Filtering and forecasting commodity futures prices under an HMM framework

Paresh Date[†]   **Rogemar Mamon**[‡,*]   **Anton Tenyakov**[‡]

[†]*Department of Mathematical Sciences*
Brunel University
London, United Kingdom

[‡]*Department of Statistical and Actuarial Sciences*
University of Western Ontario
London, Ontario, Canada

Abstract

We propose a model for the evolution of arbitrage-free futures prices under a regime-switching framework. The estimation of model parameters is carried out using the hidden Markov filtering algorithms. Comprehensive numerical experiments on real financial market data are provided to illustrate the effectiveness of our algorithm. In particular, the model is calibrated with data from heat oil futures and its forecasting performance as well as statistical validity is investigated. The proposed model is parsimonious, self-calibrating and can be very useful in predicting futures prices.

Keywords: Markov chain, change of measure, multivariate HMM filtering, oil future prices

---

[*]**Corresponding Author:** Mailing Address: Department of Statistical and Actuarial Sciences, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada N6A 5B7; E-mail: rmamon@stats.uwo.ca.

1 Introduction

In recent years, there have been various deregulations occurring in the electricity, oil and natural gas markets. Apparently, prices of these commodities reflect financial risks that are borne out by the market participants (sellers and buyers). Such risks are important considerations when proposing a model for the price evolution of these commodities especially in designing energy derivative contracts.

The modelling of commodity futures prices and their underlying variables was studied by various authors in the light of various financial modelling considerations and objectives. Cortazar, et al. [8] proposed a multicommodity model for futures prices that allows the use of long-maturity futures prices available for one commodity to estimate futures prices of another commodity; Kalman filtering was used in the model estimation. Nakajima and Ohashi [26] put forward a commodity pricing model that incorporates the effect of linear relations among commodity spot prices, and provided a condition under which such linear relations represent cointegration; using crude oil and heating oil market data, Kalman filtering was also utilised to estimate the model parameters. In Antonio et al. [1], the inclusion of jump components is carried out to explain the behaviour of oil prices; this, however, creates difficulties in the estimation of state variables, and so particle filters were applied instead of Kalman filters. Mirantes, et al. [25] formulated a generalised multi-factor model (n non-seasonal factors and m seasonal factors) for the stochastic behaviour of commodity prices, which nests the deterministic seasonal models; the seasonal factors are trigonometric components driven by random processes. A one-factor regime-switching model was developed in Chen and Forsyth [7] but the objective was to capture the risk-adjusted natural gas spot price dynamics; regression was used in model calibration using both market data on futures and options on futures. Back, et al. [2] conducted an extensive analysis covering samples of soybean, corn, heating oil and natural gas options, and provided evidence that seasonality in volatility is an important aspect to consider when valuing futures contracts; an appropriate seasonality adjustment significantly reduces pricing errors in these markets and yields more improvement in valuation accuracy than increasing the number of stochastic factors.

The contributions of this paper differ from the previous works mentioned above. We aim to address (i) the development of a model for the evolution of arbitrage-free futures prices suitable for valuation of commodity derivatives and (ii) provision of a regime-switching framework with HMM-based dynamic estimation for the modelling of multivariate commodity price time series along with the investigation of its various implementation issues.

2

Specifically, we propose an approach for the estimation of latent state variables in a model employed in futures pricing. The model is adapted from Manoliu and Tompaidis [24] under a framework that is consistent with no-arbitrage pricing. The methodology in [24] leads to a state-space formulation of the futures price model suited for Kalman filtering and maximum likelihood method. More specifically, the state variable for the spot and futures model is an Ornstein-Uhlenbeck process designed to capture mean-reversion and observed term structure of volatilities and correlation. Under this model, the futures prices are lognormally distributed. We start from a lognormal spot price process and derive a multivariate equation for futures prices. Instead of using a constant parameter (possibly multi-factor) mean-reverting process, we allow the model parameters to be modulated by a finite-state hidden Markov chain in discrete time. The parameters could then switch dynamically amongst economic regimes representing the interactions of various factors including mean-reversion and cyclical patterns (seasonality) in commodity prices.

The usual method of finding the maximum likelihood parameter estimates (MLEs) in conjunction with Kalman filtering is to numerically maximise the likelihood function. In Elliott and Hyndman [13], a filter-based implementation of the expectation maximisation (EM) algorithm that can be used to find the MLEs is presented. Such an approach makes use of the change of measure technique to evaluate filters under an ideal measure and relate the calculations back to the real-world through the Bayes' theorem. In recent years, linear and non-linear filtering have found a large number of applications in finance. A recent survey of developments in this area along with various implementation details in the context of financial modelling is featured in Date and Ponomareva [9].

Considering the multivariate nature of datasets for correlated futures prices, we utilise the filtering and estimation method for vector observations put forward in Erlwein, et al. [18]. The novelty of this work stems from the utilisation of all possible price information from the futures market to obtain model parameters. Our estimation procedures are designed to suitably calculate the $h-$step ahead forecasts of various related financial variables. We formulate a model that is compatible with the framework of Erlwein, et al. [18].

This paper is structured as follows. Section 2 presents the formulation of the model for the evolution of arbitrage-free futures prices. In section 3, the filtering algorithms for parameter estimation are outlined. We provide in section 4 a numerical implementation by applying the algorithms to a dataset of futures prices. We investigate the forecasting performance of our approach in predicting log returns and future prices. Finally, some concluding

remarks are given in section 5.

2 Arbitrage-free evolution of futures prices

In this section, we provide a brief outline of the development of an arbitrage-free model of futures price dynamics. Modelling the arbitrage-free price evolution is essential to appropriately price securities in the commodity markets such as spread options. Specification of an arbitrage-free model is necessary to be consistent with the risk-neutral approach in pricing. The development is based on reference [24] and the omitted proofs follow using a univariate version of Itô's lemma in a straightforward fashion.

We assume that the log-spot price ζ_t follows a single-factor mean-reverting process under the risk neutral measure (or $\mathbb{Q}$ measure, in conventional notation), i.e.,

$$d\zeta_t = (\alpha - \kappa\zeta_t)dt + \theta dW_t. \tag{1}$$

Here, α, κ and $\theta > 0$ are assumed constants and W_t is a $\mathbb{Q}$-Wiener process. The spot price is considered to be a latent state, i.e., unobservable. We assume further that at each time prices of m futures are available with maturities $T_1, T_2, \cdots, T_m$. The price of the futures contract with maturity T_i, at time $t < \min_i(T_i)$, is denoted by $F^i(t)$ and can be written as

$$F^i(t) = \mathbb{E}_{\mathbb{Q}}(e^{\zeta^i}|\mathcal{F}_t) = \exp\left(\mathbb{E}_{\mathbb{Q}}(\zeta^i|\mathcal{F}_t) + \frac{1}{2}\operatorname{Var}_{\mathbb{Q}}(\zeta^i|\mathcal{F}_t)\right),$$

where we denote ζ_{T_i} by ζ^i for notational brevity and use the fact that e^{ζ^i} is log-normal. Here, $\{\mathcal{F}_t\}$ is the filtration generated by W_t. This leads to a closed-form expression for $F^i(t)$ given by

$$F^i(t) = \exp\left(e^{-\kappa(T_i-t)}\zeta_t + \frac{\alpha}{\kappa}(1 - e^{-\kappa(T_i-t)}) + \frac{\theta^2}{4\kappa}\left(1 - e^{-2\kappa(T_i-t)}\right)\right). \tag{2}$$

Our modelling formulation is consistent with the log-spot price modelling assumptions in Manoliu and Tompaidis's paper [24]. Under the assumption of no-arbitrage valuation, the benefits from holding the physical asset, called convenience yield, are reflected in the futures price (cf. Hull [22]). Convenience yield in a commodity is analogous to a dividend in an asset that provides a known income; and dividends would naturally lead to a corresponding adjustment in the price of the underlying asset of a futures contract. Whilst we did not explicitly model the dynamics of the convenience yield as it is not the intent of this paper to quantify it, we see that it is implicitly taken into account through the log-spot price ζ in

the closed-form expression for the futures price in equation (2).

Furthermore, at a fixed time t, $F^i(t)$ can be an increasing or a decreasing function of maturity T_i, depending on the choice of parameters, which can easily be seen from equation (2). Futures prices decreasing (respectively, increasing) with maturity reflects backwardation (respectively, contango). Typically, being able to model both these situations adequately is a reason for modelling net convenience yield explicitly (possibly as a stochastic process). We can achieve a switch between the two situations in our framework through updating model parameters via self calibration as well as through regime switching, as will be made clear in the subsequent sections.

Next, we assume that the log-spot price follows a mean-reverting process under the objective measure (or $\mathbb{P}$ measure) given by

$$d\zeta_t = (\widetilde{\alpha} - \kappa\zeta_t)dt + \theta d\widetilde{W}_t, \tag{3}$$

where $\widetilde{W}_t$ is a $\mathbb{P}$-Wiener process. Note that the arbitrage-free-market assumption is equivalent to the existence of a $\mathbb{Q}$ measure which is equivalent to the $\mathbb{P}$ measure, such that the futures price processes are martingales under $\mathbb{Q}$ measure. Further, this assumption is also equivalent to the existence of a price of risk process λ_t such that $\tilde{\alpha} - \alpha = \lambda_t\theta$ holds. See Elliot, et al. [15] for the construction of a price of risk process λ_t for a regime-switching linear Gaussian model, using double Esscher transform, for both continuous-time and discrete-time cases. In the present situation, we will use a discretised version of the above process with a discrete-time, finite-state Markov chain governing the regime switching; thus, the discussion in [15] remains applicable in our context. For the time being, we assume $\lambda_t =: \lambda$ to be a constant. We apply Itô's lemma to $F^i(t)$, using equations (2) and (3), which leads to the following arbitrage-free dynamics for log-futures price:

$$d\left(\log F^i(t)\right) = e^{-\kappa(T_i-t)}\left(\lambda\theta - \frac{\theta^2}{2}e^{-\kappa(T_i-t)}\right)dt + \theta e^{-\kappa(T_i-t)}d\tilde{W}_t. \tag{4}$$

In the subsequent discussion, we shall assume that the parameters λ and θ are dependent on the current regime and regime-switching is allowed via a finite state hidden Markov chain. However, the discussion on regime switching is postponed to section 3 and we will assume the parameters to be constant for the purpose of this section.

