Modelling high-frequency FX rate dynamics: A zero-delay multi-dimensional HMM-based approach

Anton Tenyakov † Rogemar Mamon †, *

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

Abstract

We develop a zero-delay hidden Markov model (HMM) to capture the evolution of multivariate foreign exchange (FX) rate data under a frequent trading environment. Recursive filters for the Markov chain and pertinent quantities are derived, and subsequently employed to obtain estimates for model parameters. The rationale of zero-delay HMM hinges on the idea that with fast trading, available information must be incorporated immediately in the evolution equations of the financial variables being modelled. Our proposed model is compared with the usual one-step delay HMM, GARCH and random walk models using likelihood-based criteria and error-type metrics. Parameter estimation both under the static and dynamic settings are carried out as well as in the models used as benchmarks in a comparative analysis. Implementation details are provided. We include a numerical illustration of the methodology applied to the currency data on UK sterling pounds and US dollars both against the Japanese yen. Our empirical results demonstrate greater fitting capacity and forecasting power of the zero-delay HMM over the comparators included in our analysis.

Keywords: High-frequency trading, zero-delay model, Markov chain, change of measure, multivariate HMM filtering, Japanese yen

*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

Trading that entails days and weeks to complete decades ago are now being carried out in a fraction of seconds. There is apparently a shift in the market in recent years from the traditional long-term buy-and-hold strategy to short-term trading that involves fast execution of quote orders by computer programmes creating a cascade of buying and selling securities. Such high-frequency trading (HFT) employs modern technological tools and algorithms to achieve rapid trading. Trading firms, such as hedge funds, rake in profits from moving in and out of positions in microseconds to trade stocks, bonds and futures taking advantage of minute price differences detected by computers. Critics say high-speed trading exaggerate wild price swings, but advocates assert it helps provide price discovery and liquidity to the marketplace.

Shah [28] reported that computerised HFT has been making inroads in the market for currency derivatives. As indicated by the Boston-based consulting firm Aite Group, HFT already accounts for up to 30% of activity in the global FX market, mostly in heavily traded currencies like the dollar. In the same report, it was also noted that Credit Suisse, for example, maintains an advanced execution services system for its lines of FX products globally, and such automated system allows its traders to manage option risks. Indeed, whilst the use of HFT is widespread in equities and commodities market, banks and hedge funds have been continuously making a push into niche areas including FX derivatives and the more thinly emerging market currencies. Given these developments and the fact that the value of currency derivatives depend on the movements of FX rates, there is therefore a strong demand for fast and reliable modelling of FX rates’ future evolution as well as the accurate estimation of their volatilities.

A dependable FX rate model is essential in the pricing of FX derivatives as well as in the risk management and optimisation of portfolios containing assets and products that have FX rate exposures. In addition to the ability of the model to adequately replicate the stylised features of the FX process, there is also a need to be able to dynamically implement the model with relative ease and accessibility. As pointed out in Meese and Rogoff [23], a simple random walk (RW) describes the FX rate dynamics better than most of the suggested modelling approaches. This has the implication that majority of previously proposed FX rate models are unsatisfactory in terms of their out-of-sample forecasting and statistical fitting performance. In a paper that examines several FX rate models showing their advantages and disadvantages and illustrating the non-existence of a strong correlation between the
behaviour of major economic factors and the FX rate dynamics, Nailliu and King [24] also found that only a few models can do better than the RW model in a comparative survey covering the last 3 decades of data.

In this paper, we propose an alternative model along with its implementation methodology in explaining better the behaviour of FX rates. We show that under some performance metrics, this proposed model outperforms the RW model. Our methodology and modelling formulation reflect the present frequent trading practices of major players in the FX market as well as the mechanism to capture the appropriate state of the whole economy as time goes by. Using various statistical tests and empirical data, we show certain advantages of a zero-delay multivariate HMM in terms of its fitting and prediction capacity. As stated in [24], it is possible to predict FX spot rates more accurately than those produced by the RW model if information about the structure of the market trading is aptly incorporated in the model. Inspired by this realisation, we introduce a model, given its simplicity and implementation attainability with respect to the current computing technologies, that may be easily adopted by the practitioners. This augments the currently used time series based models such as the ARCH/GARCH and stochastic volatility (SV) models. Within the context of FX rate modelling, a vast literature has sprung from Engel’s ARCH model [13] as noted in Maheu and McCurdy [22] and a discussion of the SV approach is given in Taylor [29].

Certain information that drastically affects market prices of frequently traded assets has an instant influence in the dynamics of other financial variables such as FX rates. As elaborated in Haldane [18], traders engaging in HFT worldwide via computer programmes could automatically change their positions and bring some over-all 40,000 spot-price changes in less than a second. Thus, the assumption of current price dependence on previous asset information recorded on a time lag of one day or wider frequency, which is the usual framework for many discrete-time HMM-based approaches, is deemed insufficient. With liquid trading that affords volumes of readily available information nowadays, the utility of the zero-delay model is well justified.

In a previous study, Engel and Hamilton [14], for instance, suggested the suitability of an HMM in modelling FX rates based on raw data taken as an arithmetic average of the bid-and-ask prices for the exchange rate (in dollars per unit of foreign currency) for the last day of the quarter, beginning with the third quarter of 1973 and ending with the first quarter of 1988. Cheung and Erlandsson [6] contributed to this idea by incorporating a rigorous Monte-Carlo approach to test for multi-regime dynamics of quarterly and monthly
data. In Yuan [32], FX rates are modelled using a static HMM with a special smoothing parameter. Since many successful trading strategies are dependent on fast FX rate movements (cf. Aldridge [2]) and given the peculiar characteristics of the FX data collected recently, we regard the zero-delay HMM to be more appropriate than the commonly used one-step-delay HMM. Evidence will also show that the new resulting dynamic filtered estimators for model parameters are able to capture better the trends of the FX rate process.

We adapt the technique of Elliott, et al. [11] in establishing expressions for zero-delay filters. Such an approach uses the change of probability measure method to evaluate filters under some ideal measure and relate the calculations back to the real-world measure through the Bayes’ theorem for conditional expectation. In the derivation of the filter expressions, we also incorporate the idea explored in Erwein et al. [15] concerning the dependencies of several processes evolving in parallel on the same discrete-time Markov chain. This approach performs well in delineating the amount of fluctuations of the FX caused by volatilities from those due to regime changes. In arguing that a multi-regime behaviour is present in a given data set, we employ a sequential algorithm following closely the testing procedure of Rodionov [25]. The procedure tailored to our HMM filtering-based estimation technique provides clear support for the existence of regimes. In establishing a statistically correct quantity of regimes needed to model the data, we use the Bayesian Information Criterion (BIC), which is a penalised version of the Akaike Information Criterion (AIC). The justification based on BIC is further reinforced by the CHull criterion.

We examine the fitting and prediction performance of our proposed model and estimation approach against popular models using FX rate data on two major currency pairs, namely the Japanese yen (JPY) versus US dollar (USD) and JPY versus UK sterling pound (GBP). The JPY is chosen as the base currency in our study owing to its reputation as one of the most volatile currencies in the world, and hence, regime-switching model is a reasonable model to investigate the characterisation of its dynamics.

