High-Frequency Bayesian Modeling and Analysis of Stochastic Volatility in Finance

By: DERRICK HANG

Trinity College, Duke University, Durham, North Carolina 27708, U.S.A.

derrick.hang@duke.edu

April 2011

Abstract

In the literature, countless models exist that attempt to provide systematically superior forecasts of asset movements than the ones adopted by the market. Such a model, assuming it is successful, has numerous applications in finance, including derivative pricing and portfolio allocation. This paper explores the Bayesian multivariate dynamic linear model (DLM) with variance matrix discounting and assesses its ability to model a set of high-frequency foreign exchange rates on the US dollar in the first half of 2009, suggesting modifications to the basic specifications. Forecasting models are then constructed from the DLM fit and optimal day-by-day portfolio weightings are calculated from the perspective of an average trader. This research has several findings: (1) DLMs may outperform simple autoregressive returns models due to the inclusion of co-movements among variables across related times-series, (2) DLMs using intraday high-frequency data may lead more accurate price forecasts and better portfolio returns, (3) DLMs may benefit from the creation of a hybrid model that integrates the patterns found in models of various intraday data frequencies to produce a more-informed forecast estimate, and (4) further additions to the DLM literature must be made in order to fully benefit from use of intra-day data.

Keywords: high-frequency, intraday, Bayesian dynamic linear models, optimal portfolio allocation, variance discounting, foreign exchange, volatility co-movement

Introduction

Background

In finance, it is generally believed that complex models for returns and implicitly prices cannot systematically provide more accurate forecasts than those that simply specify returns to follow a random walk with zero mean and some variance V. From a cursory glance, this assumption appears to be valid in modeling the returns from a specified time series. However, limiting analysis to a single time series can be problematic; for the most part, assets are not used as single entities in investment, but instead are grouped into baskets. Furthermore, single time series analysis does not adequately incorporate the co-movements between other related time series and the underlying asset itself into the model. Studying changes in the relationships among a basket of assets, particularly volatility interactions, can uncover invaluable information that can be used in several applications such as determining optimal portfolio allocation and providing more accurate forecasts of price movements.

Following the work of Cai (2009), we chose to model a basket of currencies using a multivariate matrix normal dynamic linear model with variance matrix discounting. The basis of this model are found in Quintana and West (1987) and chapter 16 of West and Harrison (1997). The model specification that we employ in this analysis is outlined in Chapter 9 and 10 of Prado and West (2010). The multivariate framework is able to utilize information about the relationship between currencies in determining forecast estimates and also allows us to add other factors that we feel would improve the model fit.

Motivation

As alluded to above, understanding the interactions among a basket of assets, particularly the relationships with their volatility, can allow us to better understand the movements seen in the financial markets. Applications include determining more informed portfolio allocations, derivative pricing, and short-term forecasts. The matrix dynamic linear model with variance matrix discounting is able to accomplish this analysis since it allows for multiple times series, interactions, and a dynamic variance covariance matrix. In this paper, we posit that a basket of currencies would be a good candidate for this model since the interconnection between currencies are well-known and we expect these relationships to be relatively stable and dynamic. We utilize high frequency data with the rationale of obtaining more accurate short-term forecasting. Moreover, rather than returns, we directly model prices in order to capture the influence of significant predictors that might be masked when taking the first logged difference in prices.

Methods

Data

This dataset consists of 9 foreign exchange rates to the US dollar as well as Brent Crude Oil futures prices and Comex Gold future prices from January 2, 2009 to June 30, 2009, approximately 6 months worth of data provided by *www.forextickdata.com*. The data is high-frequency, providing 5-minute prices from 9:35AM to 3:55PM, excluding weekends. The currencies considered are listed below.

AUD	Australian Dollar
CHF	Swiss Franc
EUR	Euro
GBP	British Pound
JPY	Japanese Yen
NZD	New Zealand Dollar
CAD	Canadian Dollar
NOK	Norwegian Krone
ZAR	South African Rand

(1)

Currencies were chosen to be representative of the world markets although consideration was given to the expected magnitude of the effects of oil and gold on a given currency. For the initial model fit, we used the entire dataset to construct the dynamic linear model. In a later section, we will reserve a specified number of days to measure forecasting ability.

Model

For our analysis, we chose to model directly on prices rather than returns. The advantage of modeling on prices is that it preserves the influence of significant predictors whose forecasting abilities get differenced away once we transform prices to returns. This disappearance of predictive power occurs when the particular regressor remains relatively stable during a period of time while the dependent variable moves due to influences from other factors. The general intuition behind the regression model is staright-forward.

