Filtering of an HMM-driven multivariate Ornstein-Uhlenbeck model with application to forecasting market liquidity

Anton Tenyakov  Rogemar Mamon *  Matt Davison

Department of Statistical and Actuarial Sciences
University of Western Ontario
London, Ontario, Canada

Abstract

This paper investigates the modelling of risk due to market and funding liquidity by capturing the joint dynamics of the Treasury-Eurodollar spread, VIX and a metric derived from S&P 500 time series. We put forward a two-regime mean-reverting model in explaining the behaviour of the liquidity levels in the financial markets. Expectation-maximisation algorithm in conjunction with multivariate filters is employed to construct optimal parameter estimates of the proposed model. The selection of modelling set-up is justified by balancing the best-fit criterion and model complexity. Using market data, the model performance is demonstrated by producing accurate prediction of market illiquidity states.

Keywords: Ornstein-Uhlenbeck process, Markov chain, change of measure, multivariate HMM filtering, TED, VIX, S&P 500, financial distress.

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

An asset’s ability to be sold without producing drastic movements in the price and with loss of value at a minimum is termed as market liquidity. Examples of liquid assets that can be used to meet immediate needs and wants through buying, selling, or paying debt, are cash and cashable instruments. Although currencies are liquid assets, they can suffer, even for major ones, from illiquidity (i.e., loss liquidity) under some significant liquidation events; liquidation is the exchange of a less liquid asset with a more liquid asset. For example, the US dollar and US dollar-linked assets could experience market illiquidity when countries holding trillions of dollars of US bonds start dumping US dollar bonds. The sale of an asset has, to a certain degree, an effect on the market. The business’s capacity to own sufficient liquid assets for the purpose of meeting its financial obligations is also referred to as liquidity. The importance of dealing with liquidity problem, and hence this paper, is akin to recent developments based on Basel III, which the US Federal Reserve uses as a liquidity requirement guideline for financial institutions. Basel III directives also require the diversification of counterparty risk and stress testing that could identify unusual market liquidity conditions. The goal of regulation is to prevent investments that are particularly susceptible to sudden liquidity shifts.

In 2007, the world is deemed to have experienced the worst financial crisis since the 1930s. It originated in the United States and spread across the global financial markets within less than a year. Some big financial organisations and banks declared bankruptcy. The downfall of Lehman Brothers is perhaps the most calamitous high-profile default from this crisis. Whilst many financial market events in 2007-2008 were considered to be of direct consequence of improper credit risk management, it is also believed that the main trigger of economic turmoil was the inability to predict liquidity in the markets; the AIG is a case in point. It is widely well-accepted that, ironically, efforts to mitigate credit/counterparty risk could create additional illiquidity, which on its own causes instability in the financial industry. In this research, we propose a method of quantifying and forecasting illiquidity in the financial market. As financial turbulence cannot be avoided, warning systems that aid the prediction of economic crunches are necessary to prepare market participants deal with future instability.

It is noted in Boudt, et al. [3] that the T-bill–Eurodollar (TED) spread is directly correlated with market stability. TED is calculated as the difference between the interest rate linked to interbank loans and the yield on short-term US T-bills. Currently, its computation
makes use of the three-month London Interbank Offer Rate (LIBOR) and the three-month T-bill yield rates. An increasing TED spread usually portends a meltdown of a stock market as it is taken as a sign of liquidity withdrawal. The TED spread, as described in the Bloomberg site, can gauge perceived liquidity risk in the general economy since T-bills are risk-free instruments and the funding liquidity risk of lending to commercial banks is encapsulated in the LIBOR. The rising of TED spread indicates that lenders view the default counterparty risk to be rising as well. Thus, lenders either require a higher rate of interest or settle for lower returns on safer instruments such as T-bills. Conversely, when the default risk of banks is decreasing, TED spread is falling; see Krugman [19].

We aim to use hidden Markov models (HMMs) driving a mean-reverting process in the analysis of the joint movements of important economic indicators to forecast liquidity and illiquidity states of the financial market. In this paper, we utilise observed TED spread data as market signals and filter out the state of the economy and subsequently the extent of liquidity level. Filtering results could be useful in assessing near-future market stability. The proposed idea is very similar to that of Abiad [1] wherein a regime-switching approach is used as an early warning device in identifying and characterising periods of currency crises. It has to be recognised, however, that a noticeable TED spread movement cannot be taken as a pure indication of a fallout. Whilst fluctuations in the TED spread happen due to some underlying factors, these fluctuations are sometimes merely caused by pure noise. In the late 1990s with the world battling the dot-com bubble and other financial upheavals, more instability and uncertainty in the behaviour of the TED spread ensued.

The second indicator for liquidity levels that we consider is VIX. This is a trademarked ticker symbol for the Chicago Board Options Exchange (CBOE)’s market volatility index and measures the implied volatility of S&P 500 index options. Using historical data, VIX appears to capture some periods of illiquidity that were not picked up by the TED spread. Finally, the third indicator we look at is a metric based on the evolution of the S&P 500. At the end of October 2012, market illiquidity was felt brought about by cautious trading as speculators and traders’ anxiously anticipated the result of the US presidential election result. Such presence of illiquidity was captured by an S&P 500-based metric but not by the TED spread. For this reason, a reasonably adequate study and modelling of liquidity can be accomplished by investigating the TED spread dynamics along with other indicators such as the VIX and an S&P 500-driven measure.