For calibration and forecasting purposes of a multivariate time series of futures prices, we use a moment-matching procedure to implement equation (4) in discrete time. We suppose

that observation times $t_1 \leq t_2 \leq \ldots \leq t_n$ are equally spaced and $t_{k+1} - t_k =: \Delta$. To write the dynamics of a vector of futures prices in a compact form, let $\text{vec}\{a^i\}$ denote a vector with a^i at its ith element. Then, at time k, the arbitrage-free evolution of the futures price vector of log-returns is given by

$$\text{vec}\left\{y_k^i\right\} = \text{vec}\left\{f_k^i\right\} + \text{vec}\{q_k^i\}z_k, \tag{5}$$

where

$$f_k^i := \frac{\lambda\theta}{\kappa}e^{-\kappa(T_i - t_k)}\left(1 - e^{-\kappa\Delta}\right) - \frac{\theta^2}{4\kappa}e^{-2\kappa(T_i - t_k)}\left(1 - e^{-2\kappa\Delta}\right), \tag{6}$$

$$y_k^i := \log\frac{F^i(t_k)}{F^i(t_{k-1})}, \tag{7}$$

$$q_k^i = \theta e^{-\kappa(T_i - t_k)}\sqrt{\frac{1 - e^{-2\kappa\Delta}}{2\kappa}} \tag{8}$$

and $\{z_k\}$ is a sequence of independent Gaussian random variables with zero mean and unit variance. The above discrete-time implementation preserves the exact distribution of $\log F^i(t_{k+1})$, conditional on $\mathcal{F}_k$ (*i.e.*, information up to time t_k). Hence, it is preferred over more conventional Euler discretisation.

Although our formulation assumptions are similar to those in Manoliu and Tompaidis in [24], our approach differs significantly. The work in [24] relies on modelling the latent spot price evolution explicitly and then using Kalman filtering methodology to extract this latent price. Single as well as multi-factor models are used in [24] and a non-parametric seasonality adjustment is suggested for forecasting. In our case, the spot price does not feature in the futures equation and is modelled only implicitly. We incorporate regime switching to allow for factors such as seasonality and use a self-calibrating filter which is adapted from Erlwein et al. [18] for our particular model structure. It will be demonstrated that a single-factor, two-regime model gives satisfactory results for a chosen dataset.

Our aim is to obtain estimates for the parameters λ, κ and θ along with the transition probabilities if any of these parameters are governed by a discrete-time finite-state Markov chain. We solve this problem in two steps:

1. Initial parameter estimation: In this step, we assume that the transition probability matrix is identity over the observed time series and identify the initial parameter estimates for λ, κ and θ by maximising the likelihood of the observed time series. Let these initial estimates be denoted by λ_0, κ_0 and θ_0, respectively. The details of this step are given in section 3.1.

2. Update of parameters and transition probabilities using a self-calibrating filter: To make the implementation tractable, we assign an appropriate fixed value for κ and hence, κ is assumed to be independent of the Markov chain. The estimates of λ and θ are updated using a self-calibrating filter. The implementation steps for this filter are derived in section 3.2.

3 Filtering and model parameter estimation

3.1 Initial estimates of parameters

To find the initial estimates of parameters, we assume that the system is operating in a single regime, i.e., the transition probability matrix for the Markov chain is identity. Suppose that data on futures prices is available for times $k = 1, 2, \ldots, n$, from which a vector-valued time series of log returns, $\mathrm{vec}(y_1^i), \mathrm{vec}(y_2^i), \ldots \mathrm{vec}(y_n^i)$ can be constructed for $i = 1, 2, \ldots, m$.

In equation (5) the components f_k^i and q_k^i are parametrised by θ, κ and λ. Since we are assuming that z_k are IID standard normal random variables, the likelihood function of y_k^i is given by

$$L\left(y_k^i; \lambda, \kappa, \theta\right) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi} q_k^i} \exp\left(-\frac{(h_k^i)^2}{2(q_k^i)^2}\right),$$

where

$$h_k^i := y_k^i - f_k^i. \tag{9}$$

For each futures maturity T_i, we can therefore find estimates of the parameters λ, κ and θ by maximising the likelihood of observations, *i.e.*, by solving the non-convex optimisation problem

$$\min_{\lambda, \kappa, \theta > 0} \sum_{k=1}^{n} \left(\log q_k^i + \frac{(h_k^i)^2}{2(q_k^i)^2}\right). \tag{10}$$

To find a common set of parameters λ, κ and θ which maximise the likelihood of observations for all futures with all available maturities $(T_1, T_2, \cdots, T_m)$ *simultaneously*, one may follow multi-objective optimisation approach and seek a set of parameters such that likelihood for any T_i cannot be improved upon without decreasing the likelihood of observations for a different T_i. As we are going to update the parameters using a self-calibrating filter later, we reject this numerically involved approach. Instead, a simpler approach is sought to obtain a set of parameters which minimise the sum of negative log likelihood of observations for all

futures, *i.e.* we solve the following optimisation problem:

$$\min_{\lambda,\kappa,\theta>0} \sum_{i=1}^{m} \sum_{k=1}^{n} \left(\log q_k^i + \frac{(h_k^i)^2}{2(q_k^i)^2} \right). \tag{11}$$

To obtain initial values, let $\{\lambda_0, \ \kappa_0, \theta_0\}$ be any set of locally minimising arguments. In the succeeding dynamic estimation, *i.e.*, updating, we shall keep κ fixed and assume that the remaining two parameters λ and θ depend on a finite-state Markov chain. The estimation and update of the transition probabilities of this Markov chain and values of other parameters corresponding to different states are discussed in the next section.

3.2 Derivation of self-calibrating filter

We reformulate the problem in the standard form used in the literature on regime-switching models; see, for example, Buffington and Elliott [3], Elliott, et al. ([11], [12] and [16]), Erlwein, et al. ([17] and [18]), amongst others. Recall that the ith component of the vector in equation (5) can be written as

$$y_{k+1}^i = f_k^i + q_k^i z_{k+1}. \tag{12}$$

Let $\mathbf{x}_k$ be a finite-state homogeneous Markov chain in discrete time, i.e., $k = 0, 1, 2, \dots$. The semi-martingale representation of $\mathbf{x}_k$ is given by

$$\mathbf{x}_{k+1} = \mathbf{\Pi}\mathbf{x}_k + \boldsymbol{\epsilon}_{k+1}, \tag{13}$$

where $\mathbf{\Pi}$ is the transition probability matrix and $\boldsymbol{\epsilon}_{k+1}$ is martingale increment. Our observation process is $m-$dimensional (one price observation for each maturity) and the component i follows the dynamics given in (12). As mentioned earlier, z_k are IID random variables which are also independent from the Markov chain $\mathbf{x}_k$ driving the regime-switching dynamics of the mean and volatility parameters for each observation component.

To simplify considerably the algebra involved in the filtering equations, we associate the state space of $\mathbf{x}_k$ with the canonical basis of $I\!\!R^N$, which is the set of unit vectors $\mathbf{e}_r$, $r = 1, 2, \dots, N$ and $\mathbf{e}_r$ is a vector having 1 in its rth entry and 0 elsewhere. So in equation (12), $f^i(\mathbf{x}_k) = \langle \mathbf{f}_k^i, \mathbf{x}_k \rangle$ and $q^i(\mathbf{x}_k) = \langle \mathbf{q}_k^i, \mathbf{x}_k \rangle$, where $\mathbf{f}_k^i = (f_k^i(1), f_k^i(2), \dots, f_k^i(N))^\top \in I\!\!R^N$ and $\mathbf{q}_k^i = (q_k^i(1), q_k^i(2), \dots, q_k^i(N))^\top \in I\!\!R^N$. The notation $\langle \cdot, \cdot \rangle$ is the usual scalar product and $\top$ denotes the transpose of a vector.

To obtain the optimal estimates of $\mathbf{x}_k$ using the observation process, we employ the change

of reference probability technique. In this technique, we perform the calculations under $\widetilde{P}$ measure under which the observations y_k's are $N(0,1)$ IID sequence of random variables, and y_k is independent from $\mathbf{x}_k$. Under the reference probability the Markov chain dynamics are unchanged. All components of the $m-$dimensional observation process have the same underlying Markov chain.

The change of reference probability in our framework utilises a discrete-time version of Girsanov's theorem. The real-world measure P, under which we observe our measurements, can be recovered from $\widetilde{P}$ through the construction of the Radon-Nikodym derivative

$$\Lambda_k := \left.\frac{dP}{d\widetilde{P}}\right|_{\mathcal{F}_k^y} = \prod_{i=1}^{m}\prod_{l=1}^{k} \lambda_l^i, \ k \geq 1,$$

where

$$\Lambda_0 = 1 \ \text{ and } \ \lambda_l^i = \frac{\phi[q^i(\mathbf{x}_{l-1})^{-1}(y_l^i - f^i(\mathbf{x}_{l-1}))]}{q^i(\mathbf{x}_{l-1})\phi(y_l^i)}$$

with ϕ being the $N(0,1)$ density. All parameters are dependent on the same Markov chain and react to the same underlying price information. In some sense, the components are correlated through the Markov chain. Whilst the noise terms for the individual components are uncorrelated, the correlation structure of the futures prices are encapsulated in each noise. This simplification is made to make the model tractable.