Consider the traditional returns equation:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \tag{2}$$

which is essentially the percentage change in price from time t - 1 to t. For small changes in price, the returns equation can be approximated by the following:

$$r_t = \log \frac{p_t}{p_{t-1}} \tag{3}$$

$$r_t = \log p_t - \log p_{t-1} \tag{4}$$

Now consider the following time-varying regression:

$$y_t = \alpha_t + \beta_t y_{t-1} + \gamma_t x_{t-1} + \nu_t, \ \nu_t \sim N(0, V_t)$$
(5)

where *y* is log prices of the dependent series, α is the intercept, β is the coefficient for the first lagged prices of the dependent series, *x* is the log prices of any "external" predictive time series, γ is the coefficient for *x*, and ν is the error term which follows a normal distribution with mean zero and variance V.

If the times series x had no predictive power on y, then $\gamma = 0$. If indeed all of the predictive ability for y comes from its first lagged variable y_{t-1} , $\beta = 1$. Letting $\alpha = 0$, we obtain the equation:

$$y_t = y_{t-1} + \nu_t, \ \nu_t \sim N(0, V_t)$$
 (6)

$$r_t = y_t - y_{t-1} = \nu_t, \ \nu_t \sim N(0, V_t)$$
(7)

which says that returns are modeled as a random walk. In general, it is expected that $\beta = 1$, $\alpha = 0$, $\gamma = 0$. Through our analysis, we will show that these values are close, but significantly different from their expected values, indicating that there is another useful predictor in the model beside the first lagged price of the dependent variable.

The specification of the mutlivariate matrix normal dynamic linear models is developed in Chapter 10 of Prado and West (2010). We directly borrow the notation and equation from the model summary from Cai (2009).

$$\mathbf{y}_{\mathbf{t}}' = \mathbf{F}_{\mathbf{t}}' \boldsymbol{\Theta}_{\mathbf{t}} + \nu_{\mathbf{t}}', \ \nu_{\mathbf{t}} \sim N(\mathbf{0}, \nu_t \boldsymbol{\Sigma}_{\mathbf{t}})$$
(8)

$$\Theta_{t} = \Theta_{t-1} + \Omega_{t}, \ \Omega_{t} \sim N(0, W_{t}, \Sigma_{t})$$
(9)

where $\mathbf{F}_t, \mathbf{\Sigma}_t, \mathbf{W}_t$ must be specified. \mathbf{F}_t is the $p \times 1$ vector of predictors, $\mathbf{\Sigma}_t$ is the $q \times q$ variance matrix, \mathbf{W}_t is the $p \times p$ system variance matrix, and Θ_t is the $p \times q$ state matrix of coefficients.

With respects to each component term in $\mathbf{y}'_{\mathbf{t}} = \langle y_{t,1}, ..., y_{t,q} \rangle'$, we obtain q univariate dynamic linear models at each t,

$$y_{t,j} = \mathbf{F}'_{\mathbf{t}} \theta_{t,j} + \nu_{t,j}, \ \nu_{t,j} \sim N(\mathbf{0}, \nu_t \sigma_{t,j})$$
(10)

$$\theta_{t,j} = \theta_{t-1,j} + \omega_{t,j}, \ \omega_{t,j} \sim N(\mathbf{0}, \sigma_{t,j} \mathbf{W}_{\mathbf{t}})$$
(11)

The multivariate model consists of exchangable time series and is optimized with variance matrix discounting. We specify a matrix normal, inverse Wishart prior for each t,

$$(\Theta_{t-1}, \Sigma_t | Data_{t-1}) \sim NIW(M_{t-1}, C_{t-1}, \beta h_{t-1}, \beta D_{t-1})$$
(12)

where C_t is the covariance matrix for Θ_t , h_t is the cumulative information degrees of freedom, β is the innovation of the covariance matrix C_t , and D_t is the sum-of-squares matrix. We lose $(1-\beta)100\%$ of the information in D_t at each time step.

Rather than dealing with the covariance matrix directly, we opted to work with the precision matrix by letting $\Phi_t = \Sigma_t^{-1}$ to obtain the prior,

$$(\mathbf{\Phi}_{t-1}|Data_{t-1}) \sim W(h_{t-1}, \mathbf{D}_{t-1}^{-1})$$
(13)

The sequential updating equations as presented in Cai (2009) are presented below.