There has been many attempts to model and explain illiquidity such as those put forward
in van der End and Tabbae [24], Machini et al. [20], and Vayanos et al.[25], amongst others. Whilst these proposed modelling approaches include Monte Carlo simulations to demonstrate their implementability, they are nonetheless built on simple assumptions for tractability and do not offer the capacity for dynamic calibration using market data. This leaves a huge gap between theoretical approaches and real data’s stylised features along with model implementation. In this work, we attempt to address this gap by explaining how to fit the model with the data. With the aid of filtered estimates, we provide a description of the data dynamics with emphasis on the effect of illiquidity shocks.

In forecasting illiquidity, a discrete-time Markov chain is assumed to modulate the parameters of a mean-reverting process so that several economic regimes can be embedded into the model. As mentioned in Brunnermeier [4], the economy has a “spiral effect” and it is, therefore, reasonable to look at the Ornstein-Uhlenbeck process as a simple model for the TED spread and thus liquidity level in general. Goyenko [17] showed a strong correlation between TED and VIX as major indicators in illiquidity estimation; in the same paper, a mid-point measure for the evaluation of stock illiquidity was also presented. Our work is based on similar assumptions, but instead of finding correlation between the major indicators of illiquidity, we incorporate as much as possible information in our model by using simultaneously three market variables integrated by a set of multidimensional dynamic filters. For stock illiquidity, the S&P 500-based spread is used. The main consideration of this paper is the prediction of illiquidity level based on previous information contained in a joint time series of indicators. The dynamic filtering algorithm’s structure enables the finding of expected state probability of the illiquidity level at the next time steps.

The paper is organised in the following way. Section 2 gives an overview of the modelling set up including the HMM formulation and introduction of the change of measure concept for the filtering technique. A description of the mathematical filtering equations is also presented. We specify in section 3, the data used for the numerical estimation and prediction experiments. The process of recursive parameter calculation together with the discussion of the econometric interpretation of the dynamics of estimates are delineated in section 4. Finally, section 5 concludes.
2 Modelling setup

An Ornstein-Uhlenbeck (OU) process $r_t$ is any process that satisfies the stochastic differential equation (SDE)

$$dr_t = \theta(\mu - r_t)dt + \sigma dW_t,$$

where $W_t$ is a standard Brownian motion defined on some probability space $(\Omega, \mathcal{F}, P)$, and $\theta, \mu$ and $\sigma$ are constants independent of $W_t$. The parameter $\mu$ is the mean-reverting level that the process is trying to attain whilst $\theta$ is the speed of mean reversion, and $\sigma$ is the volatility. In the sequel, it is assumed that $\theta, \mu$ and $\sigma$ will be time-dependent; hence, we respectively denote them by $\theta_t, \mu_t$ and $\sigma_t$.

By Itô’s lemma, the solution of (1) is given by

$$r_t = r_0 e^{-\theta t} + (1 - e^{-\theta t}) \mu + \sigma e^{-\theta t} \int_0^t e^{\theta s} dW_s,$$

where $r_0$ is the initial value at time $t = 0$. To capture the switching of economic regimes, we assume that the values of parameters $\theta_t, \mu_t$ and $\sigma_t$ are modulated by a discrete-time Markov chain with a finite-state space. We regard the state of the underlying Markov chain as the regime of an economy or more specifically a liquidity regime dependent on major factors causing economic turbulence. For example, the scenario when $\theta_t, \mu_t$ and $\sigma_t$ are in the “lowest” or “worst” regime corresponds to the most unstable period of the global financial crisis; in this instance, $\mu_t$ reaches its minimum value whilst $\sigma_t$ will considerably spike up creating a completely unstable behaviour for $\theta_t$.

A particular contribution of this paper is the detailed implementation of parameter estimation under a multivariate OU setting thereby extending the one-dimensional framework of Erlwein and Mamon [14]. We consider $d$ OU processes; each process is denoted by $r_t^{(g)}$ with component $g \in \{1, \ldots, d\}$. All vectors and matrices are written in bold lowercase and uppercase letters, respectively. Following the idea developed by Elliott et al. [11] let us assume that $(\Omega, \mathcal{F}, P)$ is a probability space under which $x_k$ is a homogeneous Markov chain with a finite-state space in discrete time. Thus, $x_k$ evolves according to the equation