We present the filter equations under a multivariate setting. Our goal is to provide adaptive filters for the estimates of the states and other auxiliary processes related to the Markov chain. Let $\mathcal{F}_k$ be the filtration generated by the log-returns process y_k. To find the conditional distribution of $\mathbf{x}_k$ given $\mathcal{F}_k$ under P, we write

$$\widehat{p}_k(r) := P(\mathbf{x}_k = \mathbf{e}_r | \mathcal{F}_k^y) = E[\langle \mathbf{x}_k, \mathbf{e}_r \rangle | \mathcal{F}_k^y]$$

and $\widehat{\mathbf{p}}_k = (\widehat{p}_k(1), \widehat{p}_k(2), \ldots, \widehat{p}_k(N))^\top \in I\!\!R^N$. Now,

$$\widehat{\mathbf{p}}_k = E[\mathbf{x}_k | \mathcal{F}_k^y] = \frac{\widetilde{E}[\Lambda_k \mathbf{x}_k | \mathcal{F}_k^y]}{\widetilde{E}[\Lambda_k | \mathcal{F}_k^y]}$$

by Bayes' theorem for conditional expectation. Let $\mathbf{c}_k = \widetilde{E}[\Lambda_k \mathbf{x}_k | \mathcal{F}_k^y]$ and note that $\sum_{r=1}^{N} \langle \mathbf{x}_k, \mathbf{e}_r \rangle = 1$. Thus,

$$\sum_{r=1}^{N} \langle \mathbf{c}_k, \mathbf{e}_r \rangle = \sum_{r=1}^{N} \left\langle \widetilde{E}[\Lambda_k \mathbf{x}_k | \mathcal{F}_k^y], \mathbf{e}_r \right\rangle = \widetilde{E}\left[\Lambda_k \sum_{r=1}^{N} \langle \mathbf{x}_k, \mathbf{e}_r \rangle \middle| \mathcal{F}_k^y\right] = \widetilde{E}[\Lambda_k | \mathcal{F}_k^y]. \quad (14)$$

The construction of $\mathbf{c}_k$ along with equation (14) yields

$$\widehat{\mathbf{p}}_k = \frac{\mathbf{c}_k}{\sum_{r=1}^{N} \langle \mathbf{c}_k, \mathbf{e}_r \rangle}.$$

We denote the conditional expectation under $\widetilde{P}$ of $\Lambda_k G_k$ by $\gamma_k(G_k) := \widetilde{E}[\Lambda_k G_k | \mathcal{F}_k^y]$. The adaptive filters will enable the model parameters to adjust to current market conditions. We give the recursive filters for: (i) $(J^{sr}\mathbf{x})_k$, the process related to the Markov chain's jumps up to time k; (ii) $(O^r\mathbf{x})_k$, the process related to Markov chain's occupation time; and (iii) $(T^r(g)\mathbf{x})_k$, an auxiliary process related to $\mathbf{x}$ and for some function g.

The results for the recursive filters, which are modifications of those given in Erlwein, et al. [18] when there is only one uniform source of noise for each component of the observation vector, are presented in the Appendix. The Expectation-Maximisation (EM) algorithm is applied to calculate the optimal estimates of the model parameters. Such calculations result in expressions that involve the use of adaptive filters related to the Markov chain process provided in Proposition 1. Given the recursive filters in equations (16), (17), (18) and (19), the model parameters are updated every time new information arrives. Proposition 2 in the Appendix, whose proof is described in Erlwein, et al. [18], gives the optimal parameter estimates computed using the EM algorithm in terms of the filters.

Given the estimates for $q_k^i(x_k)$ via a self-calibrating filter, θ is found approximately using equation (8) for a particular regime. Then by substituting the given estimates of $f_k^i(x_k)$, θ and a fixed κ into equation (6), one can obtain an estimate of λ for a particular regime. This procedure ensures that the parameter constraints are satisfied, at least approximately, across all maturities.

4   Numerical implementation

We illustrate our method by applying it to the dataset of daily log-returns series of heating oil future contracts compiled by Data Stream. The data were recorded from 19 June 2009 to 28 October 2010 with ten maturity dates. These maturity dates, denoted by T_i, $i = 1, \ldots, 10$, are the last trading days in the months of October 2010, November 2010, ..., and July 2011. In the last part of this section, we perform a one-step ahead prediction using the data collected after 28 October 2010 until the expiry of the longest maturity contract on 29 July 2011, keeping κ fixed and allowing the other parameters to be updated using

	T=29/07/11	T=30/06/11	T= 31/05/11	T=29/04/11	T=31/03/11
Minimum	-0.0369	-0.0372	-0.0374	-0.0372	-0.0369
Maximum	0.0474	0.0478	0.0481	0.04812	0.0480
Median	0.0001	0.0001	0.0001	0.0000	0.0000
Mode	0.0000	0.0000	0.0000	0.0000	0.0000
Mean	0.0003	0.0003	0.0003	0.0003	0.0003
Std Dev	0.0146	0.0147	0.0149	0.0150	0.0150
Skewness	-0.0360	-0.0342	-0.0312	-0.0252	-0.0219
Kurtosis	0.0054	-0.0023	-0.0116	-0.0190	-0.0329
	T=28/02/11	T=31/01/11	T=31/12/10	T=30/11/10	T=29/10/10
Minimum	-0.0366	-0.0369	-0.0372	-0.0379	-0.0384
Maximum	0.0477	0.0479	0.0481	0.0486	0.0490
Median	0.0000	0.0000	0.0000	0.0000	0.0000
Mode	0.000	0.000	0.0000	0.0000	0.0000
Mean	0.0003	0.0003	0.0002	0.0002	0.0003
Std Dev	0.0151	0.0152	0.0154	0.0157	0.0160
Skewness	-0.0238	-0.0254	-0.0242	-0.0240	-0.0240
Kurtosis	-0.0479	-0.0579	-0.0745	-0.0899	-0.1089

Table 1: Descriptive statistics for the log-returns of futures price for the entire dataset

the adaptive filters. Note that this allows us to test as well the performance of our filters for contracts with very short-term maturity. This means that for this specific experiment, the observation vector contains 9 entries till the last trading day of November 2010, then 8 entries till the last trading day of December 2010, and so on until only one entry in July 2011.

Table 1 show the descriptive statistics of the entire futures prices data given their maturities. In Tables 2 and 3, the descriptive statistics for the log-returns of futures price for the respective periods of 29/06/2009–14/07/2009 and 15/07/2009–24/07/2009 are shown. The suitability of the regime-switching model is clearly demonstrated by Tables 2 and 3, where the sample moments are statistically different for the two non-overlapping periods.

Indeed, the log-returns of our multivariate data undergo regime changes in mean and volatility levels. To model such regime-switching behaviour, we assume that for every maturity

11

	T=29/07/11	T=30/06/11	T= 31/05/11	T=29/04/11	T=31/03/11
Minimum	-0.0268	-0.0273	-0.0278	-0.0283	-0.02870
Maximum	0.0070	0.0071	0.0072	0.0072	0.0069
Median	-0.0086	-0.0087	-0.0088	-0.0088	-0.0092
Mean	-0.0104	-0.0106	-0.0107	-0.0108	-0.0110
Std Dev	0.0106	0.0108	0.0109	0.0110	0.0111
Skewness	-0.0353	-0.0481	-0.0518	-0.0694	-0.0960
Kurtosis	-0.8898	-0.8852	-0.8639	-0.8379	-0.8404
	T=28/02/11	T=31/01/11	T=31/12/10	T=30/11/10	T=29/10/10
Minimum	-0.0292	-0.0297	-0.0304	-0.0314	-0.0324
Maximum	0.0066	0.0064	0.0062	0.0057	0.0055
Median	-0.0096	-0.0101	-0.0104	-0.0108	-0.0109
Mean	-0.0112	-0.0114	-0.0116	-0.0119	-0.0122
Std Dev	0.0111	0.0113	0.0114	0.0116	0.0119
Skewness	-0.1342	-0.1574	-0.1832	-0.2347	-0.2802
Kurtosis	-0.8255	-0.8300	-0.8142	-0.7892	-0.7674

Table 2: Descriptive statistics for the log-returns of futures price for the period 29/06/2009–14/07/2009

	T=29/07/11	T=30/06/11	T= 31/05/11	T=29/04/11	T=31/03/11
Minimum	0.0009	0.0009	0.0009	0.0009	0.0009
Maximum	0.0265	0.0268	0.0270	0.0271	0.0271
Median	0.0114	0.0115	0.0116	0.0117	0.0117
Mean	0.0128	0.0130	0.0132	0.0132	0.0132
Std Dev	0.0091	0.0091	0.0093	0.0094	0.0093
Skewness	0.0293	0.0352	0.0396	0.0390	0.0389
Kurtosis	-1.3725	-1.3875	-1.4187	-1.4349	-1.4351
	T=28/02/11	T=31/01/11	T=31/12/10	T=30/11/10	T=29/10/10
Minimum	0.0009	0.0009	0.0009	0.0009	0.0009
Maximum	0.0270	0.0270	0.0272	0.0275	0.0276
Median	0.0116	0.0116	0.0117	0.0118	0.0120
Mean	0.0132	0.0132	0.0133	0.0135	0.0136
Std Dev	0.0093	0.0093	0.0094	0.0095	0.0096
Skewness	0.0385	0.0387	0.0291	0.0289	0.0226
Kurtosis	-1.4360	-1.4355	-1.4264	-1.4420	-1.4953

Table 3: Descriptive statistics for the log-returns of futures price for the period 15/07/2009–24/07/2009

date, log-return's mean f and volatility q, the daily price process $F^i(t_k)$ is of the form

$$y_{k+1}^i = \ln \frac{F^i(t_k)}{F^i(t_{k-1})} = f^i(\boldsymbol{x}_k) + q^i(\boldsymbol{x}_k)z_{k+1}$$

corresponding to each maturity date T_i.