- Posterior at t-1: $(\Theta_{t-1}, \Sigma_t | Data_{t-1}) \sim NIW(M_{t-1}, C_{t-1}, \beta h_{t-1}, \beta D_{t-1})$ and $(\Phi_{t-1} | Data_{t-1}) \sim W(\beta h_{t-1}, D_{t-1}^{-1})$
- Prior at t: $(\Theta_t, \Sigma_t | Data_{t-1}) \sim NIW(\mathbf{a_t}, \mathbf{R_t}, h_{t-1}, \mathbf{D_{t-1}})$ where $\mathbf{a_t} = \mathbf{M_{t-1}}$ is the $p \times q$ mean matrices and $\mathbf{R_t} = \mathbf{C_{t-1}} + \mathbf{W_t}$ is the $p \times p$ covariance matrix. $(\Phi_t | Data_{t-1}) \sim W(\beta h_{t-1}, (\beta \mathbf{D}_{t-1})^{-1})$
- One-step forecast at t 1: $(\mathbf{y}_t | Data_{t-1}) \sim N(\mathbf{f}_t, q_t \boldsymbol{\Sigma}_t)$ where $\mathbf{f}_t = \mathbf{a}'_t \mathbf{F}_t$ is the forecast mean $q \times 1$ vector with the scalar variance factor $q_t = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t$.

 Posterior at time t: (Θ_{t-1}, Σ_t|Data_t) ~ NIW(M_t, C_t, βh_t, βD_t) where M_t = a_t + A_te'_t, C_t = R_t - A_tA'_tq_t, A_t = R_tF_t/q_t and e_t = y_t - f_t. e_t is the vector that represents forecast errors. (Φ_t|Data_t) ~ W(βh_{t-1}, (βD_t)⁻¹) where h_t = βh_{t-1} + 1 and D_t = βD_{t-1} + e_te'_t/q_t.

rm1	1	1 .	1	• 1	1		. •	•	C		1 1			C 1	1
INP	dene	ndent	and	Inde	nend	ent	time.	ceriec	tor	011r	model	are	າເ	tol	1011/6.
IIIC	ucpu	mucin	anu	muc	pena	CIIL	unic-	SCIICS	101	our	mouci	arc	as	101	10 10 .
					1										

Dependent	Independent
AUD_t	AUD_{t-1}
CHF_t	CHF_{t-1}
EUR_t	EUR_{t-1}
GBP_t	GBP_{t-1}
JPY_t	JPY_{t-1}
NZD_t	NZD_{t-1}
CAD_t	CAD_{t-1}
NOK_t	NOK_{t-1}
ZAR_t	$ $ ZAR $_{t-1}$
Oil_t	Oil_{t-1}
Gold_t	$Gold_{t-1}$

Discount Factors

We must specify the discount factors β and δ for the variance-covariance matrix and the sum of square matrix respectively in order to implement the model. These factors specify the value of the observations, where the most recent observations contribute more to the estimates than past observations. Generally, β and δ will be close to 1, indicating steady change in variance covariance matrices.

Using the same method found in Cai (2009), we determine the optimal discount factor by comparing the marginal likelihood obtained from different values of the discount factors. The contour plot of our model is shown below:



Figure 1: Marginal likelihood function $p(y|\delta,\beta)$.

We find that optimal discount factors for both beta and delta are 0.9993. This value raises some suspicions in the adequacy of the model since it is extremely close to 1 and allows for very little change in the variance covariance matrices.

Although the value for the theta coefficients eventually stabilizes around a value with these discount factors, the DLM takes around a month worth of 5-minute data points before adapting to the data. The could be due to difficulties with distinguishing between real movements in level and microstructure noise. However, despite this, we still see the coefficients for the regressors to be significant from their expected limiting values of 0 or 1, indicating the presence of some useful predictors for the currency set besides their respective first lagged time series.



Figure 2: GBP on lagged GBP coefficient.(left) GBP on lagged JPY coefficient.(right)

Bayesian Monitoring for Discount Factors

The difficulties in specifying a single discount factor for both β and δ can be due to the use of high-frequency data and the interference of intra-day microstructure noise. Furthermore, days of unusually high volatility, or jump days, induced by an unexpected macroeconomic event can bias several of the 5-minute data points within that day. To control for these effects, we consider the prospect of including time-varying discount factors in our model. We will compare two models: one with a high discount factor and one with a relatively lower discount factor. To determine the periods in which a particular model is preferred, we use the Bayes factors method of model assessment and sequential monitoring developed for West (1986) and summarized in chapter 11 of West of Harrison (1997).

