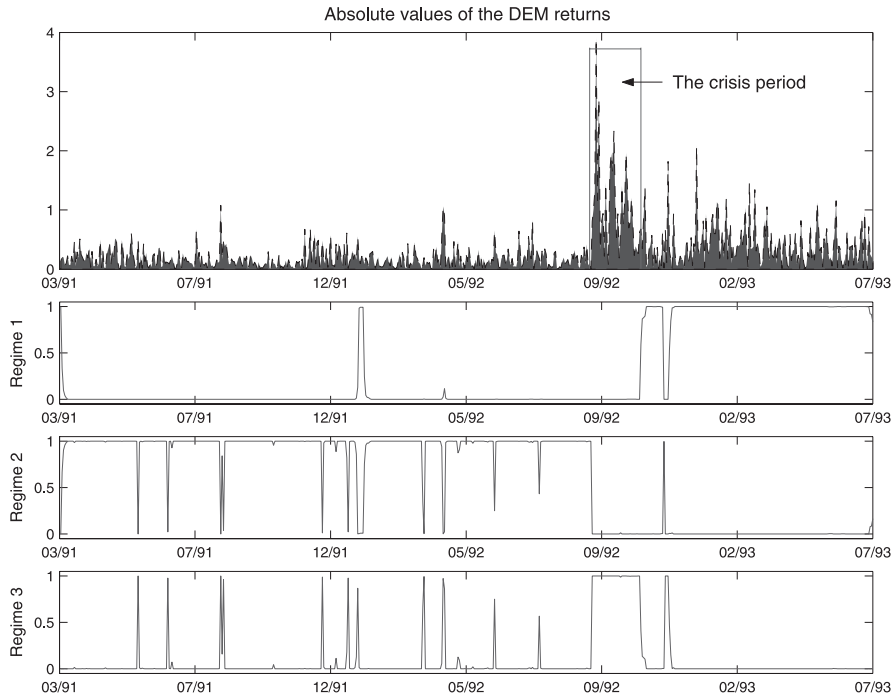

Figure 7 *Graphics 2, 3, 4*: Filtering probabilities $\gamma_j(t)$ given by the Forward-Backward algorithm.

of parameter estimates.

Our illustrative analysis first explores uncertainty about the number of factors and hidden states as in the foregoing simulated data analyses. With this intention, we have considered models with 1, 2 and 3 latent factors within a structure characterized by 1, 2 and 3 latent regimes. The BIC criterion argues that the covariance structure could be modeled by two common factors and three markovian regimes. The estimation results of this model are shown in Table 4.

Using this specification, Figure 7 shows how the model is capable of accurately detecting abrupt changes in the DEM time series structure and, in particular, the severe disruption by the violent storm which hit the European currency markets in September and October 1992. This segmentation shows that the third model is responsible for the high volatility segments, the second model is mainly responsible for the time period before August 1992, and the first one for the lower volatility segments after October 1992. This figure shows also that the average duration stay in the first regime is about 34.6 weeks versus 75 in the second and 10.4 in the third.

Some other interesting points arise from this analysis.

1. From Figure 8, It appears, that the second factor is responsible for time evolving movements in the variances of the European currencies, FRF, CHF, ITL