4.1   Computing initial parameter estimates

As mentioned in subsection 3.1, we first assume that the system operates under a one-regime setting. This allows us to find starting values λ_0, κ_0 and θ_0. To simplify the implementation, we assume that κ is constant. Thus, using $\kappa(t) = \kappa_0$ for all time $t > 0$, the evolution of the process λ_t is derived under the multi-regime modelling set-up. This would in turn update θ_t given an estimate of f^i as shown in equation (6).

Whilst we made the assumption that κ is constant to achieve simplicity in the implementation, such assumption is actually justified empirically. That is, κ appears not to depend on a Markov chain and remains constant through time for any quantity of regimes. We found that if the size of the data used to estimate the initial parameters is varied, both λ_0 and θ_0 change but κ appears stable. As it is computationally intensive to calculate starting values of λ, κ, θ for various combinations of window sizes for our dataset, we randomly select a few sub-dataset samples from the original data and then use these samples to estimate initial values. The processing of data via a moving lag window is discussed in subsection 4.2. Our numerical results show that there is not much perturbation in the estimated κ values. This indeed supports the assumption that κ can be taken as constant; see Table 4.

Following the calculation of the initial parameters of the model outlined in section 3.1, we first solve the minimisation problem in (11) using a standard function fminsearch in MATLAB. The built-in algorithm in fminsearch is quite fast, but it provides arguments which minimise the function in (11) only locally. Since this function exhibits fluctuating behaviour, we have to deal with many local extreme points. This implies that one has to search for a global minimum at least within a reasonably big subset of $I\!\!R^3$. We observed that the function fminsearch always finds an extreme point which is closest to the initial value supplied. A subset of $I\!\!R^3$ for the space of the initial values is therefore considered and a minimum point in that space is determined. The results of our estimations for five sub-dataset samples are displayed in Table 4 and again, it is clear that the assumption of a constant κ is reasonable. In our HMM filtering implementation, we set $\kappa = 0.00057$, being the average of the calculated initial values for κ_0. It has to be noted that the parameter

λ_0	κ_0	θ_0
0.0618	5.832×10^{-4}	0.0140
0.0166	5.458×10^{-4}	0.0161
0.0528	5.411×10^{-4}	0.0153
0.0248	5.594×10^{-4}	0.0149
0.0343	5.693×10^{-4}	0.0165

Table 4: Estimation of initial values of κ, λ and θ for five sub-dataset samples

θ does not vary a lot either. This fact provides additional support for the validity of the model and accuracy of the filtering algorithms. As one will see in section 4.2, the values of q_k^i do not change significantly after only several iterations. This is consistent with equation (8) since an estimate of q_k^i gives an update of θ.

4.2   Implementation of self-calibrating filters

The HMM filtering algorithms were implemented with a moving lag window of various time steps to obtain estimated values for f^i and q^i. That is, we experimented to process data points in batches of 1-5 data points per batch for each algorithm step. In particular, we apply the recursive filtering equations in Proposition 1 in processing a batch of data points. Consequently, this gives estimates of the filters for various quantities related to the Markov chain that are used to provide EM estimates for model parameters in accordance with Proposition 2. The two-step process of calculating the filters and computing EM estimates constitutes the completion of one algorithm step. The two-step process is then repeated for the next moving lag window of data points. The final filtered values in the previous algorithm step are employed as initial values for the filtering equations in the succeeding algorithm step.

Each moving lag window for our filtering is assessed on the basis of goodness-of-fit metric, which is the root mean square error (RMSE). The model corresponding to a given number of regimes and moving lag window with the lowest RMSE is deemed as the most appropriate for the dataset. The results of our RMSE computations are summarised in Tables 5 and 6 for the one-regime and two-regime settings, respectively. The starting values of q are uniform for all regimes as these do not affect the convergence of the filtering algorithms.

Given any number of regimes, a pattern emerges from Tables 5-6 in that the best RMSE value is for a window of size four. For any starting value of f^i and q^i, we observe erratic

15

Window size	Number of regimes	Starting value of f	Starting value of q	RMSE
1	1	0.0000	0.02	0.312131
2	1	0.0000	0.02	0.250943
3	1	0.0000	0.02	0.342138
4	1	0.0000	0.02	0.341751
5	1	0.0000	0.02	0.350081
1	1	-0.0001	0.04	0.156124
2	1	-0.0001	0.04	0.143903
3	1	-0.0001	0.04	0.120332
4	1	-0.0001	0.04	0.110002
5	1	-0.0001	0.04	0.123421

Table 5: RMSE results given number of regimes, size of filtering window and starting values model parameters under a one-state setting

Window size	Number of regimes	Starting value of f_i	Starting value of σ_i	RMSE
1	2	[-0.01 +0.01]	0.02	0.155199
2	2	[-0.01 +0.01]	0.02	0.105657
3	2	[-0.01 +0.01]	0.02	0.115599
4	2	[-0.01 +0.01]	0.02	0.091993
5	2	[-0.01 +0.01]	0.02	0.127700
4	2	[-0.01 +0.01]	0.03	0.090471
4	2	[-0.02 +0.02]	0.04	0.084289
4	2	[-0.05 +0.05]	0.08	0.121808
4	2	[-0.03 +0.03]	0.04	0.101739
4	3	[-0.02 0 +0.02]	0.02	0.089001
4	3	[-0.03 0 +0.03]	0.04	0.132655

Table 6: RMSE results given number of regimes, size of filtering window and starting values model parameters under a two-state and three-state settings

No. of regimes N	Likelihood value L	No. of parameters d	BIC
1	9278.33	2	9270.15
2	9691.38	6	9666.86
3	9700.23	8	9663.45

Table 7: Likelihood-based model selection analysis

trends of the transition probability matrix. In some instances, it has constant zeros and ones especially for models with low number of states. But starting values for parameters f and q can be chosen randomly. The only restriction for the initial values is to obtain convergence in the first iteration of the filtering algorithm. These starting values for f^i and q^i cannot be either too small or too large. Working with simulated data, we also found that that algorithm converges faster to the "true" values provided the number of regimes is chosen correctly. In our case, the numerical implementation of filters under the two-regime model yields the most stable parameter estimates.

4.3   Discussion of numerical results

The initial parameters were estimated using random subsets of the whole data focusing on the first 6 months. The HMM filtering algorithms were implemented to the remaining datasets with a moving window of four time steps to obtain estimated values of f^i and q^i for one-, two- and three-regime models. The evolution of transition probabilities is depicted in Figure 1. For the contract with longest maturity (expiry of 29 July 2011), the plots of f and q are exhibited in Figure 2, and they behave as expected. Considering that the data corresponds to the period when the economy was slowly recovering from the subprime financial crisis, we observed as anticipated, slowly decaying values of mean levels and constant behaviour after that period. A similar pattern for volatilities is obtained and there is a relatively big uncertainty at earlier periods but they leveled off not too long after a certain point. From the graph of f (Figure 2a) under one-regime framework, it is noticeable that the "true" values of f^is are always underestimated.

To evaluate the statistical significance of the RMSE values, we perform an F test. The estimated value of the F statistic for the comparison of RMSEs between one- and two-regime models is higher than the quantile value of the F distribution based on a 95% confidence level where the p-value is 3.91×10^{-7}. For the comparison of RMSEs between two- and three-regime models, the p-value is 3.57×10^{-11}. Hence, there is merit in using a regime-switching

framework. Considering that a two-state setting produces the best RMSE in Table 6, we conclude that the two-state Markov switching model is the most appropriate for our data.

This is further backed up by a likelihood-based selection criterion. The popular criterion for model selection is the Akaike information criterion (AIC). However, it is argued in Schwartz [27] that AIC may underestimate the optimal number of parameters, and thus, the Bayesian information criterion (BIC) is proposed as a robust alternative. The BIC metric is given

$$\text{BIC} = \ln L - \frac{1}{2} d \ln b,$$

where L is the likelihood function, d is the number of parameters in a model and b is the number of datapoints, respectively. The main idea is to choose the model with the highest BIC value. From Table 7, the BIC analysis, which combines goodness-of-fit criterion and penalty for model complexity, indicates that the two-regime model is the best given the dataset examined in this paper.
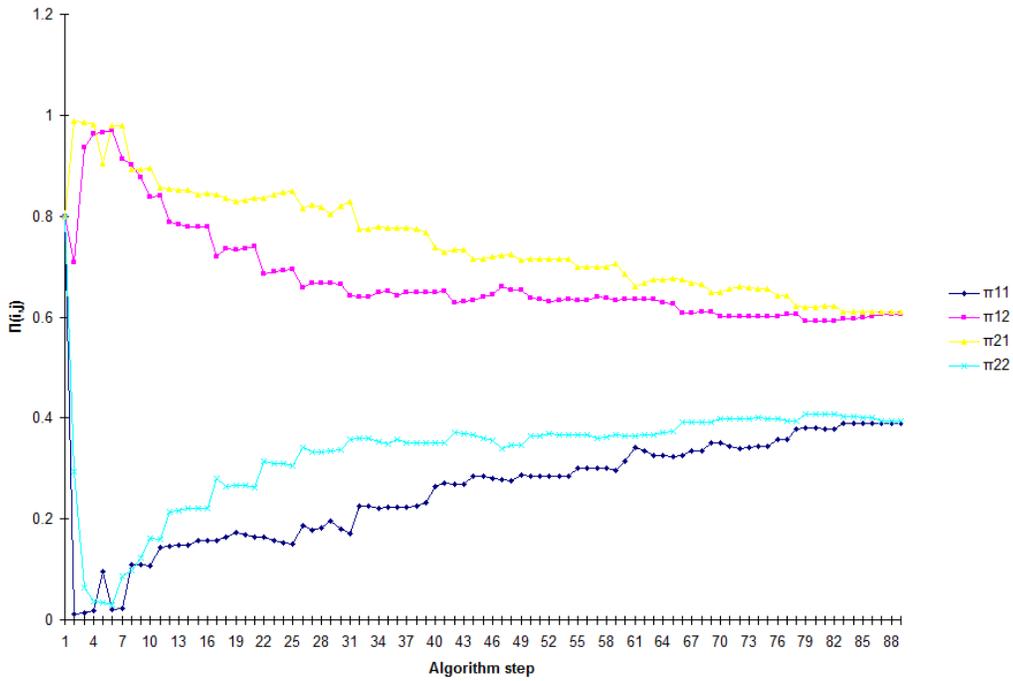
Figure 1: Evolution of transition probabilities