To achieve our objectives, this paper is structured as follows. Section 2 describes the modelling framework and filtering recursions of a zero-delay model. The multi-dimensional filtering scheme is then considered and estimates of the parameters for the proposed model are determined using the EM algorithm. In section 3, we provide numerical evidence that favours the proposed model over other competing models with respect to goodness-of-fit measures and log likelihood-driven criteria. The paper culminates with some concluding remarks in section 4.
2 Model formulation

If $s_k$ is the FX rate at time $t_k$, we posit that over a very short-time period, i.e., over several seconds or minutes, the dynamics of its log returns evolve according to the equation

$$y_k := \ln \frac{s_k}{s_{k-1}} = \mu \Delta t_k + \sigma \sqrt{\Delta t_k} \epsilon_{t_k}.$$  \hspace{1cm} (1)

In equation (1), the parameters $\mu$ and $\sigma$ are the respective drift and volatility parameters under the real-world measure $P$, where $\epsilon_{t_k} \sim N(0,1)$, i.e., $\epsilon_{t_k}$ is a standard normal random variable and $\Delta t_k = t_k - t_{k-1}$.

Given the stochastic nature of the FX process, $\mu$ and $\sigma$ do not only change with time but also depend on the state of world according to a finite-state discrete-time Markov chain $x_{t_k}$ incorporating the over-all interaction of various factors. Therefore, equation (1) can be rewritten as

$$y_k := \ln \frac{s_k}{s_{k-1}} = \mu(x_{t_k}) \Delta t_k + \sigma(x_{t_k}) \sqrt{\Delta t_k} \epsilon_{t_k},$$ \hspace{1cm} (2)

$$x_{t_k} = Ax_{t_{k-1}} + v_{t_k},$$ \hspace{1cm} (3)

where $A$ is a transition probability matrix and $v_{t_k}$ is a martingale increment. This formulation is extended to the multivariate setting below. For brevity, we shall write $x_k$ instead of $x_{t_k}$ and similar notational adjustments will made to other variables.

Let $(\Omega, \mathcal{F}, P)$ be a probability space under which $x_k$ is a homogeneous Markov chain with a finite-state space. Without loss of generality, let $\Delta t_k = 1$. The dynamics of the $d-$dimensional observation process is then

$$y^g_k = \mu^g(x_k) + \sigma^g(x_k) \epsilon^g_{t_k}, \hspace{0.5cm} 1 \leq g \leq d,$$ \hspace{1cm} (4)

where $\{\epsilon^g_{t_k}\}$ is a sequence of independent standard Gaussian random variables. In particular, they are independent for each component of the row vector $y_k$. To considerably simplify the algebra involved in the filtering equations, we associate the state space of $x_k$ with the standard basis of $\mathbb{R}^N$, which is the set of unit vectors $e_h$, $h = 1, 2, \ldots, N$ and $e_h$ is a vector having 1 in its $h^{th}$ entry and 0 elsewhere. So, in equation (4), $\mu^g(x_k) = \langle \mu^g_k, x_k \rangle$ and $\sigma^g(x_k) = \langle \sigma^g_k, x_k \rangle$. The notation $\langle \cdot, \cdot \rangle$ is the usual scalar product and $\mu^g = (\mu^g(1), \mu^g(2), \ldots, \mu^g(N))^\top, \sigma^g = (\sigma^g(1), \sigma^g(2), \ldots, \sigma^g(N))^\top \in \mathbb{R}^N$, where $\top$ denotes the transpose of a vector.
To find the optimal estimate for the state of $x_k$, we use a change of measure approach described in [11]. We perform the calculations under a new, ideal measure $\tilde{P}$ under which all observations $y_k$'s are IID standard normal random variables and $y_k$'s are independent from $x_k$. Using a discrete-time version of the Girsanov's theorem, the real-world measure $P$ can be recovered via the Radon-Nikodým derivative

$$\Lambda_k := \left. \frac{dP}{d\tilde{P}} \right|_{\mathcal{F}_k} = \prod_{g=1}^{d} \prod_{l=1}^{k} \lambda_{gl}^g, \quad k \geq 1,$$

$$\Lambda_0 = 1 \quad \text{and} \quad \lambda_{gl}^g = \frac{\phi[\sigma^g(x_{l-1})^{-1}(y_{gl}^g - \mu^g(x_{l-1}))]}{\sigma^g(x_{l-1})\phi(y_{gl}^g)},$$

where $\phi(\cdot)$ is the density function of an $N(0,1)$ random variable. The filtration $\mathcal{F}_k$ is given by $\mathcal{F}_k := \mathcal{F}_k^x \vee \mathcal{F}_k^y$, where $\mathcal{F}_k^x$ and $\mathcal{F}_k^y$ are the filtrations generated by the Markov chain $x_k$ and observation process $y_k$, respectively.

We construct the filters for relevant quantities related to the Markov chain under the multivariate setting. Model parameter estimates will then be obtained in terms of these filters. Define the conditional probabilities of $x_k$ given $\mathcal{F}_y$ under $P$ as

$$p_k^i := P(x_k = e_h | \mathcal{F}_y) = E[\langle x_k, e_h \rangle | \mathcal{F}_y].$$

Write $\tilde{p}_k := (\tilde{p}_k(1), \tilde{p}_k(2), \ldots, \tilde{p}_k(N))^\top \in \mathbb{R}^N$ and so the optimal estimate for $x_k$ given the available information up to time $k$ is

$$\hat{p}_k = E[x_k | \mathcal{F}_y^u] = \tilde{E}[\Lambda_k x_k | \mathcal{F}_y^u] / \tilde{E}[\Lambda_k | \mathcal{F}_y^u]$$

by the Bayes' theorem for conditional expectation. Define $c_k := \tilde{E}[\Lambda_k x_k | \mathcal{F}_y^u]$ and note that

$$\sum_{h=1}^{N} \langle x_k, e_h \rangle = 1.$$

Thus,

$$\sum_{h=1}^{N} \langle \Lambda_k, e_h \rangle = \sum_{h=1}^{N} \langle \tilde{E}[\Lambda_k x_k | \mathcal{F}_y^u], e_h \rangle = \tilde{E} \left[ \Lambda_k \sum_{h=1}^{N} \langle x_k, e_h \rangle \mid \mathcal{F}_y^u \right] = \tilde{E}[\Lambda_k | \mathcal{F}_y^u].$$

The construction of $c_k$ along with equation (6) yields

$$\hat{p}_k = \frac{c_k}{\sum_{h=1}^{N} \langle c_k, e_h \rangle}.$$
In addition to the state’s optimal estimate, we also consider the following quantities:

\[
J^{sr}_{k+1} = \sum_{n=1}^{k+1} \langle x_{n-1}, e_r \rangle \langle x_n, e_s \rangle \tag{7}
\]

\[
O^{r}_{k+1} = \sum_{n=1}^{k+1} \langle x_{n-1}, e_r \rangle \tag{8}
\]

\[
T^r_{k+1}(f(y^g_{k+1})) = \sum_{n=1}^{k+1} \langle x_{n-1}, e_r \rangle f(y^g_n), \quad 1 \leq r \leq N, 1 \leq g \leq d \tag{9}
\]

with \( f(y^g_n) = y^g_n \) or \( f(y^g_n) = (y^g_n)^2 \).

Equation (7) counts the number of jumps from \( e_r \) to \( e_s \) at time \( t_{k+1} \). The amount of time up to \( t_{k+1} \) that \( x \) occupies state \( e_r \) is given by \( O^{r}_{k+1} \) in (8). The function \( T^r_{k+1}(f) \) in (9) is an auxiliary quantity that occurs in the estimation of model parameters.