The log Bayes factors are calculated as follows:

$$\log(H_t) = \log(p_0(Y_t|D_{t-1})) - \log(p_1(Y_t|D_{t-1}))$$
(15)

where p_0 is the forecast distribution for model 0 and p_1 is the forecast distribution for model 1. The cumulative log Bayes factor equation at time k is just the summation of the log Bayes factors time 1 to k.

$$\log[H_t(k)] = \log(H_t) - \log[H_t(k-1)]$$
(16)

From Jeffreys (1961), a log Bayes factor above 1 indicates evidence for model 0, a log Bayes factor below -1 indicates evidence for model 1, and a log Bayes factor of 0 indicates no evidence for either model.

The monitoring variables L_t and l_t are defined as follows.

$$\log(L_t) = \log(H_t) + \min(0, \log(L_{t-1}))$$
(17)

$$l_t = \begin{cases} 1 + l_{t-1} & \text{if } \log(L_{t-1}) < 0\\ 1 & \text{otherwise} \end{cases}$$
(18)

where L_t detects a local deviation from model 0 and l_t counts the number of consecutive periods the alternative model is preferred. To clarify, when L_t indicates strong preference for the alternative model with a value less than τ , the alternate discount is used at that particular time point to allow for more uncertainty in the model; also when l_t indicates that a string of weak indicators for the alternative ($\tau < \log(L_t) < 0$) have occurred, the model switches to the alternate discount at the point after a specified threshold. In any other case, the original discount is used, which is expected for the majority of the points.

For ease of modeling, the delta and beta discounts are set to be equal. The discounts chosen for the primary model are 0.9993, as specified by comparing marginal likelihoods of a window of values as described in the previous section. The alternate discounts are specified using a similar approach; the original discount is compared to a window of possible lower discount factors and the discount that had the most point preferred to the original is chosen.



Figure 3: Alternate Discount Model Preference

The discount of 0.9813 had the highest frequency of preference with 1003 ticks and was, therefore, chosen to be the alternate discount value.

Recent literature on high frequency modeling have hinted at the existence of an intraday volatility smile, which states that there is higher volatility at the opening and closing periods of the market than at the midday. Looking at the empirical results of the Bayes Factor tracking model, we indeed find some weak evidence that is consistent with this phenomenon. If we re-specify the model to switch to the alternate if there is any evidence against the primary model (in other word, $\tau = 0$), we find that the beginning of all days prefer the lower discount factor, which flattens the posterior distributions to allow more variation at the time point. However, if we returned to setting τ to -1, a value that makes more a conservative assessment of model adequacy, the instances of this beginning-of-the-day event drop significantly from all start of days to a mere 28 days. Taking the results from both of these specifications into consideration, we decided to implement an automatic lower discount factor for the first two ticks of each day while keeping τ at the more conservative value of -1. Interestingly enough, with this design, the tracking model does not pick up any additional ticks in favor of the alternate model whereas, without the systematically lower start-of-day discount, it indicates 1003 points; in other words, only the first two points of everyday, or 252 ticks, use the alternate discount.

As an additional comment on this topic, although the basic sequential monitoring method from West (1986a) allows for model comparison, it does not provide a direct method of finding the optimal lower discount factor for the alternative model, nor is it able to consider more than two models. More research can be done to develop a more comprehensive Bayesian monitoring technique; however, this is not the main focus of this paper.

Day-to-Day Forecasting

The relative optimality of the weighting of our portfolio is closely linked to the ability of our model to forecast prices accurately. In using high frequency data, we hypothesized that these additional intra-day points will assist in making more precise estimates of future prices. Here, we look at forecasting from the standpoint of the average at-home trader who cannot make quick or frequent interactions but has access to historical high-frequency data. In this case, the trader is interested in forecasting in terms of days instead of hours or minutes.

k-Step-Ahead Model

Five-Minute Data

Using the fitted five-minute model, we must forecast 77 time-steps in order to obtain the estimate for the following day. However, there are two main problems associated with this notion: (1) the Bayesian DLM model as specified in Prado and West (2010) was constructed to only make one-step-ahead forecasts, and (2) the discounting of information accumulates with each additional forecast, which will result in extremely flat ending posterior distributions, and thus, unreliable ending predictions. Despite these major shortcomings, we will still attempt to make the 77-step forecasts. The n-step-ahead forecasting equations are defined as follows:

$$W_t = C_t(1/\delta - 1) \tag{19}$$

$$\Sigma_{t+n} \sim IW(D_t, n_t - q + 1) \tag{20}$$

$$\Theta_t \sim N(M_t, C_t, \Sigma_{t+1}) \tag{21}$$

$$\Omega_{t+n} \sim N(\mathbf{0}, W_t, \Sigma_{t+n}) \tag{22}$$

$$\Theta_{t+n} = \Theta_{t+n-1} + \Omega_{t+n} \tag{23}$$

$$\nu_{t+n} \sim N(0, \Sigma_{t+n}) \tag{24}$$

$$Y_{t+n} = Y'_{t+n-1}\Theta_{t+n} + \nu_{t+n}$$
(25)