$$x_{k+1} = \Pi x_k + v_{k+1},$$

where $\Pi$ is a transition matrix and $v_{k+1}$ is a martingale increment, i.e., $E[v_{k+1}|\mathcal{C}_k] = 0$, where $\mathcal{C}_k = \mathcal{F}_k \vee \mathcal{R}_k$. Here, $\mathcal{F}_k = \sigma\{x_0, x_1, \ldots, x_k\}$ is the filtration generated by $x_0, x_1, \ldots, x_k$ and $\mathcal{R}_k$ is the filtration generated by the $\{r_k\}$ process.
With the closed-form solution in (2), each component of the \(d\)-dimensional observation process can be written as
\[
\begin{align*}
    r_k^{(g)} &= r_k e^{-\theta(g)(x_k)\Delta t} + (1 - e^{-\theta(g)(x_k)\Delta t}) \mu^{(g)}(x_k) \\
    &\quad + \sigma^{(g)}(x_k)e^{-\theta(g)(x_k)\Delta t} \int_{t_k}^{t_{k+1}} e^{\theta(g)(x_k)s} dW_s,
\end{align*}
\]
where \(\mu^{(g)} = (\mu_1^{(g)}, \mu_2^{(g)}, \ldots, \mu_N^{(g)})^\top\), \(\sigma^{(g)} = (\sigma_1^{(g)}, \sigma_2^{(g)}, \ldots, \sigma_N^{(g)})^\top\), \(\theta^{(g)} = (\theta_1^{(g)}, \theta_2^{(g)}, \ldots, \theta_N^{(g)})^\top\) \(\in \mathbb{R}^N\) and \(\Delta t = t_{k+1} - t_k\). For ease of calculation, the state space of \(x_k\) is associated with the canonical basis of \(\mathbb{R}^N\), which is the set of unit vectors \(e_h\), \(h = 1, 2, \ldots, N\) that is, \(e_h\) is a vector having 1 in its \(h^{th}\) entry and 0 elsewhere. So in equation (4), \(\mu^{(g)}(x_k) = \langle \mu_k^{(g)}, x_k \rangle\), \(\theta^{(g)}(x_k) = \langle \theta_k^{(g)}, x_k \rangle\) and \(\sigma^{(g)}(x_k) = \langle \sigma_k^{(g)}(x_k), \rangle\), where \(\langle \cdot, \cdot \rangle\) is the usual scalar product and \(\top\) denotes the transpose of a vector.

If \(x_k\) is constant on a small time interval \(\Delta t\) then using the property of a Gaussian distribution and the Itô’s isometry, the variance of \(r_{k+1}^{(g)}\) in (4) is
\[
\int_{t_k}^{t_{k+1}} e^{2\theta(g)(x_k)s} ds = \frac{1 - e^{-2\theta(g)(x_k)\Delta t}}{2\theta(g)(x_k)}.
\]
Equation (4) has the representation
\[
r_k^{(g)} = \nu^{(g)}(x_k)r_k^{(g)} + \xi^{(g)}(x_k) + \xi^{(g)}(x_k)\omega_k^{(g)}, \quad 1 \leq g \leq d,
\]
where \(\omega_k^{(1)}, \omega_k^{(2)}, \ldots, \omega_k^{(d)}\) are independent standard Gaussian random variables and
\[
\begin{align*}
    \nu^{(g)}(x_k) &= e^{-\theta(g)(x_k)\Delta t}, \\
    \xi^{(g)}(x_k) &= (1 - e^{-\theta(g)(x_k)\Delta t})\mu^{(g)}(x_k), \\
    \xi^{(g)}(x_k) &= \sigma^{(g)}(x_k)\sqrt{\frac{1 - e^{-2\theta(g)(x_k)\Delta t}}{2\theta(g)(x_k)}},
\end{align*}
\]
The succeeding calculations are inspired by the approach described in Elliott et al. [11], where filters are derived under some equivalent probability measure \(\tilde{P}\). Under this ideal measure, the observations are independent and identically distributed random variables making the calculations of conditional expectations easy. The filters, which are conditional expectations, are then related back to the real-world by the use of the Bayes’ theorem for conditional
expectation. The ideal measure \( \tilde{P} \) is equivalent to the real-world measure \( P \) via the Radon-Nikodym derivative constructed as

\[
\Lambda_K = \frac{dP}{d\tilde{P}} \bigg|_{C_K} = \prod_{g=1}^{d} \prod_{k=1}^{K} \lambda_{g,k}, \; K \geq 1, \; \Lambda_0 \equiv 1, \tag{10}
\]

where

\[
2 \ln(\lambda_{g,k}^{(g)}) = -\frac{r_{g,k-1}^{(g)}(x_{k-1}) + \zeta^{(g)}(x_{k-1}) - \left(r_{g,k-1}^{(g)}(x_{k-1}) + \zeta^{(g)}(x_{k-1})\right)^2}{\xi^{(g)}(x_{k-1})^2}. \tag{11}
\]

Write the conditional probability of \( x_k \) given \( C_k \) under \( P \) as

\[
p_i^k := P(x_k = e_h | C_k) = E[(x_k, e_h) | C_k],
\]

where \( \hat{p}_k = (\hat{p}_1^k, \hat{p}_2^k, \ldots, \hat{p}_N^k)^\top \in \mathbb{R}^N \). Now,

\[
\hat{p}_k = E[x_k | C_k] = \frac{E[\Lambda_k x_k | C_k]}{E[\Lambda_k | C_k]}
\]

by the Bayes’ theorem for conditional expectation. Let \( c_k = E[\Lambda_k x_k | C_k] \) and note that

\[
\sum_{i=1}^{N} \langle x_k, e_h \rangle = 1.
\]

Thus,

\[
\sum_{i=1}^{N} \langle c_k, e_i \rangle = \sum_{i=1}^{N} \langle E[\Lambda_k x_k | C_k], e_i \rangle = E \left[ \Lambda_k \sum_{i=1}^{N} \langle x_k, e_i \rangle | C_k \right] = E[\Lambda_k | C_k]. \tag{12}
\]

Consequently, equation (12) implies that

\[
\hat{p}_k = \frac{c_k}{\sum_{i=1}^{N} \langle c_k, e_h \rangle}.
\]

Similar to Erlwein, et al. [16] or Erlwein and Mamon [14], we define the following quantities:

\[
\mathcal{J}_{k+1}^j x = \sum_{n=1}^{k+1} \langle x_{n-1}, e_j \rangle \langle x_n, e_s \rangle \tag{13}
\]

\[
\mathcal{O}_{k+1}^j x = \sum_{n=1}^{k+1} \langle x_n, e_j \rangle \tag{14}
\]

\[
\mathcal{T}_{k+1}^j(f) x = \sum_{n=1}^{k+1} \langle x_{n-1}, e_j \rangle f(r_n), \; 1 \leq j \leq N. \tag{15}
\]
Equations (13) and (14) are the respective number of jumps from $e_s$ to $e_j$ and the amount of time that $x$ occupies the state $e_j$ up to $k+1$. The quantity $T_{k+1}^j(f)$ is an auxiliary process that depends on the function $f$ of the observation process; in our case, $f$ takes the form $f(r) = r$, $f(r) = r^2$ or $f(r) = r_{k+1} r_k$.