The proposed model in this paper is estimated by the maximum likelihood method. Hence, it appears that a likelihood ratio test (LRT) would be more appropriate in the comparison of embedded models, i.e., a lower dimensional model (say 2 regimes) is a restricted version of the higher dimensional model (say 3 regimes). However, it is noted in Hardy [21] that

18
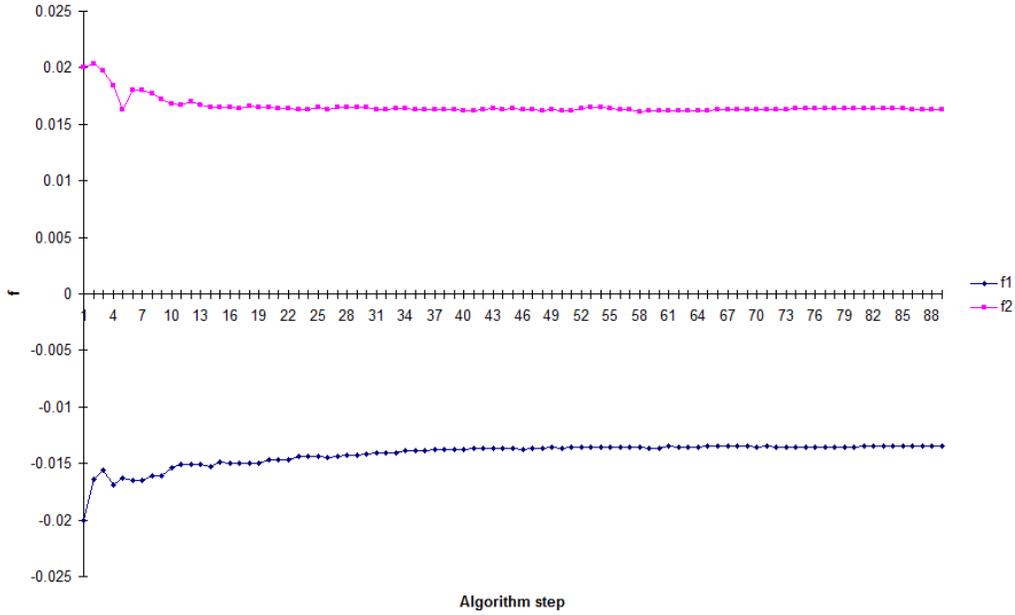
(a) Dynamics of $f(r)$ levels
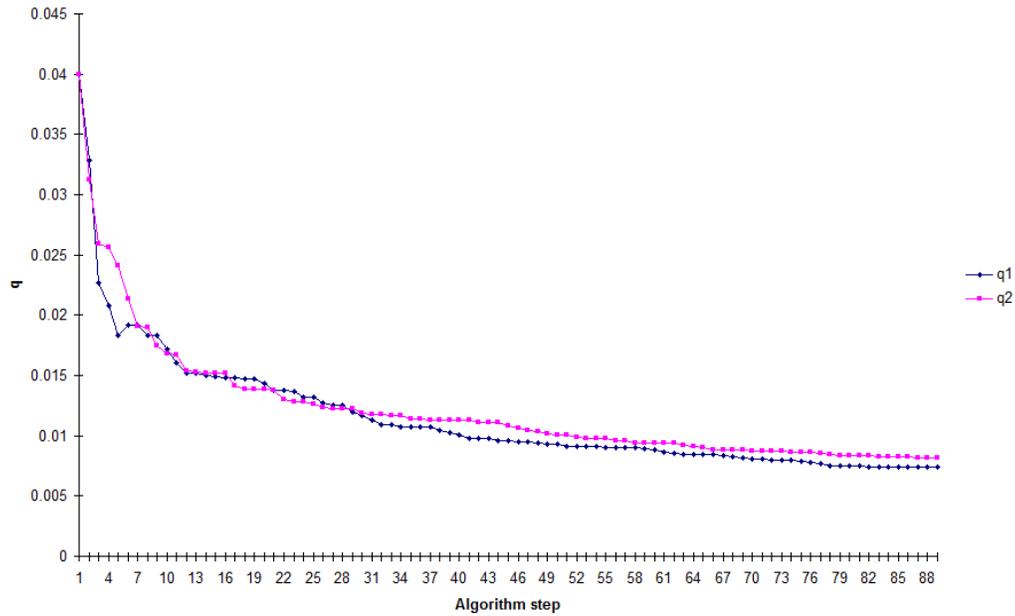
(b) Dynamics of $q(r)$ levels

Figure 2: Parameter estimates using data prices of futures contracts with expiry 29/07/2011

the LRT is not a valid test for the number of regimes in a regime-switching model. The critical issue about the LRT concerns the asymptotics of the estimator for its test statistic. The legitimacy of such asymptotics under regime-switching models is questionable. On
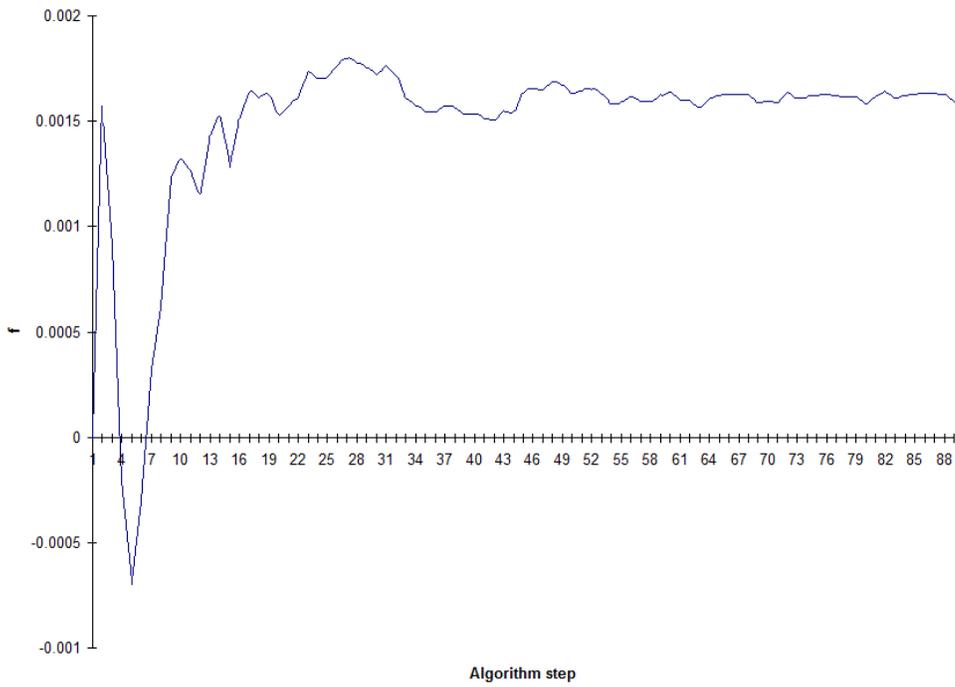
(a) Dynamics of f
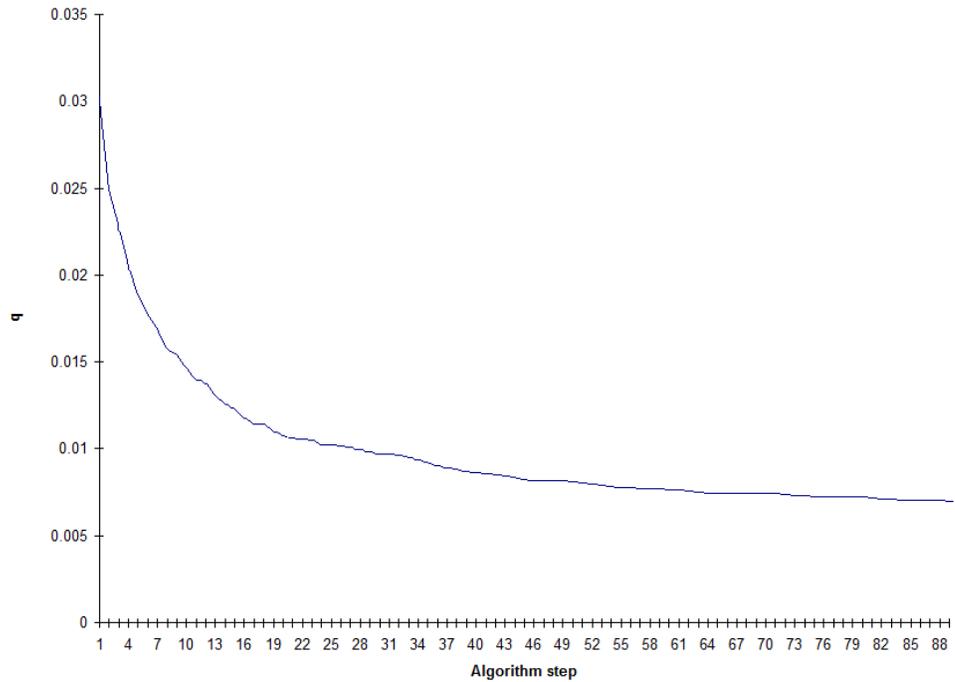
(b) Dynamics of q

Figure 3: Parameter estimates using data prices of futures contracts with expiry 29/07/2011 under a one-regime Markov chain

the other hand, the the BIC, although based on likelihood functions, does not involve any

20

test statistics and thus, there is no convergence/asymptotics to worry about. BIC is just a fitting-penalty-based criterion pretty much similar to RMSEs and other similar model-fitting metrics.