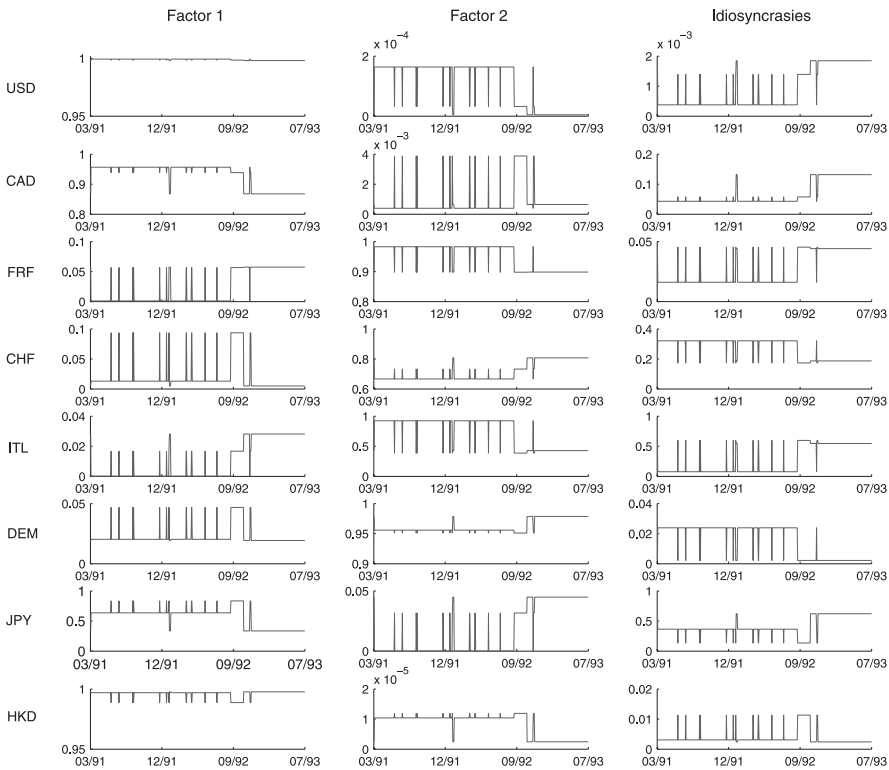

Figure 8 Two factor model: Proportion of the time series variances explained by each of the factors (common and specific), from 03/05/1991 to 07/05/1993.

and DEM. From this figure it can be observed that the first common factor explains at least 99% of the USD and HKD currency's variances at all times and at least 97% for the CAD before the crisis. This factor explains also 60% of the Japanese currency variances before August 1992. The contribution of the second factor in the variance of these currencies is negligible. It can also be observed that the second factor is responsible for about 98% of the variances of the FRF and DEM currencies and about 68% of the variance of the CHF currency before the 1992 crisis. For the CHF, the contribution of the second factor is about 80% after August 1992.

2. Figure 9 shows that all the correlations between the European currencies have increased just after August 1992. This is the effect of financial contagion that can be defined as a significant increase in co-movement of financial prices experienced by a group of countries, after controlling fundamentals (domestic news, ...) and common shocks, following a crisis elsewhere.

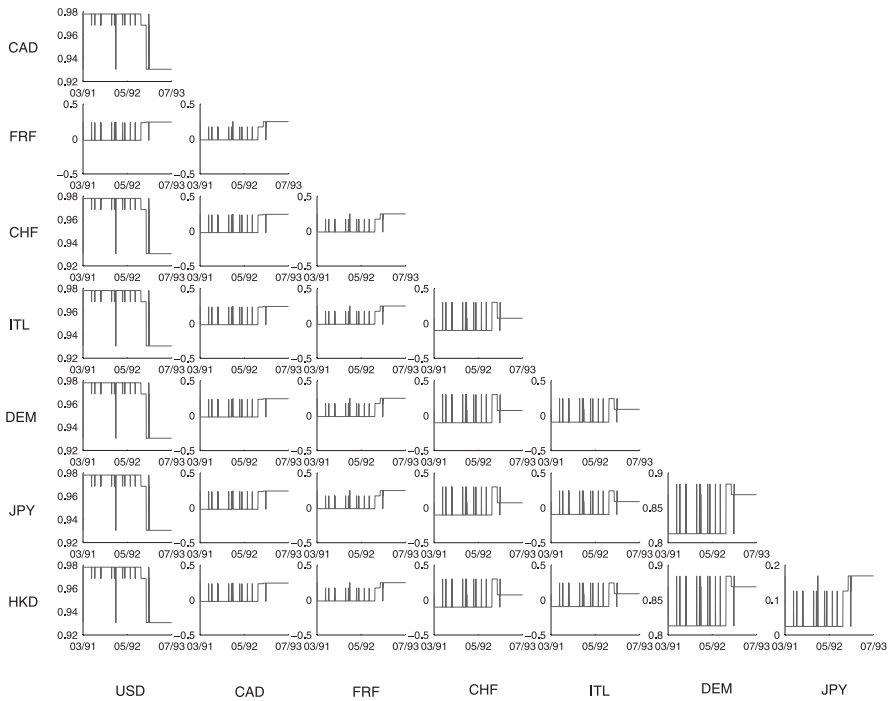

Figure 9 Two factor model: Time series co-dependence structure from 03/05/1991 to 07/05/1993.