Other than generalising the framework in Erlwein and Mamon [14], our contribution includes expressing recursive filtering equations compactly though matrix notation. This allows efficient computation and decreases parameter estimation time using vector-optimised mathematical packages (e.g., MATLAB by Mathworks). Define the diagonal matrix $D(r_k)$ with elements $d_{i,j}$ by

$$
(d_{ij}(r_k)) = \begin{cases} 
\prod_{g=1}^{d} \exp \left( -\frac{r^{(g)} (r^{(g)} + \nu^{(g)} + \zeta^{(g)})) - (r^{(g)} + \nu^{(g)} + \zeta^{(g)})^2}{2\xi^{(g)}} \right) & \text{for } i = j \\
0 & \text{otherwise.} 
\end{cases}
$$

(16)

For any process $G_k$, we denote the conditional expectation, under $\bar{P}$, of $\Lambda_k G_k$ by $\gamma(G)_k := E[\Lambda_k G_k | C_k]$. We provide recursive filters for $c_k$, $\gamma(J^{j,i} x)_k$, $\gamma(O^{i} x)_k$ and $\gamma(T^{i}(f)^{(g)} x)_k$.

**Theorem 1:** Let $D$ be the matrix defined in (16). Then

$$c_k = \Pi D c_{k-1}$$

(17)

$$\gamma(J^{j,i} x)_k = \Pi D(r_k) \gamma(J^{j,i} x)_{k-1} + \langle c_{k-1}, e_i \rangle \langle D(r_k) e_i, e_j \rangle \pi_{ji} e_j$$

(18)

$$\gamma(O^{i} x)_k = \Pi D(r_k) \gamma(O^{i} x)_{k-1} + \langle c_{k-1}, e_i \rangle \langle D(r_k) e_i, e_i \rangle \Pi e_i$$

(19)

$$\gamma(T^{i}(f)^{(g)} x)_k = \Pi D(r_k) \gamma(T^{i}(f)^{(g)} x)_{k-1} + \langle c_{k-1}, e_i \rangle \langle D(r_k) e_i, e_i \rangle f(r^{(g)}_k) \Pi e_i.$$  

(20)

**Proof** The proof follows similar derivations of the filtering equations in Elliott [11], Erlwein, et al. [16], Erlwein and Mamon [14] or Mamon and Tenyakov [23].

To obtain the model parameter estimates, we use the Expectation-Maximisation (EM) algorithm [9]. The EM estimation for the multi-regime setting is very similar to that in the one-dimensional case illustrated in Erlwein and Mamon[14], and the proof of the next theorem is omitted.

As indicated in the above discussion, the model parameters $\nu^{(g)}$, $\zeta^{(g)}$ and $\xi^{(g)}$ have estimates
that depend on the filters of quantities given in Theorem 1. These dynamic parameter estimates are given as follows.

**Theorem 2:** If multivariate data set with row components \(r^{(g)}_1, r^{(g)}_2, \ldots, r^{(g)}_K\) \(1 \leq g \leq d\) is drawn from the model described in equation (6) then the EM parameter estimates are

\[
\hat{\pi}_{ji} = \frac{\gamma (J^j_i)^k}{\gamma (O^j_i)^k} \\
\hat{\nu}^{(g)}_i = \frac{\gamma (T^i (r^{(g)}_{k+1}, r^{(g)}_k))_k - \zeta^{(g)}_i \gamma (T^i (r^{(g)})_k)}{\gamma (O^j_i)^k} \\
\hat{\zeta}^{(g)}_i = \frac{\gamma (T^i (r^{(g)}))_{k+1} - \hat{\nu}^{(g)}_i \gamma (T^i (r^{(g)}))_k}{\gamma (O^j_i)^k} \\
\hat{\xi}^{(g)}_i = \frac{\gamma (T^i (r^{(g)}_{k+1}, r^{(g)}_k))_k + \hat{\xi}^{(g)}_i \gamma (T^i (r^{(g)}))_{k+1} + \hat{\nu}^{(g)}_i \hat{\zeta}^{(g)}_i \gamma (T^i (r^{(g)})_k)}{\gamma (O^j_i)^k}.
\]

**Proof** The derivations of (21) - (24), which generalise the filters for the univariate OU case, are straightforward based on Erlwein and Mamon [14].

**Remarks**

1. To implement the recursive equations in Theorem 1 in providing the dynamic updating of the estimates (21)–(24) under Theorem 2, note that \(\gamma (H^i)_k = \gamma (H^i \langle 1, x_k \rangle) = \langle 1, \gamma (H^i x_k) \rangle\), for some function \(H\). In our case, \(H\) can be replaced by \(J\), \(O\) or \(T\).