In particular, there are theoretical problems concerning the consistency of the estimator (i.e., asymptotic level of the test) since the regularity conditions are not satisfied under the null hypothesis; thus, the chi-square theory underpinning the LRT does not apply, see Gassiat and Keribin [19]. This is further substantiated in Chen, et al. [4] and Chen, et al. [5], where it is explicitly stated that the required LRT's regularity conditions are not fulfilled under mixture problems. In fact, it is also asserted in Chen and Kalbfleish [6] that the asymptotic properties of likelihood ratio statistics for testing the number of subpopulations are complicated and difficult to establish. So, whilst there are modified LRTs (with elaborate implementation requirements) tailored to regime-switching models, we opted to keep the model selection assessment simple by adhering to the BIC-based evaluation.

Finally, as can be seen from equation (6), the evolution of the market price of risk λ through time and maturity can be inferred and constructed once the estimates of other parameters are fully determined. Our numerical implementation with the use of the HMM filtering techniques and estimation of initial parameter values produce the λ dynamics given in Figure 4.

4.4   Prediction performance

The futures prices are treated as a 10-dimensional observation process, which includes contracts that mature from October 2010 onwards. The data up to 19th June 2009 (i.e., first 6 months as stated in subsection 4.3) is used for initial parameter estimation and is excluded from the one step ahead prediction. The prediction can be viewed as an 'out-of-sample' prediction (since κ is fixed after the initial parameter estimation). From equation (12), we have the dynamics of the vector process of price returns $y_k^i = \ln \frac{F^i(t_k)}{F^i(t_{k-1})}$ $i = 1, 2, \ldots, 10$. Therefore, the one-step ahead forecasts for $F_{t_{k+1}}^i$ is obtained through the forecast equation

$$E[F_{t_{k+1}}^i | \mathcal{F}_{t_k}^i] = F_{t_k}^i \sum_{r=1}^{N} \langle \hat{\mathbf{x}}_k, e_r \rangle \exp \left(f_{t_k}^i(r) + \frac{q_{t_k}^i(r)^2}{2} \right), \qquad (15)$$

where $\hat{\mathbf{x}}_k$ is the estimate for the unconditional distribution of the Markov chain. We utilise the estimates of $f_{t_k}^i(r)$ and $q_{t_k}^i(r)$ to get one-step ahead forecasts for $F_{t_{k+1}}^i$. The results are shown in Figure 5a. To complement the RMSE metrics in Tables 5 and 6, the criteria in
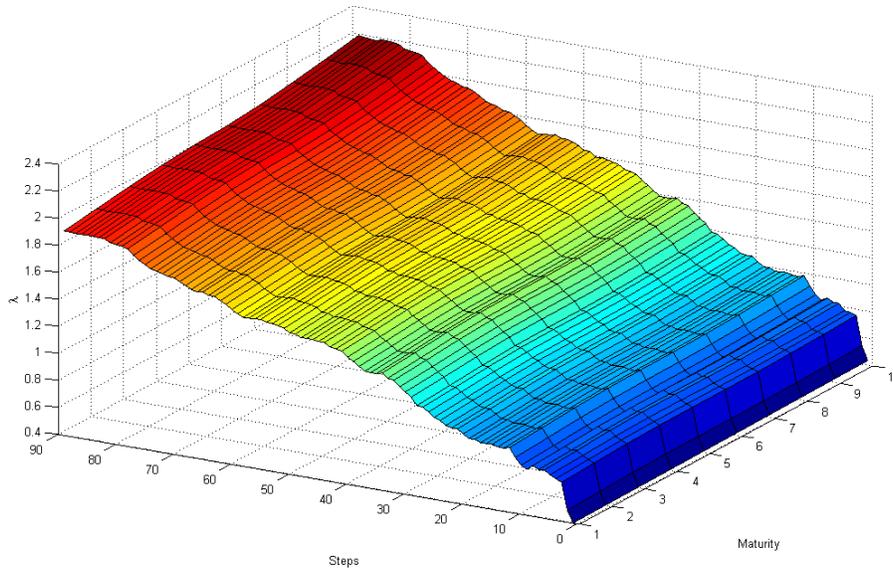
(a)
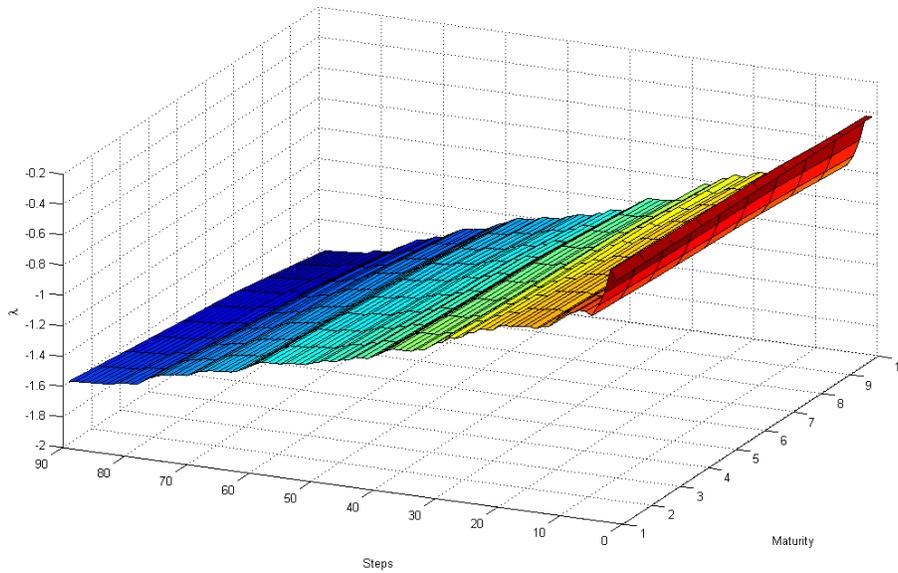
(b)

Figure 4: Dynamics of λ_t process under a 2-state setting corresponding to regime 1 in (a) and regime 2 in (b)

Hyndman and Koehler [23] in assessing the goodness of fit of the one-step ahead forecasts are adopted. The mean absolute percentage error (MAPE), median absolute percentage error (MdAPE) and median relative absolute error (MdRAE), for the 1-, 2- and 3-state HMM-based models are evaluated. The models are compared using these three criteria. The results of this error analysis are presented in Table 8. The two-state model outperforms both

Model setting	RMSE	MAPE	MdAPE	MdRAE
One-state model	0.11000	0.02769	0.02090	1.18122
Two-state model	0.08429	0.02395	0.01964	0.88347
Three-state model	0.13266	0.02590	0.01921	0.94448

Table 8: Further error analysis

the one-state and three-state models under the MAPE and MdRAE. The three-state model outperforms the two-state model under the MdAPE albeit the improvement is minimal.

Figure 5a displays the plots of the actual and one-step ahead predictions. A Q-Q plot depicted in Figure 5b strongly supports the initial assumption of using Brownian motion as a source of uncertainty in the model. From the plot of residuals against time shown in Figure 6a as well as Figure 6b, it is apparent that the assumption of constant variance is very reasonable. Whilst we would not generally expect this behaviour, one can see that from our dataset the variance does not change much for all regimes; see Figure 2b.