For any process \( G_k \), we denote the conditional expectation under \( \tilde{P} \) of \( \Lambda_k G_k \) by \( \gamma(G_k) := \tilde{E}[\Lambda_k G_k | J_k^g] \). The adaptive filters enable the updating of model parameters that will incorporate past and current market conditions. Taking advantage of the semi-martingale representation of \( x_k \), we obtain the recursive filters for \( \gamma(J^i x_k) \), \( \gamma(O^i x_k) \) and \( \gamma(T^i (f(y^g_n)) x_k) \), and they are presented in the results that follow.

**Proposition 1:** Suppose \( A = (a_{ji}) \) is the transition matrix, \( a_i = A e_i \), and

\[
\Gamma^i = \Gamma^i(y^g_{k+1}) = \Gamma^i(y^1_{k+1}, \ldots, y^d_{k+1}) = \prod_{g=1}^{d} \phi \left( \frac{y^g_{k+1} - \mu^g(i)}{\sigma^g(i)} \right).
\]
Then
\[ \gamma(x_{k+1}) = \sum_{i,j=1}^{N} \langle \gamma(x_k), e_i \rangle a_{ji} \Gamma^j a_i \]  

\[ \gamma(J^i_{k+1} x_{k+1}) = \sum_{m,l=1}^{N} \langle \gamma(J^i_k x_k), e_m \rangle a_{lm} \Gamma^l a_m + \langle \gamma(x_k), e_i \rangle a_{ji} \Gamma^j a_i \]  

\[ \gamma(O^i_{k+1} x_{k+1}) = \sum_{m,l=1}^{N} \langle \gamma(O^i_k x_k), e_m \rangle a_{lm} \Gamma^l a_m + \sum_{l=1}^{N} \langle \gamma(x_k), e_i \rangle a_{li} \Gamma^l a_i \]  

\[ \gamma(T^i_{k+1} (f(g^i_{k+1})) x_{k+1}) = \sum_{m,l=1}^{N} \langle \gamma(T^i_k (f(g^i_k)) x_k), e_m \rangle a_{lm} \Gamma^l a_m \\ + f(g^i_{k+1}) \sum_{l=1}^{N} \langle \gamma(x_k), e_i \rangle a_{li} \Gamma^l a_i. \]  

**Proof:** See Appendix.

We apply the Expectation-Maximisation (EM) algorithm (cf. Dempster, et al. [10]) to find the optimal estimates of the model parameters. As the next result shows, the transition probabilities \( a_{ji} \), and the levels of the drift \( \mu(i) \) and volatility \( \sigma(i) \) are expressed in terms of the quantities in (7)-(9). The following results are stated without proof. It has to be noted that although we work under the zero-delay modelling set-up, the reasoning behind the proof of the following results are similar to those in the one-step delay model case (see Erlwein, et al. [15]), and hence it is easily reproducible.

**Proposition 2:** Consider a multivariate dataset \( y^g_1, y^g_2, \ldots, y^g_k, 1 \leq g \leq d \) observed up to time \( k \). The respective EM estimates of parameters \( \{a_{ji}, \mu^g(i), \sigma^g(i)\} \)

\[ \hat{a}_{ji} = \frac{\gamma(J^i_k)}{\gamma(O^i_k)} \]  

\[ \hat{\mu}^g(i) = \frac{\gamma(T^i_k (y^g_k))}{\gamma(O^i_k)} \]  

\[ \hat{\sigma}^g(i) = \sqrt{\frac{\gamma(T^i_k ((y^g_k)^2) - 2y^g_k \gamma(T^i_k (y^g_k)) + (y^g_k)^2 \gamma(O^i_k))}{\gamma(O^i_k)}} \]  

Now, in terms of the recursive filters given in Proposition 1, the optimal estimate of the
Markov chain is then
\[ \hat{p}_k = \frac{c_k}{\sum_{h=1}^N \langle c_k, e_h \rangle} = \frac{\gamma(x_k)}{\sum_{h=1}^N \langle \gamma(x_k), e_h \rangle}. \]

Furthermore, for any process \( G_k \), we have
\[ \gamma(G_k) = \gamma(G_k \langle x_k, 1 \rangle) = \langle \gamma(G_k x_k), 1 \rangle. \]

Hence, when \( G_k = \mathcal{J}_k, \mathcal{O}_k \) or \( \mathcal{T}_k \), equations (14)-(16) in Proposition 2 are fully determined, and Proposition 1 gives the dynamic updates of the model parameters.

3 Numerical case study

In this section, we analyse the performance of our proposed model and estimation method. An implementation to FX rate data compiled by Bloomberg is conducted. Two data sets are considered, namely, the JPY/USD and JPY/GBP, spanning the period of 0935 HRS, 06 July 2012 – 1840 HRS, 11 July 2012 with a five-minute interval between each observation. The JPY was selected as the base currency as it is known to possess random wild fluctuations in log returns or increments. This characteristic is deemed suitable for an HMM to capture.

The distributional structure of Japan’s natural resources and electricity supply combined with its geographical location implies that the JPY movement is mostly affected by internal factors (e.g., occurrence of natural disasters, supply and demand of raw materials, political climate, institutional policies, manufacturing sector stability, amongst others). Data sourced out from the IMF and World Bank in 2011 [30] reveals that Japan is the third largest economy in terms of GDP per capita and is also the third largest automobile manufacturing country. The Honda Motor Corporation’s 2012 Report explained that the variations in Honda’s stock price volatility are caused by various factors including fierce and increasing competition, short-term fluctuations in demand, changes in tariffs, import regulations and other taxes, and shortages of certain materials, amongst others. Since Honda is one of the biggest contributors to the national GDP, it is recognised that its stock price is highly correlated with the JPY currency. Therefore any major price movements in the value of major car companies have strong association with large price movements of JPY relative to the USD, GBP and other major currencies trading in the FX market. Apparently, the above-mentioned circumstances surrounding Japan could contribute to the volatile nature of its currency against major currencies such as the GBP and USD.
In this paper, we take FX rates as inputs to a model and assume that they contain latent information. These FX rates are “filtered” to extract the information essential in the characterisation of the “best” estimates of parameters for our proposed model. These “best” estimates can in turn be utilised for pricing derivatives, risk management and forecasting over a short horizon. From the actual data, we generate the observed log returns $y^g_k$ for $g = 1, 2$ ($1 \equiv$ JPY/GBP, $2 \equiv$ JPY/USD), $k = 1, \ldots, 1000$. It is assumed that the log return process is decomposed into two parts: the mean $\mu$ and volatility $\sigma$, which are both driven by an unobserved HMM $x_k$. The goal is to estimate $\mu, \sigma, x_k$ and the matrix $A$ in an optimal way.

### 3.1 Regime-switching assumption in the data

Before starting the model implementation, we first validate the important modelling assumption of multi-regime behaviour in our data. In addition to simple visual check of spikes, which shows instances of regime shifts in the mean and volatility of log returns, there are a few formal statistical tests described in the literature that could be used to determine the presence of regime switches; see for example Rodinov [26]. We choose a sequential algorithm adapted from the works of Rodionov [25] and apply the testing procedure with a 99.9 percent confidence level on every data point. This test essentially checks if the difference between the mean values of two consecutive regimes are statistically significant according to the principles of the Student’s $t$-test. This test of significance for each of the mean and variance levels is a two-step test. In the first step, the size of a sample window encompassing the first few data points is chosen, and the mean and variance of the data in this sample window are calculated. These statistics (sample mean and variance) are employed in establishing a “test interval”. In the second step, data points that belong outside the “test interval” are compared to a computed regime shift index (RSI). An inference of whether a possible regime switch occurred is made on the basis of this RSI.