More broadly, we can argue that the first factor represents the value of sterling relative to a basket of currencies in which the HKD, USD and CAD are dominant. Results presented in Table 4 show that the USD, CAD and HKD are roughly equally weighted, which is expected as CAD and HKD rates are heavily determined in international markets by USD rates. This first factor may be termed the North American factor. The second factor may be similarly termed the European Union factor. It represents a restricted basket of currencies dominated by the European currencies, with a relatively reduced weighting on JPY. USD, HKD and CAD are practically absent from this factor with $x_{1,2j}$, $x_{2,2j}$ and $x_{8,2j}$ for $j = 1, 2, 3$ indicating very small values. Inferences about idiosyncratic variances strengthen and extend these conclusions. Those of USD and DEM are very small, indicating that these two currencies play determining roles in defining their sector factor. FRF and ITL have larger idiosyncratic variances (during the crisis period), indicative of their departures from their sector factors.

To check the robustness of our estimated structural change dates to the base currency used to define the exchange rates, we re-estimated our switching model using the US dollar as the base currency. We use the posterior probabilities of the hidden states or the Viterbi algorithm to choose the number of structural breaks as

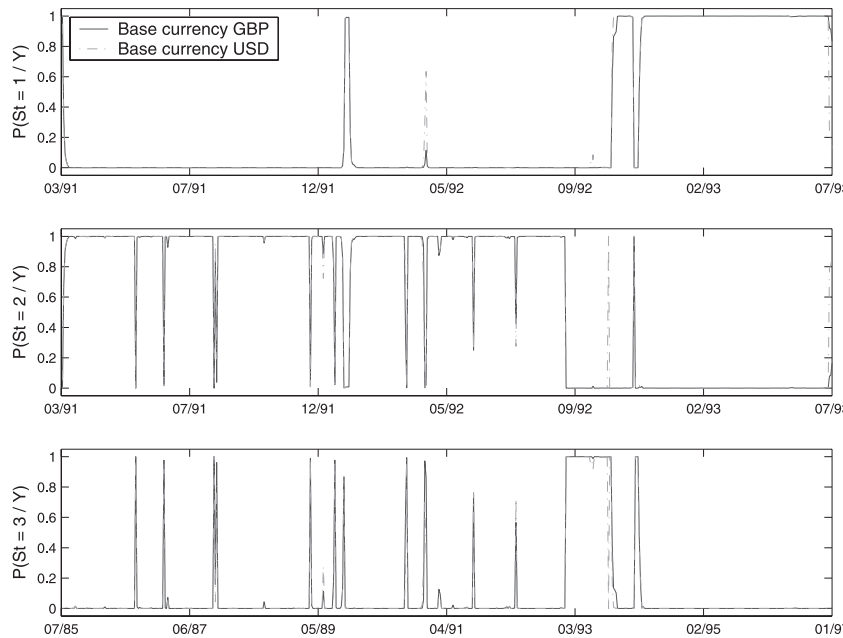

Figure 10 Posterior probabilities of the hidden Markov states $\gamma_j(t)$.

described in Sections 3.2 and 3.5, respectively. For all countries the break dates in the spot prices using the US dollar as the base currency are quite similar to the breaks found using the British Pound as the base currency (Fig. 10).

We find 33 break dates for all currencies, and, except the 269-*th*, 426-*th* and 598-*th* break dates, these breaks are very similar to those found using the British Pound as the base currency. However, as shown in Figure 11, the behavior of the spot prices in terms of the British pound is quite different than the spot prices in terms of US dollar, but the location of structural break dates are quite similar.