2. Equation (22) in Theorem 2 contains the parameter \(\zeta^{(g)}\), which must be known prior to achieving a workable recursion. In practice, the sequence of equations (22) and (23) can be implemented in a reverse order. That is, \(\hat{\zeta}^{(g)}\) can be estimated using the previous knowledge of \(\hat{\nu}^{(g)}\). This latter implementation was adopted in the empirical part of this paper, and we got significant stability in parameter estimates.

### 3 Description of data for implementation

To model the levels of liquidity, we use three monthly time series data covering the period of 30 April 1998–30 April 2013; data points are recorded at the last trading day of each month.
These data sets are: (i) TED spread obtained from Bloomberg, (ii) S&P 500 VIX compiled by the CBOE, and (iii) calculated average spread of S&P 500 based on the data collected by Bloomberg. The indicator in (iii) is adopted from Goyenko [17], and given by

\[ \text{MktIll} = 2 \frac{\text{Bid} - \text{Ask}}{\text{Bid} + \text{Ask}} \]  

(25)

where \( \text{Bid} \) and \( \text{Ask} \) are the respective bid and ask prices.

![Figure 1: Plot of TED, VIX and MktIll × 100](image)

The data set for our filtering applications is formed by creating a matrix with a dimension of \( 181 \times 3 \) over the period 30 April 1998 – 30 April 2013 with the TED, VIX and MktIll in the first, second and third columns, respectively. Figure 1 displays a visualisation of the movements of the TED spread, VIX and MktIll ×100 variables. Note that we use MktIll ×100 to scale the MktIll magnitude and make it comparable to that of the TED and VIX. The instability of the TED spread in the late 1990s - early 2000s is caused by the information technology bubble, political crisis in the United States, political and financial crises in the post-Soviet Russia, and recession in Japan.

The dot-com bubble, which was a price bubble, covers the period of 1997-2000 climaxing in March 2000. It is worth noting that these three indicators pin down the occurrence of
the financial market instability directly affecting liquidity. However, each indicator captures this instability at different moments and with different durations. These measures appear to “track down” the bubble inflation but some are more sensitive to others at certain times. The superiority of one measure over the others is therefore not clear. This is usually the case whenever the duration of the market crash is short and fast recovery of the economy is expected in general.

On the other hand, the subprime mortgage crisis in 2008-2009 was captured by all three measures altogether. The noticeably unusual spike in the TED spread though seems to clearly herald the coming of extreme financial meltdown that happened in August-September 2008. From the plot of the trivariate series, we also observe the cyclical behaviour of the economy shifting from stable to unstable states. This provides support for using a multivariate version of the OU process in modelling the generating process of the underlying data.

4 Numerical application

4.1 Calculation of estimates and other implementation assumptions

Several approaches may be employed to find initial parameters for filtering algorithms. These include the methods in Erlwein and Mamon [14], Erlwein, et al. [15], Mamon and Tenyakov [23], and Date and Ponamareva [7], amongst others. Good starting parameter values are necessary to stabilise the filtering algorithm procedure. However, estimating initial parameters is not straightforward considering the nature of the data and other factors. Whilst none of the initial-value estimation algorithms must be disregarded, the choice mainly depends on (i) achieving stable performance and (ii) relative ease of implementation. In this paper, we combine the above-mentioned approaches to generate reasonable initial estimates.

To choose the number of states in a regime-switching model, statistical inference-based methods such as the Akaike criterion information (AIC) [2], Bayes’ information criterion (BIC) described in Schwarz [21] and Hardy [18], or the CHull mentioned in Ceulemans and Kiers [6] may be utilised. These criteria are independent of the nature of the data; they are general tools that can be applied to any data set with an ultimate goal of selecting the model that optimally balances goodness of fit and model complexity. To make the discussion meaningful and the mathematics tractable, one may posit that the economy can only have two states - “high” and “low”. A transitional state may be created and persists over some
time due to the weighted combination of volatilities under two particular regimes.

Brokers who have long-term positions on different securities behave differently under expectations of future crash or upturn in their portfolios. In the trading world the value of the financial contract can only go up, go down or stay the same. In general, it is assumed that every stock always has a positive growth, i.e., it earns more money than what one can get from a risk-neutral investment. Therefore, even when the potential growth on a stock, mutual fund and other risky investment portfolios is minimal but the level of liquidity is high, the economy is still deemed to be in the good (or “high”) state. Consequently, when the percentage change in the value of the index or any other major indicators of the financial state of the country (GDP, for example) is relatively close to the risk-neutral rate, we regard the economy to be in the “high” state. Furthermore, we rely on the results of Boudt [3] and Dionne [10] advancing two-state models in investigating liquidity. Any three-state model will be shown later as a special case of the two-regime framework, where the third state is a state in between the “high” and “low” states.