4.5 Further out-of-sample forecasting

As indicated in the beginning of this section, a one-step ahead prediction exercise using the data collected after 28 October 2010 was conducted. The new results are reported in Table 9 and show that they do not differ qualitatively from those given in Table 8. This particular experiment includes the examination of the prediction performance for very liquid futures contracts with short times to maturity since the dataset covers contracts with one month maturity (e.g., Nov 2010 expiration and Dec 2010 expiration). The dataset also contains contracts with medium-term maturities between 3 and 5 months (i.e., January 2011 to March 2011), and finally contracts with long-term maturities between 6 and 9 months (i.e., April 2011 to July 2011)

Since we are dealing with a multivariate dataset, it is worth emphasising that the values for the RMSE, MAPE, MdAPE and MdRAE in Table 8 are the average RMSEs, MAPEs, MdAPEs and MdRAEs, respectively, for 10 error vectors. In a similar manner, the error measures in Table 9 were obtained by considering 9 error vectors.
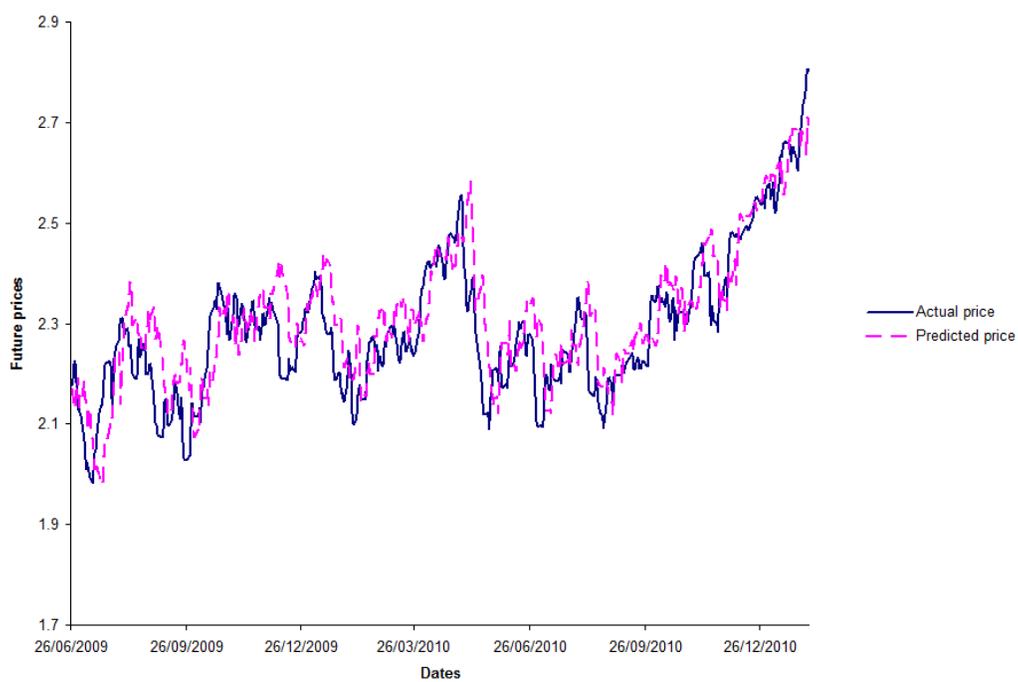
(a) One-step predictions
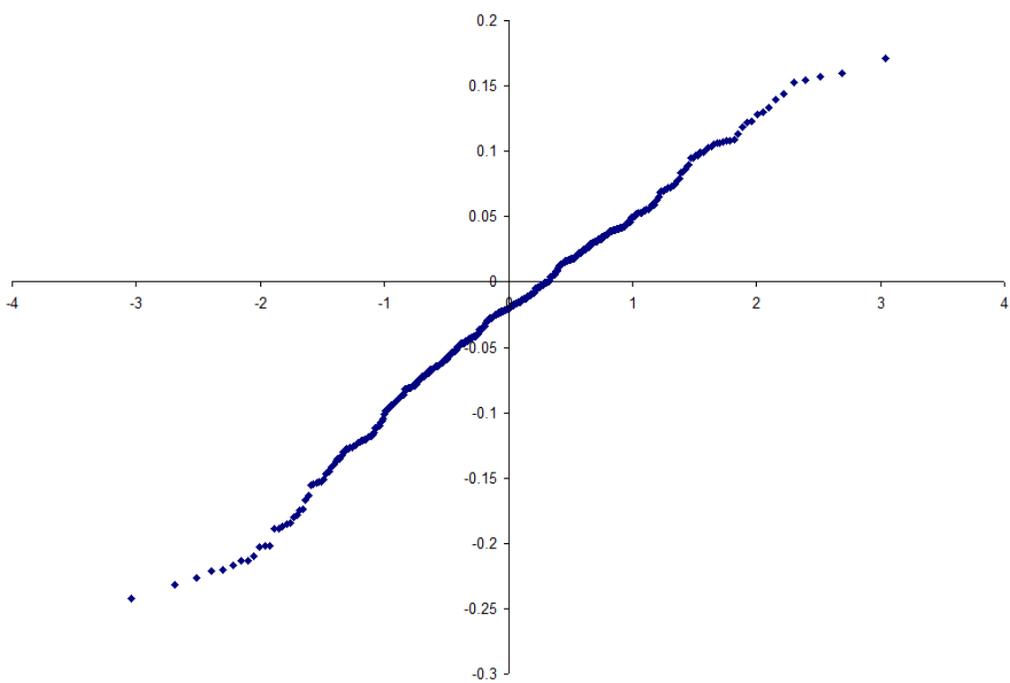
(b) Q-Q plot

Figure 5: One-step ahead forecasts and normal analysis of residuals
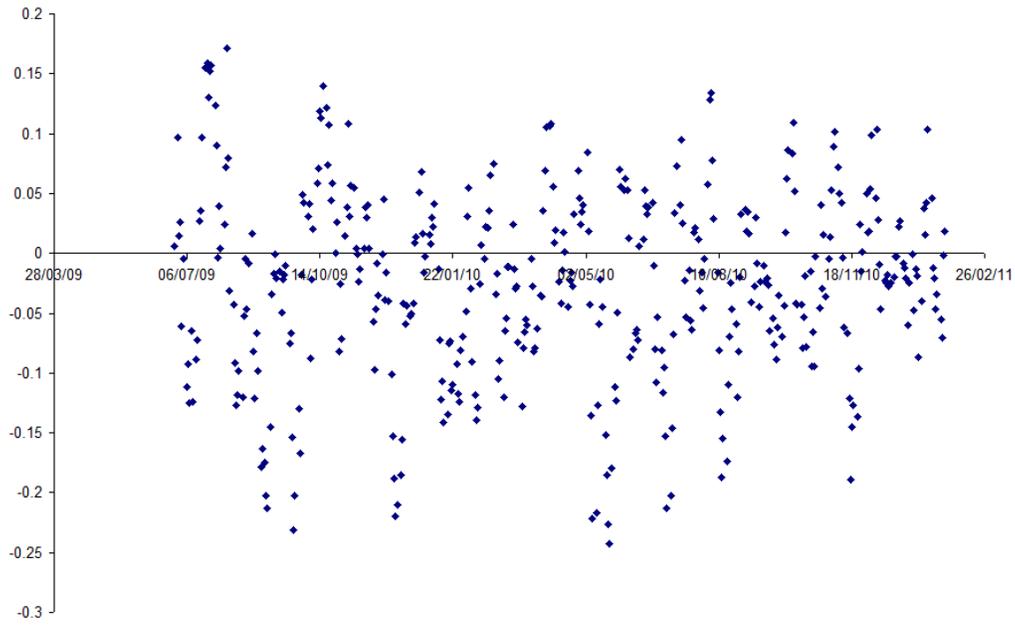
(a) Residuals
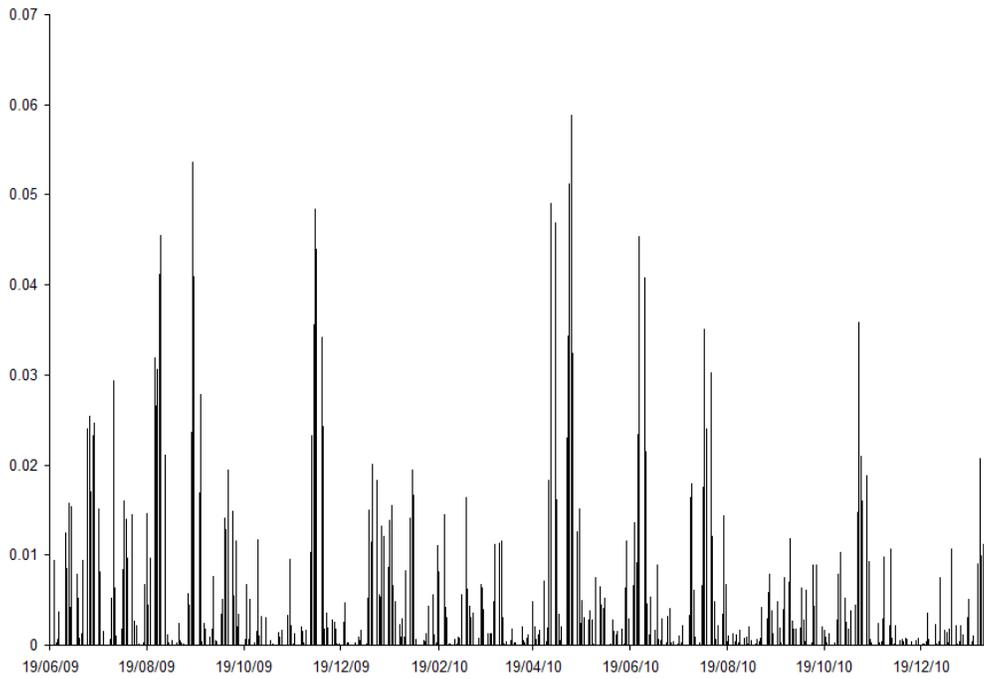
(b) Squared residuals

Figure 6: Residual analysis supporting the one-step ahead forecasting

Model setting	RMSE	MAPE	MdAPE	MdRAE
One-state model	0.08731	0.02491	0.01901	1.01019
Two-state model	0.08502	0.02045	0.01723	0.89197
Three-state model	0.09957	0.02902	0.01934	0.99129

Table 9: Further error analysis for observations after 28 October 2010 until 28 July 2011

5   Conclusion

In this paper, we performed an integration of various modelling ideas to model the evolution of arbitrage-free futures prices. In particular, the initialised one-state model gives poor prediction, whilst the use of a multi-state regime switching model will necessitate finding the rate of mean reversion κ using a mechanism other than direct self calibration. A calibration of a multiple regime model with a good prediction performance is possible only through a combination of the two calibration methodologies as elaborated in sections 3.1 and 3.2. We provided an approach and algorithms capable of providing parameter estimates for an arbitrage-free commodity futures price model. Given a dataset of futures prices, the recovery of parameter estimates is carried out under the assumption that parameters shift dynamically according to the state of the economy modulated by a hidden Markov chain.