For the detection of regime shifts in the variance of the log returns, the data points are first normalised, i.e., the data is transformed to have a zero mean. The test for the shifts in variance is conducted in a manner similar to that of the regime-shift determination for the mean of the log returns. Needless to say, the size of the sample window used to construct the RSI and the “test interval” have direct impact on the detection of mean and variance shifts of the underlying process; see Rodionov [26]. For the purpose of providing support to the Markovian assumption in this paper, we heuristically find that a window of size 6 data points (30-minute interval) is adequate.
Figures 1 and 2 depict episodes of regime switches in the mean of log returns. Regime changes occur at times where the vertical bars are drawn. The orientation of the vertical bars (above or below the horizontal axis) indicate whether the regime change in the mean that occurred was going up or down. A salient point to notice is that within the time interval [09:15, 11:15] on 11 July 2012 (right-end portion of Figures 1 and 2), there were noticeable simultaneous changes in the mean of FX rates’ log returns for both currency pairs. This strongly supports our assumption that the joint FX rates’ behaviour in the mean is driven by the same Markov chain.

As portrayed in Figures 3 and 4, there is also evidence of regime switches in volatility happening closely in tandem, at 01:40 on 10 July 2012 and at 01:45 on 10 July 2012 (middle section of Figures 3 and 4) for the JPY/GBR and JPY/USD, respectively. This instance serves another indication that a multidimensional HMM may work well to model our bivariate FX rate data.

In general, the result of any statistical test in establishing the presence of regime switches is a function of confidence level, size of “test interval”, and the data possessing extreme
Figure 2: Illustrating the occurrence of regime switches in the mean of log returns with a 99.9 % confidence level for JPY/USD covering the period 09:35, 10 July – 19:20, 11 July 2012.

Figure 3: Illustrating the occurrence of regime switches in the volatility with a 99.9 % confidence level for JPY/GBR covering the period 09:35, 10 July – 19:20, 11 July 2012.
volatility movement (or the lack of it). It has to be noted though that when the data is recorded with very short frequency, such as in our case, it may not be that straightforward to differentiate between the shifts in the volatility parameter and movements of volatility itself. Sudden increases in the variance are normally adjudged as regime switches although such assessment result is still test-specific. In certain sequential tests, some spikes in the data are simply treated as outliers even though these may be considered as triggers for a regime switch in other detection tests for regime switching. Relying only on Rodionov’s regime-switching test procedure [25], our aim is not to pursue other test procedures in detecting multi-regime dynamics but rather on showing enough evidence that our HMM approach is suitable for the FX-rate data sets that we are examining in this paper.

In the next subsections, we describe several algorithms to model the distinguishing distributional characteristics of FX rates, propose approaches for model calibration, and perform model validation via goodness-of-fit tests and forecasting exercise.

3.2 Benchmarking the zero-delay HMM

We evaluate the performance of several competing HMMs based on several statistical criteria. We also make comparisons, through the assessment of fit and prediction performance, of our
proposed modelling approach with the widely used generalised autoregressive conditionally heteroskedastic (GARCH) [4] and the RW model. In addition to the zero-delay filters developed in section 2, we consider two more HMM filtering algorithms, viz.: the static estimation approach presented in Hamilton [19] and the dynamic filters for the one-step delay model derived in Erlwein, et al. [15]. The models included in the benchmarking of our zero-delay HMM are as follows.

1. Static independent log-normal model (ILN). It is assumed that the log returns process is lognormally distributed with constant mean and variance, i.e.,

\[ y_k = \mu + \sigma \epsilon_k, \text{ where } \epsilon_k \sim N(0,1), \]

where \( N(0,1) \) stands for a standard normal random variable.

2. Static regime-switching one-step delay log-normal model with \( N \) regimes (RSLN-N). This extends the ILN model to multi regimes. The mean and variance of log returns are not constant. Rather, they are driven by discrete-time Markov chain with finite state space. Thus,

\[ y_{k+1} = \mu_S(x_k) + \sigma_S(x_k)\epsilon_{k+1}, \text{ where } \epsilon_{k+1} \sim N(0,1). \]

Under this model, a static estimation is performed by processing the entire data set, and only one set of model estimates are obtained. The notation \( \mu_S \) and \( \sigma_S \) signify estimates calculated from this static estimation procedure. Similar explanation goes for the subscript notation in the next two models.

3. Dynamic regime-switching one-step delay log-normal model with \( N \) regimes (DRSLN-N). Under the DRSLN-N,

\[ y_{k+1} = \mu_D(x_k) + \sigma_D(x_k)\epsilon_{k+1}, \text{ where } \epsilon_{k+1} \sim N(0,1). \]

This model features a dynamic estimation procedure that produces a sequence of parameter estimates evolving through filtering-algorithm steps.

4. Dynamic regime-switching zero-delay log-normal model with \( N \) regimes (ZDRSLN-N). This model has all the properties of the DRSLN-N with the assumption that the distribution of \( y_k \) depends without delay on \( x_k \), i.e., a zero-time lag dependence. That is,

\[ y_k = \mu_Z(x_k) + \sigma_Z(x_k)\epsilon_k, \text{ where } \epsilon_k \sim N(0,1). \]
Remark 1: We recognise that we do not explicitly model the correlation amongst the currency pairs. Nevertheless, the currency pairs are governed by the same HMM, and are therefore implicitly correlated. Filters with appropriate correlation structure for the various white noise governing the log returns of currency pairs will most likely be better. Notwithstanding the limitation of the proposed model, this research investigation can be treated as a lower bound for the credibility of a study that takes explicitly into account correlated white noise (or correlated Brownian motions in the case of continuous-time modelling set up).

Remark 2: Under the one-step delay model (3rd model), the reaction to \( x_k \) is not instantaneous. With the zero-delay model (4th model), there is no delay in the reaction to \( x_k \). Under the latter framework, this is tantamount to relabelling the observation process. So, in the calculation of filters under the zero-delay model in Proposition 1, if \( \{ F^y_t \} \) is the complete filtration generated by the observations then \( F^y_t = F^y_{t+1} \), where \( F^y_{t+1} \) is the filtration generated by the observations under the one-step delay model. This implies that \( \tilde{E}[x_{k+1}|F^y_t] = \tilde{E}[x_{k+1}|F^y_{t+1}] \). See further the Appendix.

3.3 Numerical implementation

3.3.1 Optimal number of states in HMM

Our choice of the optimal number of states for the HMM is based on the maximisation of log likelihood with a penalty for increasing the complexity of the model, i.e., adding more parameters. This is further complemented by a few error analyses choosing the number of states that yield the lowest value for a given error metric. Such evaluation of log likelihood-maximisation with penalty criteria and error metrics assume that we are able to successfully obtain models’ parameter estimates. This is further discussed in subsection 3.4.

3.3.2 Initial parameter estimates

Various approaches may be used to find initial parameters in the implementation of filtering algorithms. Least-square method (see, for example, Erlwein and Mamon [17]), likelihood maximisation if it can be accomplished (cf. Date and Ponomareva [9]) and use of first two sample moments are the most popular.