Finding parameter estimates via the likelihood maximisation procedure is a tedious endeavour but such procedure provides best results for dynamic modelling if it can be accomplished. We shall use the first 40 points of the multidimensional data set to calculate the starting parameters for our filters. It is assumed at the outset that the set of parameters

$$\Xi = \{\pi_{ij}, \nu^{(g)}(x_i), \zeta^{(g)}(x_i), \xi^{(g)}(x_i)\}$$

is homogeneous, i.e., the values of the set $\Xi$ do not change when subsets of the data are chosen. The likelihood function, conditional on knowing which state the process $x_i$ is in, is given by

$$L = \prod_{g=1}^{d} \prod_{i=1}^{K} \frac{1}{\sqrt{2\pi((\xi^{(g)}(x_i)))^2}} \exp \left( -\frac{(r_{i+1}^{(g)} - \nu^{(g)}(x_i)r_i^{(g)} - \zeta^{(g)}(x_i))^2}{2(\xi^{(g)}(x_i))^2} \right)$$

or

$$L = \prod_{g=1}^{d} \prod_{i=1}^{K} \phi_{g,i},$$

where

$$\phi_{g,i} = \phi \left( \frac{r_{i+1}^{(g)} - \nu^{(g)}(x_i)r_i^{(g)} - \zeta^{(g)}(x_i)}{\xi^{(g)}(x_i)} \right),$$

and $\phi$ stands for the density of the standard Gaussian distribution.
As the sequence of the states $x_i$ is hidden, a recursive algorithm similar to the one proposed in Hardy [18] is used. The idea of the algorithm is to calculate the most probable set $\Xi$ by building the likelihood function using recursions and to apply standard computer routines for maximisation of the function over a desired set of parameters. Although Hardy’s method was designed for the geometric Brownian motion model, we extend it in a straightforward manner to handle our multidimensional data set assumed to follow the OU process. Additionally, the definition of the log-likelihood function is extended by using the sum of the log-likelihood functions for the TED, VIX and MktIll data sets. Adopting the notation of this article, the density function of the process at time $t$ in Hardy’s algorithm, given the whole set of parameters including the state of the Markov chain at time $t$, is changed to $\phi_{g,i}(r_{i+1}, r_i, \Xi)$.

The results of the initial parameter estimation under the two-state model are provided in Table 1. The values of parameters $\nu_1$ and $\nu_2$, encapsulating the speed of mean reversion, can be considered to be almost the same for all three variables. This empirical result that these parameters have uniformly close estimated values is no coincident and gives additional strong support to the hypothesis about the dependency of TED, VIX and MktIll on the same underlying factor.

<table>
<thead>
<tr>
<th></th>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$\zeta_1$</th>
<th>$\zeta_2$</th>
<th>$\xi_1$</th>
<th>$\xi_2$</th>
<th>$\pi_{12}$</th>
<th>$\pi_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TED</td>
<td>0.5178</td>
<td>1.4795</td>
<td>0.1478</td>
<td>-0.6130</td>
<td>0.6119</td>
<td>1.3191</td>
<td>0.6010</td>
<td>0.0076</td>
</tr>
<tr>
<td>VIX</td>
<td>0.5709</td>
<td>1.5484</td>
<td>0.4852</td>
<td>-0.1123</td>
<td>0.0212</td>
<td>0.4137</td>
<td>0.6010</td>
<td>0.0076</td>
</tr>
<tr>
<td>MktIll</td>
<td>0.5428</td>
<td>1.3090</td>
<td>0.0079</td>
<td>0.0016</td>
<td>0.0006</td>
<td>0.0001</td>
<td>0.6010</td>
<td>0.0076</td>
</tr>
</tbody>
</table>

Table 1: Initial parameter estimates for the filtering algorithms under the two-state setting

The HMM filtering algorithms can only give a local maximum, and at times could be extremely unstable to implement. Such limitation can be rectified by choosing initial estimates that fit the data very well and working in double precision arithmetics. It is also possible to employ some symbolic packages such as Mathematica by Wolfram Research, but in that case the speed of the computation drops dramatically. Whilst there is a variety of methods to determine starting values, the static log-likelihood maximisation approach appears to yield initial parameters that afford appreciable stability for our OU-based filters.

The starting values for the one-regime model are obtained by simply maximising the likeli-
Table 2: Initial parameter estimates for the filtering algorithms under the one-state setting

<table>
<thead>
<tr>
<th></th>
<th>ν</th>
<th>ζ</th>
<th>ξ</th>
</tr>
</thead>
<tbody>
<tr>
<td>TED</td>
<td>1.2201</td>
<td>-0.1475</td>
<td>2.0761</td>
</tr>
<tr>
<td>VIX</td>
<td>0.5911</td>
<td>0.1295</td>
<td>0.0223</td>
</tr>
<tr>
<td>MktIll</td>
<td>0.8580</td>
<td>0.0023</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

hood function (26), taking into account that \( x = 1 \), i.e., the system always operates under one state. The results of this optimisation are exhibited in Table 2. As expected, the initial parameter values for the single state model lie between the corresponding estimates for the two-regime model. The only parameter which does not follow this observation is \( \xi \); but even then such \( \xi \) values corresponding to the three indicators produce a stable convergence for the filtering procedure outlined in the next subsection.

4.2 Filtering procedure

In the estimation of the parameters of the underlying OU process, we employ the method described in section 2 using the starting parameters in Tables 1 and 2. The data set described in section 3 contains columns consisting of \( g = 1, 2, 3 \) (1 \( \equiv \) TED, 2 \( \equiv \) VIX and 3 \( \equiv \) MktIll ). There are 141 time points considered in our filtering application. The first 40 time points for the three vectors of data are used for the initialisation discussed in subsection 4.1. The prediction power of the model is tested on the last 60 monthly observations from the middle of the financial crisis (30 May 2008) up to the end of the time series data (30 April 2013). All results are analysed and evaluated using a combination of both intuitive and rigorous statistical approaches for decision making.