In our illustrative analysis, one critical aspect connected with the use of this model to analyze and forecast financial data is that the estimates dont take into account the heteroscedasticity and correlation in the data. An interesting direction for further research is to allow a dynamic structure for the conditional variances of the underlying factors in order to investigate possible time-varying latent processes, and their implications in modeling changes in covariance matrices over time. This is partly motivated by the fact that financial markets volatility changes over time (see Engle's (1982) work on Autoregressive Conditional Heteroscedasticity (ARCH) and Bollerslev's (1986) Generalized ARCH (GARCH)). Hence, in this situation, a switching Factor-ARCH framework can provide a plausible and parsimonious parametrization of the time varying variance-covariance structure of asset returns. For an exhaustive literature review on Factor-ARCH models with-

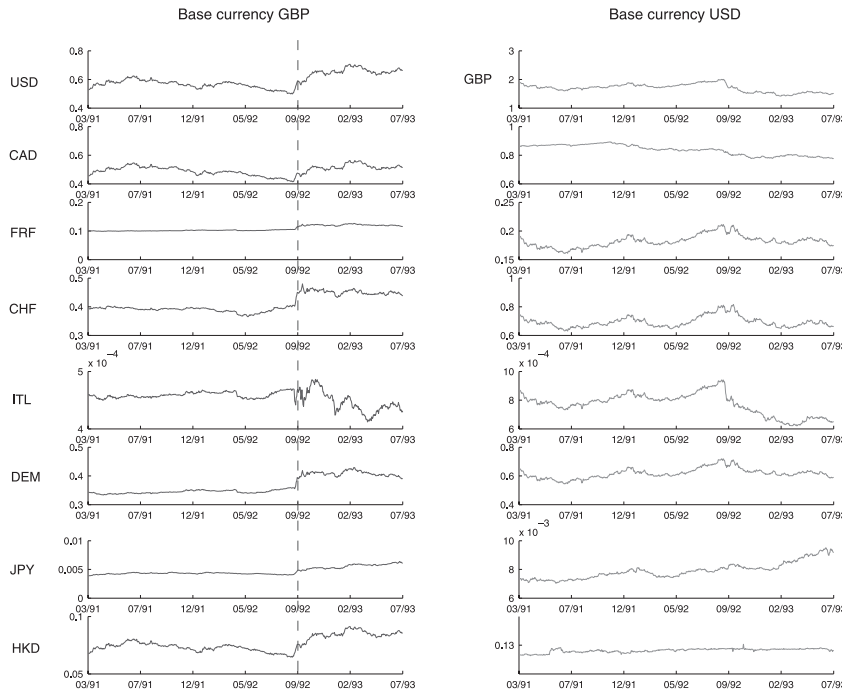

Figure 11 Real daily observed exchange rates from 03/05/1991 to 07/05/1993 (600 observations). The vertical line represents 31/08/1992.

out regime switching, the interested reader is referred to the works by Engle et al. (1990); Engle and Ng (1993); Diebold and Nerlove (1989); Engle and Susmel (1993) and Lin (1992). Engle et al. (1990) apply such structures to model the pricing of Treasury bills. A similar model is used by Engle and Ng (1993) to study the dynamic behavior of the term structure of interest rates. Diebold and Nerlove (1989) use a latent factor ARCH model to describe the dynamics of exchange rate volatility. Engle and Susmel (1993) use the factor ARCH to test for common volatility in international equity markets. Alternative estimation procedures for such models are, also, investigated by Lin (1992) on the basis of Monte Carlo comparisons.

5. Conclusion

In this paper we proposed a model that combines factor models and HMMs. We formulated the model and developed maximum likelihood estimates for its parameter. Our preliminary experiments have demonstrated promising results in classification of some latent behavior. Using two model selection criteria, we demonstrated

accurate discrimination between specifications characterized by different hidden structures. An interesting direction for further research is the generalization of this model to one where one allows the common and idiosyncratic variances to be time-varying. The study of such models would provide a further step in the extension of hidden Markov models to probabilistic factor analysis and allow for further flexibility in financial applications, where accurate on-line predictions of the time varying covariance matrices are very useful for dynamic asset allocation.