In our case, we first assume that the time-series data are stationary. Then under the one-
state setting, we find the estimates \( \hat{\mu}^g \) and \( \hat{\sigma}^g \) using (i) the sample moments (sample mean and sample standard deviation) and (ii) maximum likelihood. For both sample moment-based and maximum-likelihood methods, we identify the maximum and minimum of the data set and assign the starting values as \( \mu^g(1) = \frac{\text{max}(y^g_k) - \hat{\mu}^g}{2} \) and \( \mu^g(2) = \frac{\text{min}(y^g_k) + \hat{\mu}^g}{2} \). This procedure may be generalised for the \( N \)-regime case by “spreading” starting values of \( \mu^g(i) \) evenly around \( \hat{\mu}^g \). Initial values of \( \sigma^g(i) \) are typically taken to be \( \hat{\sigma}^g \) in either regime and as the data are filtered, two volatility estimates emerge, which may or may not be different for each of the two regimes. The EM algorithm’s rate of convergence is very fast so long as the starting parameters are close to the “true” parameter values. But as these “true” values are unknown to begin with, we rely on data processing’s stability and speed with the aid of the filtering equations attaining convergence in each algorithm step.

We choose \( 1/N \) for the initial value of each entry in the transition probability matrix, and this yields a stable convergence. An alternative approach described in Hamilton [19] and Hardy [20] is useful for the two-state RSLN model. Such approach works very fast because it is computationally easy to maximise a smooth function (exponential function in this case) over the six parameters. It is worth noting that whilst the results of sample moment-based and likelihood maximisation approaches in obtaining initial values (Tables 1 and 2) are different, they still make the filtering algorithms converge and are able to replicate identical dynamics of \( a_{ij} \), and \( \mu(i) \) and \( \sigma(i) \) for each pair of currencies.

By construction, the dynamic filters, with a single underlying unobserved HMM, take into account the behaviour of two FX rates jointly. Hence, the starting values for \( a_{12} \) and \( a_{21} \)
were taken as the average of the corresponding values in Tables 1 and 2 in each method. Although the likelihood maximisation (ML) is an ideal method to initialise parameter values given its statistical formality and grounding, its implementation in practice may pose one insurmountable challenge. The maximum likelihood value found using numerical methods may not necessarily be a global maximum, and proving that is not clear-cut if not impossible. Such a problem akin to ML estimation becomes more pronounced under HMM with four or more states where computational time increases with dimensionality.

We further remark that instead of employing the maximal and minimal data values in the above assignment of initial value estimation, appropriate quantiles could be considered instead. This is because outliers can substantially affect this procedure. In spite of the fact that the use of sample moments to select initial estimates is rather heuristics, they enable the filtering algorithms to yield parameter estimates that have identical evolution to those produced by the ML method in the 2- and 3-state settings.

3.3.3 Filters, data processing and estimation

In the implementation of the ILN and RSLN-N models, we employ the estimation method proposed in Hardy [20]. This method, however, was formulated not taking into account the joint dynamics of two FX rates. Thus, parameter estimates are obtained by applying the method separately to each data set. This implies a simplifying assumption that each FX rate data series is independent of each other. The same approach and assumption are taken in the implementation of the GARCH model, where we utilise the standard-fitting and forecasting procedures from Matlab’s Econometrics Toolbox.

The dynamic filtering algorithms put forward in this paper on the other hand were designed to work on two-dimensional data of FX rates collected at same time points. Dynamic filters process a moving window of several univariate or multivariate data points. This gives rise to certain filtered quantities related to the Markov chain that subsequently produce a sequence of model parameter estimates using Proposition 2. The processing of a window of data points is termed as one complete pass or algorithm step. The initial values are used to generate parameter estimates in the first algorithm step, which in turn serve as the initial values to obtain the parameter estimates after the second pass through the data, and this process continues until the last algorithm step. The size of the moving window is chosen based on some criterion, which is the maximisation of the log-likelihood function in our case.
Our findings show that the numerical values of log-likelihood function are not monotonic with respect to the window size, and the models with the best fit (in the context of penalty-based information criterion) do not always produce the best out-of-sample forecasts. This issue is linked to an associated instability of the log-likelihood maximisation due to various factors, most notably the numerical error from the division of quantities close to zero. The selection of the appropriate window size in our case is made on the basis of (i) maximising log-likelihood values and (ii) reproducing the nonlinear evolution of the probabilities $\hat{p}$; refer to equation (5). We display the plots of $\hat{p}$ in Figures 5 and 6 through the algorithm steps. The behaviour of $\hat{p}$ under the zero-delay HMM (Figure 6) looks very similar to the behaviour of $\hat{p}$ under the one-step delay model (Figure 5).

From the above window-size selection guidelines, our numerical experiments demonstrate that the most suitable moving window must contain 7 bivariate data points, which correspond to a 35-minute interval. A wider moving window smooths out the prediction process. That is, with more data points in the moving window, the graphs of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{p}$ would look like straight lines, which are contrary to the essence of nonlinear dynamics of parameters. Moreover, a large sample window decreases the quality of the out-of-sample forecasts as FX rate dynamics are not also fully captured. On the other hand, smaller window size magnifies the effect of volatility, which markedly dominates the drift component. Consequently, this results to large numerical errors making the predictions meaningless.

![Figure 5: Evolution of $\hat{p}$ under the one-step delay HMM](image)

Figure 5: Evolution of $\hat{p}$ under the one-step delay HMM


3.4 Comparison of numerical results

In Tables 3 and 4, we present various statistical measures to gauge the goodness of fit of the proposed model in comparison with other popular existing models. We evaluate the Akaike information [1] and Bayesian information [27] criteria, abbreviated as AIC and BIC, respectively, that highlights the maximisation of log likelihood and penalty for model complexity. The model with the highest AIC and BIC is most preferred. The AIC and BIC metrics are given by

\[
AIC = \ln L - b
\]

\[
BIC = \ln L - 0.5k \ln m,
\]

where \(L\) is a likelihood or log-likelihood function, \(m\) is the number of observations and \(b\) is the number of parameters in a model.

Furthermore, we also evaluate the fitting performance of the models by analysing the errors when prediction values are compared to actual values. With FX rates treated as a 2-dimensional observation process, the one-step ahead forecasts using equation (1) is

\[
E[y_{k+1}^g|\mathcal{F}_k^g] = y_k^g \sum_{j=1}^N \langle \hat{\mathbf{x}}_k, e_j \rangle \exp \left( \hat{\mu}_g(j) + \frac{\hat{\sigma}_g(j)^2}{2} \right), \tag{17}
\]

where \(\hat{\mathbf{x}}_k\) is the estimate for the unconditional distribution of the Markov chain.
<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>BIC</th>
<th>AIC</th>
<th>RMSE</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>$2.6510 \times 10^{-4}$</td>
<td>n/a</td>
</tr>
<tr>
<td>ILN</td>
<td>6603.99</td>
<td>6597.08</td>
<td>6601.99</td>
<td>n/a</td>
<td>2</td>
</tr>
<tr>
<td>RSLN-2</td>
<td>6680.68</td>
<td>6659.96</td>
<td>6674.68</td>
<td>n/a</td>
<td>6</td>
</tr>
<tr>
<td>RSLN-3</td>
<td>6702.21</td>
<td>6660.76</td>
<td>6690.21</td>
<td>n/a</td>
<td>12</td>
</tr>
<tr>
<td>DILN</td>
<td>6781.19</td>
<td>6774.28</td>
<td>6779.19</td>
<td>$8.0811 \times 10^{-6}$</td>
<td>2</td>
</tr>
<tr>
<td>DRSLN-2</td>
<td>6877.03</td>
<td>6856.30</td>
<td>6871.03</td>
<td>$6.5976 \times 10^{-6}$</td>
<td>6</td>
</tr>
<tr>
<td>DRSLN-3</td>
<td><strong>6890.12</strong></td>
<td>6848.67</td>
<td><strong>6878.12</strong></td>
<td>$6.5897 \times 10^{-6}$</td>
<td>12</td>
</tr>
<tr>
<td>ZDILN</td>
<td>6804.96</td>
<td>6798.05</td>
<td>6802.96</td>
<td>$8.2164 \times 10^{-6}$</td>
<td>2</td>
</tr>
<tr>
<td>ZDRSLN-2</td>
<td>6862.17</td>
<td>6841.44</td>
<td>6856.17</td>
<td><strong>6.5533 \times 10^{-6}</strong></td>
<td>6</td>
</tr>
<tr>
<td>ZDRSLN-3</td>
<td><strong>6878.29</strong></td>
<td>6836.84</td>
<td><strong>6866.29</strong></td>
<td>$6.5832 \times 10^{-6}$</td>
<td>12</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>6667.86</td>
<td>6654.05</td>
<td>6663.86</td>
<td>$2.4720 \times 10^{-4}$</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3: Results of likelihood- and error-based fitting measures covering JPY/GBP data collected between 09:35, 06 July 2012 and 18:40, 11 July 2012