The dynamics of the estimates \( \theta, \mu \) and \( \sigma \) are computed by first producing the estimates of \( \nu, \zeta \) and \( \xi \). Then using equations (7), (8) and (9), we back out the values of the desired model parameters. The OU filters described in section 2 were implemented with a moving window spanning vectors of data similar to the idea in Tenyakov and Mamon [23] and extending the procedure in Erlwein and Mamon [14]. More specifically, vectors of data are processed through the recursive equations of the filter to obtain the best estimates (in the sense of conditional expectation) after several time points. Once the parameters are estimated from a batch of vector of data points, these estimates are used as starting values for the next recursion, and so on. The size of the processing window is determined by likelihood-maximum or other statistical criterion. Owing to the complex nature of the data and the
filtering equations, we employ the smallest window possible (3 points per window in our case) that gives stability to the algorithms. Whilst this choice results to relatively high volatility, the outputs contain an ample amount of insightful information about parameter fluctuations.

Figures 2–5 provide output for the parameter estimates of the OU process corresponding to TED spread.

An implication that can be drawn from the behaviour of transition probabilities in Figure 5 is that major illiquidity events do not happen very often, but when they do, they do not last very long and not severe up until the financial market collapse in 2007-2008. After the crisis in 2008, the structure of the economy changed completely. This fact is supported by Figure 2, wherein the mean-reverting levels are switched. This phenomenon occasionally arise in similar filtering applications (e.g., Xi and Mamon [26] or Xi and Mamon [27]). In our case, this anomaly can be explained by the presence of higher volatility levels during the times with greater uncertainty. This is substantiated by Figure 4 as the level of $\sigma$ reaches the highest level in 2008.

The other odd behaviour shown by the filtering is the negativity of $\theta$ in Figure 3. Even though the formulation of the OU process does not allow the parameter $\theta$ to be negative,
the multi-regime construction of OU process, proposed by Elliott and Wilson [13], does not restrict $\theta$ to be always positive. From an empirical perspective, getting $\theta < 0$ is justified by the fact the OU process has become unstable as can be seen during the 2007–2008 period when there was a sudden unfolding of several related financial and economic events leading to the crisis, and exacerbated by too much uncertainty and unpredictability of the economy. However, during stable periods, the speed of reversion remains positive and this parameter is interpreted in the usual sense.

The results of the dynamic parameter estimation based on filtering under the one-regime framework are illustrated in Figures 6–8. The dynamics of the parameters look similar to those of the two-regime model. This fact can be interpreted as an excellent fit of the 2-state HMM-modulated OU model to the data set. We use the AIC and BIC tailored to several previous works on filtering, in particular, Mamon and Tenyakov [23], to show that the proposed two-regime model provides a better explanation of the data compared to the one-regime setting. The AIC and BIC metrics are computed as

$$AIC = \ln L - p$$

(29)
Figure 4: Evolution of the volatility levels for the TED spread data

Figure 5: Evolution of the filtered transition probabilities obtained from the multivariate data
Figure 6: Evolution of the mean-reverting level under the one-state setting using the TED spread data

<table>
<thead>
<tr>
<th>Regimes</th>
<th>Log-likelihood</th>
<th>BIC</th>
<th>AIC</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>491.0379</td>
<td>446.4991</td>
<td>482.0379</td>
<td>9</td>
</tr>
<tr>
<td>II</td>
<td>569.6856</td>
<td>470.7104</td>
<td>549.6856</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3: Comparison of selection criteria for single- and 2-state regime models

and

\[ BIC = \ln L - \frac{1}{2} p \ln N, \]  

where \( L \) is the log-likelihood function for the entire multivariate data set, \( N \) is the number of observations and \( p \) is the number of parameters in a model. With the calculated value for the the log-likelihood function of the last 141 row vector of values, both the AIC and BIC signifies that the two-regime model significant outperforms the one-state model; see Table 3.

The general trend of the behaviour of the data during the crisis in 2007-2008 is captured by the model. Nonetheless, due to extreme volatility movements, getting a perfect fit during this period is a challenge. Given the initial parameter estimates, we report that it takes 5-6 algorithm steps for the OU filters to adjust and maintain some stability.
4.3 Forecasting illiquidity

There are many approaches in modeling illiquidity based on TED spread or some other major economic factors. However, to the best of our knowledge these approaches either look at a certain threshold of the TED spread as benchmark for illiquidity (see, Boudt, et al. [3] and Krugman [19]) or correlate the TED spread with another major economic variable (see Goyenko [17]).

In this work, we introduce a new approach which is naturally suited for dynamic filtering algorithms. It relies on the dynamics of $\hat{p}_k = E[x_k|C_k]$. As previously specified, the two-regime model is instructive in that each regime corresponds to illiquid and liquid states of the market. We put forward that if $\hat{p}_k(1) = \langle \hat{p}, e_1 \rangle >> 0.5$, i.e., the probability of being in regime 1 is very high, the market is extremely liquid and it is therefore easy to buy and sell every contract. But, if $\hat{p}_k(1) << 0.5$, which is equivalent to $\hat{p}_k(2) = \langle \hat{p}, e_2 \rangle >> 0.5$, then the market is very illiquid, which typically corresponds to recession or period of economic crisis brought about by some major financial events.
Accordingly, we use the following technique to predict the state of liquidity. Firstly, $\hat{p}_k$ is computed for some $k$, and secondly, expectation of $\hat{p}_{k+1}$ given $C_k$ is calculated as

$$E[\hat{p}_{k+1}|C_k] = \Pi_k \hat{p}_k,$$  \hspace{1cm} (31)

where $\Pi_k$ is defined by equation (3).