Acknowledgements

An earlier version of this paper was presented at the Third Nordic Econometric Meeting 2005 in Helsinki, Finland. We are grateful to the participants of the Congress for their helpful discussion. In addition, we would like to thank an anonymous referee for his extensive and constructive comments that greatly improved this paper. This joint work has been made possible by cross-invitations at INRIA Rhône-Alpes – IS2 and MISTIS teams – and Institut de Mathématiques et de Modélisation de Montpellier I3M which are also gratefully acknowledged.

References

Ang, A. and G. Bekaert (2002) "International asset allocation with regime shifts", *Review of Financial Studies* 15: 1137–1187.

Ang, A. and M. Piazzesi (2003) "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables", *Journal of Monetary Economics* 50: 745–787.

Bargmann, R. A. (1957) *A study of independence and dependence in multivariate normal analysis.* University of North Carolina, Institute of Statistics Mimeo Series N°186.

Bellman, R. (1957) *Dynamic programming.* Princeton University Press. NJ.

Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* 31: 307–327.

Cerra, V. and S. C. Saxena (2005) "Did output recover from the asian crisis?", *IMF Staff Papers* 52: 1–23.

Chauvet, M. and J. Hamilton (2005) "Dating business cycle turning points", in C. Milas, P. Rothman and D. van Dijk (eds.) *Nonlinear analysis of business cycles*. Elsevier (forthcoming).

Connor, G. and R. A. Korajczyk (1993) "A test for the number of factors in an approximate factor model", *Journal of Finance* 48: 1263–1291.

Connor, G. and R. A. Korajczyk (1995) "The arbitrage pricing theory and multifactor models of asset returns", in R. A. Jarrow, V. Maksimovic and W. T. Ziemba (eds.) *Handbooks in operations research and management science* 9. North-Holland, Amsterdam, pp. 87–144.

Dai, Q. and K. Singleton (1999) Specification analysis of affine term-structure models. Working Paper, Stanford University.

Dai, Q., K. Singleton and Y. Wei (2003) Regime shifts in a dynamic term structure model of U.S. treasury bonds. Working paper, Stanford University.

Davig, T. (2004) "Regime-switching debt and taxation", *Journal of Monetary Economics* 51: 837–859.

Dempster, A., N. Laird and D. B. Rubin (1977) "Maximum likelihood from incomplete data via the EM algorithm", *Journal of Royal Statistical Society Series B* 39: 1–38.

Diebold, F. and M. Nerlove (1989) "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model", *Journal of Applied Econometrics* 4: 1–21.

Engle, R. (1982) "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* 50: 987–1008.

Engle, R., V. K. Ng and M. Rothschild (1990) "Asset pricing with a factor-ARCH structure: empirical estimates for treasury bills", *Journal of Econometrics* 45: 213–237.

Engle, R. F. and R. Susmel (1993) "Common volatility in international equity markets", *Journal of Business and Economic Statistics* 11: 369–380.

Engle, R. and V. Ng (1993) "Time varying volatility and the dynamic behavior of the term structure", *Journal of Money Credit and Banking* 25: 336–349.

Fama, E. F. and K. R. French (1996) "Multifactor explanation of asset pricing anomalies", *Journal of Finance* 51: 55–84.

Forney, G. D. (1973) "The viterbi algorithm", *Proceedings of IEEE* 61: 268–278.

Gales, M. J. F. (1998) "Maximum likelihood linear transformations for HMM-based speech recognition", *Computer Speech and Language* 12: 75–98.

Garcia, R., R. Luger and E. Renault (2003) "Empirical assessment of an intertemporal option pricing model with latent variables", *Journal of Econometrics* 116: 49–83.