One way to examine the model’s goodness of fit is to use the root-mean-square-error (RMSE) metric given by

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{k=1}^{m} (s_k^g - \hat{s}_k^g)^2},$$

where $s_k^g$, $\hat{s}_k^g$ and $m$ are the actual value, predicted value and length of the data series, respectively. The model producing the lowest RMSE is most preferred. For modelling multivariate data, the corresponding appropriate RMSE metric is given, for example, in Date, et al. [7].

**Remark 3:** We do not consider the RMSE for multivariate modelling here because the implementation of ILN and GARCH models was performed using univariate-based algorithms. Thus, to make our model comparison valid, the univariate version of the RMSE is employed on each time series of FX rates.

Under both the AIC and BIC, our empirical results reveal that for the JPY/GBP data, the multi-regime models with dynamic filters (one-step and zero-delay models) significantly outperform other competing models. From Table 3, the DRSLN-N and ZDRSLN-N models fit the data very well judging from the chosen criteria. The ZDRSLN-N models are marginally better than the DRSLN-N models as the former have slightly smaller RMSEs than the latter. In the case of the JPY/USD data, the RW model outperforms all other models. This because the log increments of the JPY/USD spot-rate process are so small...
<table>
<thead>
<tr>
<th>Model</th>
<th>Log likelihood</th>
<th>BIC</th>
<th>AIC</th>
<th>RMSE</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td><strong>2.883 × 10^{-6}</strong></td>
<td>n/a</td>
</tr>
<tr>
<td>ILN</td>
<td>6999.81</td>
<td>6992.90</td>
<td>6997.81</td>
<td>n/a</td>
<td>2</td>
</tr>
<tr>
<td>RSLN-2</td>
<td>7065.49</td>
<td>7044.76</td>
<td>7059.48</td>
<td>n/a</td>
<td>6</td>
</tr>
<tr>
<td>RSLN-3</td>
<td>7078.31</td>
<td>7036.87</td>
<td>7066.31</td>
<td>n/a</td>
<td>12</td>
</tr>
<tr>
<td>DILN</td>
<td>7102.50</td>
<td>7095.60</td>
<td>7100.50</td>
<td><strong>4.0598 × 10^{-6}</strong></td>
<td>2</td>
</tr>
<tr>
<td>DRSLN-2</td>
<td>7230.05</td>
<td><strong>7209.32</strong></td>
<td><strong>7224.05</strong></td>
<td><strong>4.0021 × 10^{-6}</strong></td>
<td>6</td>
</tr>
<tr>
<td>DRSLN-3</td>
<td>7235.12</td>
<td>7193.67</td>
<td>7223.12</td>
<td><strong>7.9566 × 10^{-6}</strong></td>
<td>12</td>
</tr>
<tr>
<td>ZDILN</td>
<td>7200.95</td>
<td>7193.62</td>
<td>7198.53</td>
<td><strong>4.0678 × 10^{-6}</strong></td>
<td>2</td>
</tr>
<tr>
<td>ZDRSLN-2</td>
<td><strong>7239.74</strong></td>
<td><strong>7218.35</strong></td>
<td><strong>7233.07</strong></td>
<td><strong>3.1228 × 10^{-6}</strong></td>
<td>6</td>
</tr>
<tr>
<td>ZDRSLN-3</td>
<td><strong>7244.13</strong></td>
<td>7202.69</td>
<td>7232.13</td>
<td><strong>6.5533 × 10^{-6}</strong></td>
<td>12</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>7027.22</td>
<td>7013.41</td>
<td>7023.22</td>
<td><strong>2.1801 × 10^{-4}</strong></td>
<td>4</td>
</tr>
</tbody>
</table>

Table 4: Results of loglikelihood- and error-based fitting measures covering JPY/USD data collected between 09:35, 06 July 2012 and 18:40, 11 July 2012

that numerical errors muddled the calculation of the RMSE. But, under the log likelihood-driven criteria (AIC and BIC), it is clear that the ZDRSLN-2 won this comparison race of FX-rate models. The ZDRSLN-2 also ranked second in terms of the RMSE criterion. Thus, we still place a great deal of confidence in the performance of the ZDRSLN-2 model.

For both data sets, the fit of the ILN and GARCH(1,1) models is poor relative to those of the other models across all criteria. We observe significant improvement in log-likelihood values of the dynamic models over those of the static RSLN.

### 3.5 CHull criterion

In any modelling endeavour, the main goal is to select a model that optimally balances model’s goodness of fit/misfit and complexity/simplicity. To reinforce the AIC, BIC and RMSE criteria in our analyses above, we consider the model-fitting criterion called CHull (based on the concept of convex hull) in the context of mixtures of factor analyses developed by Ceulmans and Kiers [5]. We tailor this criterion to our regime-switching methodology aided by the guidelines provided in Buteel, et al. [3]. A generic but succinct synopsis of CHull is described in Wilderjans, et al. [31] comprising of two main steps: (i) determining the convex hull of the fit-measure-versus-complexity-measure plot of the models under comparison and (ii) identifying the model on the boundary of the convex hull such that in-
creasing complexity (i.e., adding more parameters) has only a small effect on the fit measure, whereas lowering complexity (e.g., having less parameters) changes the goodness of fit/misfit significantly.

In our CHull implementation, we keep track and compare the changes in the log-likelihood (simple fitting measure) when the dimensions of the models (measure of complexity) are increased or decreased. In particular, the change of the log-likelihood of the models on the convex hull is given by

\[
\frac{\left( \log L_n - \log L_{n-1} \right)}{f_n - f_{n-1}} / \left( \frac{\log L_{n+1}^* - \log L_n^*}{f_{n+1} - f_n} \right),
\]

where \( L_n \) is the log-likelihood value of the \( n^{th} \) model, with models being ordered according to the number of free parameters. Here, \( f_n \) stands for the number of free parameters in the \( n^{th} \) model. In our case, \( n = 2, 4, 6 \) and 12. In summary, the above procedure can be described as simple as plotting the values of the log-likelihood for all models versus the number of free parameters and choosing the points where the log-likelihood values are the biggest and building a convex hull using these chosen points. We pick the model that produces the greatest statistic value in equation (18).

The points corresponding to the models lying on the convex hull are connected by a line. From Figures 7 and 8, the models chosen by the BIC is on the upper boundary of the convex hull. As seen in Figures 7 and 8, the models whose log-likelihood values lie on the convex hull are the same models which are selected in accordance with the BIC criterion. This provides extra support on the appropriateness of the dynamic regime-switching models (especially the zero-delay model) in capturing the trend of the bivariate FX data.