To illustrate the numerical feasibility of our approach, we focused on two- and three- regime switching modelling frameworks. We detailed the solution of determining appropriate initial parameter values by considering an approximation to a non-convex optimisation problem. Assumptions on the model parameters are verified empirically. Self-calibrating HMM filtering algorithms are then able to produce dynamic parameter estimates reflecting the switching of economic regimes. The performance of the model is deemed adequate based on the analysis of one-step ahead forecasts and the accompanying post-model diagnostics. We benchmarked our results with those from the one-state setting and both the goodness-of-fit and information criterion metrics validate the merits of using a model with regime-switching feature. Although the focus of our application is on daily data, we experimented as well on weekly data (say, Wednesday's prices). The parameter estimates as expected would change but the trends of the volatilities and drifts as depicted by their graphs remain the same. The plots of the transition probabilities have similar decaying behaviour as in Figure 1. Nevertheless, their evolution through algorithm steps is different. We observed that the near terminal probability values would still approach the 0.6 and 0.4 limits.

Paper	Regime-switching?	Mean-reversion?	Spot prices observed?	Futures prices observed?
[13]	No	No*	No	Yes
[17]	Yes	Yes	Yes	No
[24]	No	Yes	No	Yes
This paper	Yes	Yes	No	Yes

Table 10: Comparison of this research with recent existing works
*–uses geometric Brownian motion for spot price

The relationship of this research with the existing recent works on modelling commodity futures price under some model characteristics is summarised in Table 10. Note that our method may also be employed for modelling the futures price evolution of other commodities such metals, agricultural products and raw materials. The proposed model is shown to be very parsimonious, with a two-regime one-factor model, which provides adequate performance in one step ahead out-of-sample forecasting on 10 measurement variables, having only 6 free parameters and no non-parametric seasonality adjustments. The algorithms proposed in this paper provides a very useful alternative to the existing methods of futures price modelling and forecasting. Accurate forecasting of commodity futures prices has important implications in various financial modelling applications such as the pricing of commodity spread options and calculation of quantile risk measures of commodity futures portfolios.

Acknowledgements

References

[1] Antonio, F., Aiube, L., Keshar, T., Baidya, N., Americo, E., Tito, H., 2008, Analysis of commodity prices with the particle filter. Energy Economics 30, 597–605.

[2] Back, J., Prokopzuk, M., Rudolf, M., 2013. Seasonality and the valuation of commodity options. Journal of Banking and Finance 37, 273–290.

[3] Buffington, J., Elliott, R., 2002. American options with regime switching. International Journal of Theoretical and Applied Finance 5, 497–514.

[4] Chen, H., Chen, J., Kalbfleisch, J., 2001. A modified likelihood ratio test for homogeneity in finite mixture models. Journal of the Royal Statistical Society Series B (Statistical Methodology) 63, 19–29.

[5] Chen, H., Chen, J., Kalbfleisch, J., 2004. Testing for a finite mixture models with two components. Journal of the Royal Statistical Society Series B (Statistical Methodology) 66, 95–115.

[6] Chen, J., Kalbfleisch, J., 2005. Modified likelihood ratio test in finite mixture models with structural parameter. Journal of Statistical Planning and Inference 129, 93–107.

[7] Chen, Z., Forsyth, P., 2010. Implications of a regime-switching model on natural gas storage valuation and optimal operation. Quantitative Finance 10, 159–176.

[8] Cortazar, G., Milla, C., Severino, F., 2008. A multicommodity model of futures prices: Using futures prices of one commodity to estimate the stochastic process of another. Journal of Futures Market 28, 537–560.

[9] Date, P., Ponomareva, K., 2011. Linear and non-linear filtering in mathematical finance: A review. IMA Journal of Management Mathematics 22, 195–211.

[10] Duffie, D., 1996. Dynamic Asset Pricing Theory. Princeton University Press, Princeton, New Jersey.

[11] Elliott, R., Chan, L., Siu, T., 2005. Option pricing and Esscher transform under regime switching. Annals of Finance 4, 423–432.

[12] Elliott, R., Hunter, W., Jamieson, B., 2001. Financial signal processing: A self-calibrating model. International Journal of Theoretical and Applied Finance 4, 567–584.

[13] Elliott, R., Hyndman, C., 2007. Parameter estimation in commodity markets: A filtering approach. Journal of Economic Dynamics and Control, 31, 2350–2373.

[14] Elliott, R., Chan, L., Siu, T., 2005. Option pricing and Esscher transform under regime switching. Annals of Finance 1, 423–432.

[15] Elliott, R., Siu, T., Badescu A., 2011. Bond valuation under a discrete-time regime-switching term structure model and its continuous time extension. Managerial Finance 37 1025–1047.

[16] Elliott, R., Siu, T., Chan, L., 2008. A PDE approach for risk measures for derivatives with regime switching. Annals of Finance 4, 55–74.

[17] Erlwein, C., Benth, F., Mamon, R., 2010. HMM filtering and parameter estimation of electricity spot price model. Energy Economics 32, 1034–1043.

[18] Erlwein, C., Mamon, R., Davison, M., 2011. An examination of HMM-based investment strategies for asset allocation. Applied Stochastic Models in Business and Industry 27, 204–221.

[19] Gassiat, E. and Keribin, C., 2000. The likelihood ratio test for the number of components in a mixture with Markov regime. ESAIM: Probability and Statistics 4, 25–52.

[20] Hamilton, J., 1994. Time Series Analysis. Princeton University Press, New Jersey.

[21] Hardy, M., 2003. Investment Guarantees: Modelling and Risk Management for eEquity-Linked Life Insurance. John Wiley & Sons, Inc, New Jersey.

[22] Hull, J., 2011. Options, Futures, and other Derivatives. Prentice Hall, Boston.

[23] Hyndman, R., Koehler, A., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679-688.

[24] Manoliu, M., Tompaidis, S., 2002. Energy futures: term structure models with Kalman filter estimation. Applied Mathematical Finance 9, 21–43.

[25] Mirantes, A., Poblacion, J., Serna, G., 2012. Stochastic seasonal behaviour of natural gas prices. European Financial Management 18, 410–443.

[26] Nakajima, K., Ohashi, K., 2012. A cointegrated commodity pricing model. Journal of Futures markets 32, 995–1033.

[27] Schwartz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Appendix
Recursive filters and EM updates

Proposition 1: *Define the diagonal matrix* $\mathbf{B}$ *whose* B_{ij} *entry is given by*

$$
B_{ij} = \begin{cases} \displaystyle\prod_{i=1}^{m} \frac{\phi\left(\frac{y_{k+1}^h - f_i^h}{q_i^h}\right)}{q_i^h \phi(y_{k+1}^h)} & for\ i = j \\ 0 & otherwise \end{cases} .
$$

Then

$$
\mathbf{c}_{k+1} = \mathbf{\Pi B c}_k, \tag{16}
$$

$$
\gamma\left(J^{(sr)}\mathbf{x}\right)_l = \mathbf{\Pi B}(y_l)\gamma\left(J^{(sr)}\mathbf{x}\right)_{l-1} + \langle \mathbf{c}_{l-1}, \mathbf{e}_r\rangle \left(\frac{\phi\left(\frac{y_l^i - f_r^i}{\sigma_r^i}\right)}{q_r^i \phi(y_l^i)}\right)^m \pi_{sr}\mathbf{e}_s, \tag{17}
$$

$$
\gamma\left(O^{(r)}\mathbf{x}\right)_l = \mathbf{\Pi B}(y_l)\gamma\left(O^{(r)}\mathbf{x}\right)_{l-1} + \langle \mathbf{c}_{l-1}, \mathbf{e}_r\rangle \left(\frac{\phi\left(\frac{y_l^i - f_r^i}{q_r^i}\right)}{\sigma_r^i \phi(y_l^i)}\right)^m \mathbf{\Pi e}_r, \tag{18}
$$

and

$$
\gamma\left(T^{(r)}\left(g(y)\right)\mathbf{x}\right)_l = \mathbf{\Pi B}(y_l)\gamma\left(T^{(r)}(g)\mathbf{x}\right)_{l-1} + \langle \mathbf{c}_{l-1}, \mathbf{e}_r\rangle \left(\frac{\phi\left(\frac{y_l^i - f_r^i}{\sigma_r^i}\right)}{q_r^i \phi(y_l^i)}\right)^m g(y_l^i)\mathbf{\Pi e}_r, \tag{19}
$$

where $g(y_l^i) = y_l^i$ *or* $(y_l^i)^2$.

Proposition 2: *Consider a multivariate dataset* $y_1^i,\ y_2^i, \ldots, y_k^i,\ 1 \le i \le m$ *observed up to time* k. *If the set of parameters* $\left\{\widehat{\pi}_{sr}, \widehat{f}_r^i, \widehat{q}_r^i\right\}$ *characterises the model then the EM estimates are given by*

$$
\widehat{\pi}_{sr} = \frac{\gamma\left(J^{(sr)}\right)_k}{\gamma\left(O^{(r)}\right)_k} \tag{20}
$$

$$
\widehat{f}_r^i = \frac{\gamma\left(T^{(r)}\left(y^i\right)\right)_k}{\gamma\left(O^{(r)}\right)_k} \tag{21}
$$

$$
\widehat{q}_r^i = \sqrt{\frac{\gamma\left(T^{(r)}\left(y^i\right)^2\right)_k - 2f_r^i\gamma\left(T^{(r)}(y^i)\right)_k + (f_r^i)^2\gamma\left(O^{(r)}\right)_k}{\gamma\left(O^{(r)}\right)_k}}. \tag{22}
$$