Geweke, J. F. and K. J. Singleton (1980) "Interpreting the likelihood ratio statistic in factor models when sample size is small", *Journal of the American Statistical Association* 75: 133–137.

Geweke, J. F. and G. Zhou (1996) "Measuring the pricing error of the arbitrage pricing theory", *The Review of Financial Studies* 9: 557–587.

Gopinath, R. A., B. Ramabhadran and S. Dharanipragada (1988) "Factor analysis invariant to linear transformations of data", in *Proc. Intl. Conf. on Spoken Language Processing.* ICSLP, pp. 397–400.

Gregory, A. W., A. C. Head and J. Raynauld (1997) "Measuring world business cycles", *International Economic Review* 38: 677–701.

Hamilton, J. (1988) "Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates", *Journal of Economic Dynamics and Control* 12: 385–423.

Hamilton, J. (1989) "A new approach to the economic analysis of nonstationary time series and the business cycle", *Econometrica* 57: 357–384.

Hamilton, J. (1990) "Analysis of time series subject to changes in regime", *Journal of Econometrics* 45: 39–70.

Hamilton, J. (2005) "What's real about the business cycle?", *Federal Reserve Bank of St. Louis Review*.

Jeanne, O. and P. Masson (2000) "Currency crises, sunspots, and Markov switching regimes", *Journal of International Economics* 50: 327–350.

Juang, B. H. and L. Rabiner (1985) "A probabilistic distance measure for hidden Markov models", *AT & T Tech. Journal* 64: 391–408.

Kass, R. E. and A. E. Raftery (1995) "Bayes factors", *Journal of the American Statistical Association* 90: 773–795.

Knez, P., R. Litterman and J. A. Scheinkman (1994) "Explorations into factors explaining money market returns", *Journal of Finance* 49: 1861–1882.

Lawley, D. N. (1940) "The estimation of factor loadings by the method of maximum likelihood", *Proceedings of the Royal Society of Edinburgh, Section A* 60: 64–82.

Lawley, D. N. (1942) "Further investigations in factor estimation", *Proceedings of the Royal Society of Edinburgh, Section A* 61: 176–185.

Lawley, D. N. (1943) "The application of the maximum likelihood method to factor analysis", *British Journal of Psychology* 33: 172–175.

Lawley, D. N. and A. E. Maxwell (1963) *Factor analysis as a statistical method.* Butterworths. London.

Lin, W. (1992) "Alternative estimators for factor GARCH models: a Monte Carlo comparaison", *Journal of Applied Econometrics* 7: 259–279.

Lintner, J. (1965) "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets", *Review of Economics and Statistics* 47: 13–37.

Press, S. J. (1985) *Applied multivariate analysis: using bayesian and frequentist methods of inference.* Krieger. California.

Rao, C. R. (1955) "Estimation and tests of significance in factor analysis", *Psychometrika* 20: 93–111.

Ross, S. A. (1976) "The arbitrage theory of capital asset pricing", *Journal of Economic Theory* 13: 341–360.

Rubin, D. B. and D. T. Thayer (1982) "EM algorithms for ML factor analysis", *Psychometrika* 47: 69–76.

Saul, L. and M. Rahim (2000) "Maximum likelihood and minimum classification error factor analysis for automatic speech recognition", *IEEE Transactions on Speech and Audio Processing* 8: 115–125.

Shapiro, S. S. and R. S. Francia (1972) "An approximate analysis of variance test for normality", *Journal of the American Statistical Association* 67: 215–216.

Sharpe, W. F. (1964) "Capital asset prices: A theory of market equilibrium under conditions of risk", *Journal of Finance* 19: 425–442.

Sims, C. and T. Zha (2004) Were there switches in U.S. monetary policy? Working paper, Princeton University.

Stock, J. and M. Watson (1993) "A simple estimator of cointegrating vectors in higher order integrated systems", *Econometrica* 61: 783–820.

Viterbi, A. J. (1967) "Error bounds for convolutional codes and an asymptotically optimal decoding algorithm", *IEEE Transactions on Information Processing* 13: 260–269.