3.6 Parameter estimation results and further model validation

3.6.1 Dynamics of parameter estimates

Figures 9 - 13 depict the estimation outputs of the filtering experiment for the zero-delay 2-state model. The plots for the outputs of the one-step delay 2-state model look and behave in a similar manner, and are therefore omitted. The estimated drift and volatility dynamics clearly show the multi-regime pattern of the FX-rate process. This empirically illustrates the unpredictability of the parameters’ behaviour in the models under consideration. Such unpredictability is consistent with Nailiu and King’s findings [24] regarding extreme difficulty in estimating parameters and modelling the FX-rate process in general.
Figure 7: Chull for the JPY/GBP data

Figure 8: Chull for the JPY/USD data
3.6.2 Validating the white-noise assumption

A Q-Q plot is generated and shown in Figure 14 to validate the initial assumption of an $N(0, 1)$ noise in our multivariate modelling set up. Note that this is a multivariate Q-Q plot built based on Liang and Ng [21].
3.7 Frequent trading and the ZDRSLN-N modelling set-up

Following the discussion in subsection 3.4, we focus our comparison on several dynamic regime-switching models if no delay in the reaction time of \( y_k \) to \( x_k \) in frequent trading is a pertinent assumption. We implemented the ZDRSLN-N and DRSLN-N filters using the most popular computing platforms in the industry. These include the Microsoft Visual
Figure 13: Evolution of transition probabilities under the zero-delay 2-state HMM

Figure 14: Q-Q plot for the bivariate FX rate data

C++ and Visual Basic for Applications (VBA) in Microsoft Excel. The codes written in Excel are more instructive than those in C++, and without a doubt they are more useful for practitioners keen on implementation. However, it takes a very long time to generate results
using Excel. It would only take 8.29 seconds to run the filtering together with the parameter estimation codes and produce the needed numerical outputs (e.g., plots of parameters' evolution) in C++ whilst it would take 7 minutes to perform the same tasks in VBA.

In the financial industry the design of many algorithms is compatible with Excel workbooks. We therefore implemented a C++ code to interface with Excel sheets. This endeavour imposes additional limitation on the performance of the code as Excel is used not only as a database, but also as a graph-plotting engine. If C++ is run for the purpose of calculating the parameter values alone, the computing time involved is only 2 seconds. This shows that C++ implementation is convincingly more efficient than VBA.

Given the appropriateness and seeming dependability of the zero-delay models in outperforming other models under a frequent trading environment, we now concentrate on the analysis of our complete JPY/GBP and JPY/USD FX bivariate data set. That is, all data points that were recorded, some of which occurred in time intervals of less than 5 minutes, are considered. Thus, we wish to examine time series data with a much higher frequency than the ones studied in the previous subsections. The period covered is the same as that in the previous exercise, i.e., from 09:35, 06 July 2012 to 18:40, 11 July 2012.

Even with higher frequency, the trades on our two currency pairs are still at irregular intervals and do not occur simultaneously. Moreover, they do not necessarily happen every second or even within a minute. Sometimes, there were several trades in a one-minute interval, at other times, there was none at all. So, from the original data sets, we construct two new “derived” FX data sets where data points are “observed” at synchronised time points (but finer intervals of uniform lengths) to be able to apply our filtering algorithms in conjunction with our zero-delay models.

The new-data-generation procedure combines simple heuristics of data imputation and smoothing in the following way. The time series data is binned into two-minute intervals. With the two-minute interval binning, we ended up with 35 and 63 empty bins for JPY/USD and JPY/GBP currency pairs, respectively. The bin length is justified on the basis of minimising the number of empty bins whilst attempting to retain the effect of the model's drift. If the bin size becomes smaller than two minutes, the filter is unable to recognise the fluctuations coming from the drift component. If the bin size is larger than two minutes, the rationality and benefits of the zero-delay models disappear. Expectedly, we found that increasing the bin size favours the use of the one-step delay models.
The data values are averaged in each two-minute bin interval and this average represents the data point of the modified new data set assigned at the beginning of the bin interval. If there were no trades in a given time interval, the previous average value is used, i.e., it becomes a new trading value for the next empty bin.

We concentrate on the performance analysis of the adjudged best two models from the previous subsection given the newly obtained 2-minute time-series bivariate data set. We include the RW model in the analysis. The resulting goodness-of-fit statistics from the application of both the dynamic one-step and zero-step delay filtering algorithms on the new data set are depicted in Tables 5 and 6. The results ostensibly favours the ZDRSLN-2 model in terms of BIC and AIC. The RMSEs show that the ZDRSLN-2 model is able to outperform significantly the RW model with approximately 43% decrease in RMSE for the JPY/USD data and approximately 18% decrease in RMSE for the JPY/GBP data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log likelihood</th>
<th>BIC</th>
<th>AIC</th>
<th>RMSE</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>$3.692 \times 10^{-4}$</td>
<td>n/a</td>
</tr>
<tr>
<td>DRSLN-2</td>
<td>6671.307</td>
<td>6650.584</td>
<td>6665.307</td>
<td>$8.472 \times 10^{-4}$</td>
<td>6</td>
</tr>
<tr>
<td>ZDRSLN-2</td>
<td><strong>6728.407</strong></td>
<td><strong>6707.684</strong></td>
<td><strong>6722.407</strong></td>
<td><strong>3.031 \times 10^{-4}</strong></td>
<td>6</td>
</tr>
</tbody>
</table>

Table 5: Loglikelihood- and error-based goodness-of-fit measures for the new JPY/GBP data set with 2-minute frequency covering the period 09:35, 06 July 2012 - 18:40, 11 July 2012

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>BIC</th>
<th>AIC</th>
<th>RMSE</th>
<th>Number of parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>RW</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>$1.010 \times 10^{-5}$</td>
<td>n/a</td>
</tr>
<tr>
<td>DRSLN-2</td>
<td>7115.221</td>
<td>7102.19</td>
<td>7109.221</td>
<td>$1.323 \times 10^{-5}$</td>
<td>6</td>
</tr>
<tr>
<td>ZDRSLN-2</td>
<td><strong>7204.691</strong></td>
<td><strong>7191.66</strong></td>
<td><strong>7198.691</strong></td>
<td><strong>5.780 \times 10^{-6}</strong></td>
<td>6</td>
</tr>
</tbody>
</table>

Table 6: Loglikelihood- and error-based goodness-of-fit measures for the new JPY/USD data set with 2-minute-frequency covering the period 09:35, 06 July 2012 - 18:40, 11 July 2012
4 Conclusion

In this paper, we put forward an alternative approach to jointly model the behaviour of high-frequency multivariate FX rates using an HMM. New filtering recursive equations are derived assuming a zero-delay modelling paradigm. These recursions yield a self-calibrating multi-dimensional model that practitioners may employ for various financial modelling endeavours. To demonstrate the applicability and performance of our proposed model together with its parameter estimation methodology, we consider its implementation on the JPY/USD and JPY/GBP FX rates. A comparative analysis was conducted examining various HMM competing models with increasing complexity in terms of regime dimension and lag order as well as commonly used models for simple benchmarking. Our selection criteria in assessing model performance are in adherence to balancing the goodness of fit via log-likelihood maximisation and model complexity penalty (AIC, BIC, Chull). This was complemented by evaluating the model’s RMSE and choosing the model with the least forecasting error.

Whilst we tend to emphasise model development for data under the HFT’s framework, this investigation can be viewed as bridging the gap between the usual modelling of low-frequency data (daily, weekly monthly or quarterly) and the modelling of the currently emerging data sets resulting from rapid trading (minutes or seconds). More specifically, some of the challenges unique to high-frequency data were elaborated, and certain ways to formally or heuristically rectify them were given. Common issues in the implementation of our proposed approach were detailed and addressed.