In Figure 9, we depict the dynamics of the prediction values $E[\hat{p}_{k}(1)|C_{k-1}]$ and estimates of $\hat{p}_{k}(1)$ obtained by applying the filtering algorithms on the last 60 points of the data set. The drastic change in the movement of estimated probabilities between the fourth and fifth time points corresponds to a significant drop in the values of all the three variables (TED, VIX, MktIll) in April-May 2009; such twist was perfectly captured by the dynamics of $\hat{p}_{k}(1)$.

To distinguish liquid from illiquid state, we propose a criterion that hinges on $\hat{p}_{k}(i) >> 0.5$ for $i = 1, 2$. If $\hat{p}_{k}(1) > 0.6$, it is assumed that there is enough evidence to conclude that the level of liquidity in the market is high, and traders can take positions with little or no probability of acquiring additional risks due to market or funding liquidity. So, the higher $\hat{p}_{k}(1)$ is, the higher is the liquidity level. Whenever $\hat{p}_{k}(1) < 0.4$, the financial markets are assumed to experience illiquidity, and therefore, additional capital has to be infused to deal
with the financial distress.

Based on empirical evidence, typifying exactly the liquidity state when $0.4 < \hat{p}_k(i) < 0.6$, $i = 1, 2$ is not an easy endeavour. This situation is characterised by a very high level of uncertainty regarding market directions over a short time. On the one hand, the “state of uncertainty” signals the occurrence of future hard times. On the other hand, it can also be viewed as a sign that economic stability is forthcoming after an economic downturn. The case in point here is the period of early 2009, when regulators used all possible schemes to stabilise the market sentiments and provided instant artificial liquidity to help markets function the way they were intended to be. We note that our proposed model gives somewhat overoptimistic estimates for $E[\hat{p}_k(1)]$ during the last period of the market crash in 2008, and this requires some adjustment. However, the estimates for the $\hat{p}_k(1)$ still remain at the 0.4 level, coinciding with what was previously argued concerning “state of uncertainty” and artificial liquidity. Predictors for the last 48 data-points are impressively quite accurate and they capture the dynamics of the markets reasonably well.

The “state of uncertainty” can be viewed either as a third regime in a two-state model, which is interpreted as the lowest/worst bound for the “high” regime and upper/best bound for the “low” regime. This can be explained from the econometrics point of view. Recession and upturn times in the economy are generally followed by short periods of market anxiety. During these unstable periods, liquidity can rise and fall quite frequently because speculators do not have stable expectations for the long-term horizons and short-term government interventions can provide only temporally relief. It is rather difficult to capture that “state” as it has in a way the characteristics of either regime. Of course, the stability of a possible separate three-regime model must be investigated as well, and could certainly be an alternative model. Nevertheless, preliminarily results in our case reveal that recursive algorithms do not provide even an approximate convergence for finding the starting parameters for the three-state model. Thus, we rule out the appropriateness of the dynamic three-regime setting.

5 Concluding remarks

In this work, we developed an HMM-based modelling approach in assessing levels of market and funding liquidity risks. The structure of the proposed model incorporates major econometric assumptions to deal with economic spirals. We provided a detailed methodology on how to extract information from major economic indicators, and linking these to
Figure 9: Evolution of the estimated liquidity-state probabilities and one-step ahead forecasts of liquidity-state probabilities

the short-term prediction of market illiquidity or liquidity. Such methodology made use of newly developed multivariate HMM recursive filtering algorithms expressed in matrix representations. Effects of mean-reversion and liquidity state dependency were explored as well.

The model’s implementability and forecasting performance were investigated using market data. Results were analysed against statistical metrics and interpreted by examining underlying historical financial events. We found that the one-regime model significantly underperforms compared to a two-regime model. Undoubtedly, a simple OU process cannot capture adequately the whole complexity of the liquidity risk in the financial market. A technique for liquidity-state estimation naturally consistent with dynamic HMM filtering algorithms was put forward and its validity was evaluated using past data.

An improvement that could be done with our modelling approach is the further examination of the two-state model. Its predictability of liquidity becomes uncertain if the conditional probability of the Markov chain falls in the range $[0.4, 0.6]$. Despite our empirical and economic reasoning to support our assumption and conclusions under this scenario, additional analysis of this particular aspect is a good research direction. Our primary results suggest that the three-state model cannot be fitted given the data we examined. There is a possibility though that the three-regime model may work by adding some other economic variables portraying clear multi-regime behaviour.
Our suggested modelling construction and empirical work took into account monthly data. Building on our results, further analysis of data with different frequency could be carried out to open avenues for modelling methods and insights about liquidity risk over a long or very short-time periods. These entail establishing new drivers, factors and determinants of liquidity to be included in the filtering experiments. The HMM-driven OU process may have to be tweaked to accommodate these new inputs leading to new filters.

Finally, the current recommended modelling and estimation set up can be effectively exploited under sophisticated trading-scheme environments. For example, underlying variables involved in trading, valuation or reserve calculation for financial derivative contracts, are known to follow the OU process. Our filtering equations can be employed to provide dynamic parameter estimates both for pricing and risk management. Regulators may also consider this model to study the impact of different constraints on the economy.
References