The empirical evidence of our extensive study using various filtering estimation algorithms in conjunction with model validation diagnostics shows that the proposed dynamic models, in particular the ZDRSLN-N, outperform other competing models (GARCH, RSLN, DRSLN). These zero-delay models would be beneficial in forecasting FX rates to aid practitioners in setting their trading positions. Armed with a better model in terms of fitting high-frequency data, a natural direction of this study would be the pricing of FX derivatives as well as testing the accuracy of risk measures for portfolios with exposures to FX-rate movements.

Two extensions could be pursued to probably improve further the model performance and make the filtering algorithms more flexible. (i) Smoothers could be incorporated that will allow parameters to change quickly without having the need to filter a quite large subset of data. This procedure will decrease the RMSEs and simultaneously shorten the time of parameter estimation. (ii) A correlation structure between the white noise drivers of the
data could be introduced. This would lead to a collection of enhanced filtering algorithms 
but most likely, a set of new challenges in the numerical implementation have to be tackled.

Finally, an alternative approach could be explored by considering instead the modelling 
of the price dynamics of currency or FX futures. From the estimates of FX futures prices, 
forecasts of FX spot rates and other parameters of interest can be recovered. This method-
ology is similar to the idea worked out in Date, et al. [8] in capturing the evolution of 
arbitrage-free futures prices on commodities. The methodology of modelling FX futures 
prices directly is ideal but only if backing out FX rates for one particular currency pair. 
Nonetheless, when a joint evolution of FX rates is needed, the approach in this paper is still 
deemed more appropriate and relevant.
References


Appendix

Proof of recursive filters in Proposition 1

We provide the derivation of the filters given in Proposition 1 of section 2.

Filter for the state of $x_k$

$$
\gamma(x_{k+1}) = \tilde{E}[\Lambda_{k+1}x_{k+1}|F_{k+1}^y] = \sum_{j=1}^{N} \tilde{E}[\Pi x_k \Lambda_k \langle x_{k+1}, e_j \rangle | F_{k+1}^y] \Gamma^j 
$$

(19)

$$
= \sum_{j=1}^{N} \tilde{E}[\Pi x_k \Lambda_k \langle \Pi x_k + v_{k+1}, e_j \rangle | F_{k+1}^y] \Gamma^j 
$$

$$
= \sum_{j=1}^{N} \tilde{E}[\Pi x_k \Lambda_k \langle \Pi x_k, e_j \rangle | F_{k}^y] \Gamma^j.
$$

By noting that $\sum_{j=1}^{N} \langle x_k, e_i \rangle = 1$, it follows that

$$
\gamma(x_{k+1}) = \sum_{i,j=1}^{N} \langle \gamma(x_k), e_i \rangle a_{ji} \Gamma^j a_j.
$$

Filter for the jump process $\mathcal{J}$

$$
\gamma(\mathcal{J}^{ji}_{k+1} x_{k+1}) = \tilde{E}[\Lambda_{k+1} \mathcal{J}^{ji}_{k+1} x_{k+1}|F_{k+1}^y] 
$$

(20)

$$
= \sum_{l=1}^{N} \tilde{E}[\Lambda_k \Pi x_k \langle x_{k+1}, e_l \rangle \left( \mathcal{J}_k^{ji} + \langle x_k, e_i \rangle \langle x_{k+1}, e_j \rangle \right) | F_{k+1}^y] \Gamma^l 
$$

$$
= \sum_{l=1}^{N} \tilde{E}[\Lambda_k \Pi x_k \langle x_{k+1}, e_l \rangle \left( \mathcal{J}_k^{ji} + \langle x_k, e_i \rangle \langle x_{k+1}, e_j \rangle \right) | F_{k+1}^y] \Gamma^l 
$$

$$
= \sum_{l=1}^{N} \tilde{E}[\Lambda_k \Pi x_k \langle x_{k+1}, e_l \rangle \left( \mathcal{J}_k^{ji} + \langle x_k, e_i \rangle \langle x_{k+1}, e_j \rangle \right) | F_{k+1}^y] \Gamma^l 
$$

$$
= \sum_{l=1}^{N} \tilde{E}[\Pi \langle x_k, e_l \rangle \Lambda_k x_k \mathcal{J}_k^{ji} | F_{k}^y] \Gamma^l + \tilde{E}[\Lambda_k \Pi x_k \langle x_k, e_i \rangle \langle x_k, e_i \rangle | F_{k}^y] \Gamma^j 
$$

$$
= \sum_{m,l=1}^{N} \langle \gamma(\mathcal{J}_k^{ji} x_k), e_m \rangle a_{lm} \Gamma^l a_m + \tilde{E}[\Pi x_k \Lambda_k \langle x_k, e_j \rangle | F_{k}^y] \Gamma^j 
$$

$$
= \sum_{m,l=1}^{N} \langle \gamma(\mathcal{J}_k^{ji} x_k), e_m \rangle a_{lm} \Gamma^l a_m + \langle \gamma(x_k), e_i \rangle a_{ji} \Gamma^j a_i.
$$
Filter for the auxiliary process $T$

$$\gamma(T_{k+1}^i(f(y_{k+1}^g))x_{k+1}) = \tilde{E}[\Lambda_{k+1}T_{k+1}^i(f(y_{k+1}^g))x_{k+1}|F_{k+1}^y]$$

$$= \sum_{l=1}^{N} \tilde{E} [\Lambda_k \Pi x_k \langle x_{k+1}, e_i \rangle (T_k^i(y_k^g) + f(y_{k+1}^g)(x_k, e_i)) |F_{k+1}^y] \Gamma^l$$

$$= \sum_{l=1}^{N} \tilde{E} [\Lambda_k \Pi x_k (\Pi x_k + u_{k+1}, e_i) (T_k^i(f(y_k^g)) + f(y_{k+1}^g)(x_k, e_i)) |F_{k+1}^y] \Gamma^l$$

$$= \sum_{l=1}^{N} \tilde{E} [\Lambda_k \Pi x_k (\Pi x_k, e_i) (T_k^i f(y_k^g)) + f(y_{k+1}^g)(x_k, e_i)) |F_{k+1}^y] \Gamma^l$$

$$= \sum_{l=1}^{N} \tilde{E} [\Pi (\Pi x_k, e_i) \Lambda_k x_k T_k^i (f(y_k^g)) |F_{k}^y] \Gamma^l$$

$$+ \sum_{l=1}^{N} \tilde{E} [\Lambda_k \Pi x_k, e_i \langle x_k, e_i \rangle \Pi x_k |F_{k}^y] f(y_{k+1}^g) \Gamma^l$$

$$= \sum_{m,l=1}^{N} \langle \gamma(T_k^i(f(y_k^g)))x_k, e_m \rangle a_{lm} \Gamma^l a_m$$

$$+ \sum_{l=1}^{N} \tilde{E} [\Lambda_k \Pi x_k, e_j \langle x_k, e_i \rangle \Pi x_k |Y_{k}^y] f(y_{k+1}^g) \Gamma^l$$

$$= \sum_{m,l=1}^{N} \langle \gamma(T_k^i(f(y_k^g)))x_k, e_m \rangle a_{lm} \Gamma^l a_m$$

$$+ f(y_{k+1}^g) \sum_{l=1}^{N} \langle \gamma(x)_k, e_i \rangle a_{li} \Gamma^l a_i.$$