Using the same variables as described earlier. Note that equation (21) the initialization equation. In this algorithm, the system variance matrix W_t , sum-of-squares matrix D_t , and degrees-offreedom for the distribution of the variance matrix Σ are all held constant at the last known value at time t through all k forecasts since no additional information is observed to necessitate updating. Also, the beta discounting of the information in D_t is ignored for the purposes of obtaining tighter predictions for prices. For simplicity, δ is always assumed to the original discount of 0.9993. Moreover, due to the extensive forecasting simulations required for a 77-step-ahead prediction, the diagonal of the state matrix Θ , which represent the coefficient for the first lagged price of each respective currency-pair, can wander to a value greater than 1, violating the assumptions of the model. Thus, these diagonal values are tracked for each step-ahead forecast and are reset to 1 if a simulation indicates a value greater than 1. The graphs below shows a comparison of the actual 5-minute prices for day 101 and the 77 forecasted values with the closing price previous day 100 as the last observation for two selected foreign currencies, the Swiss franc and British pound.



Figure 4: CHF forecasted vs. actual price for day 101.(left) GBP forecasted vs. actual price for day 101.(right)



Figure 5: CHF forecast absolute error for day 101.(left) GBP forecast absolute error for day 101.(right)

Looking at the forecasts for the franc, we see that our model provides an adequate prediction of the price until around the tenth forecast; the errors for the estimates for the subsequent ticks, then, becoming increasingly large. This pattern is observed for the almost all of the other time-series; the British pound being the only exception. Although the error of the forecast does increase for a period within the day, it does not stray too far away from the actual price (remaining within the .008 interval), and appears to have an element of control since it actually decreases near the end of the forecast instead of constantly increasing like the other series. The intra-day forecasts of British pound prices for day 71 and 111 were also examined to provide a basis of comparison.



Figure 6: GBP forecasted vs. actual price for day 71.(left) GBP forecasted vs. actual price for day 111.(right)



Figure 7: GBP forecast absolute error for day 71.(left)GBP forecast absolute error for day 111.(right)

Estimates for the ticks in day 111, like those for day 101, perform relatively well; however, the day 71 predictions follows the pattern prevalent in the other series; the errors are relatively small for first few ticks, and steadily increase for all subsequent ones. The explanation of this phenomenon is not readily known, but could be attributed to small values for the system variance and sum-of-square matrixes throughout a particular day. Overall, the 77-step-ahead forecasts according to our k-step model seems to provide poor one-day-ahead estimates based on examining one Monte-Carlo run for a randomly chosen day.

Fitting the DLM model with similar specifications using only daily end-of-day data on prices, we compare the one-step-ahead daily forecast obtained from this model to the 77-step-ahead daily forecasts obtained from the 5-minute model. As a side note, 100 Monte-Carlo estimates were averaged to calculate the 77-step-ahead forecast for each day. Graphs of the full forecasts as well as forcasting errors for each currency-pair and commodity can be found in the appendix.

Surprisingly, the 77-step-ahead forecast of the 5-minute data was better than the daily 1-stepahead estimate, in terms of absolute error from the actual value, for the clear majority of days for the AUDUSD, CHFUSD, GBPUSD, NZDUSD, and CADUSD, although the size of this error appears to gradually decrease and, at times, even favor the 1-step forecast near the end. Of these currencypairs, GBPUSD had the most points preferring the 5-minute method. EURUSD seems indifferent with respects to accuracy, and the forecasts for JPYUSD, NOKUSD, ZARUSD, Brent oil, and Comex gold resulted in dismal 77-step-ahead values, strongly favoring the 1-step forecasting.

For the time-series that favored the use of the 5-minute data, the decreasing difference of absolute error versus the daily 1-step might be attributed to the burn-in of the daily model. Since this model only has 127 points into total compared to the 5-minute models 9778 points, the daily model might not have enough data to make as accurate a forecast as it otherwise can. Since we do not see stability in this difference in absolute error in our model, it is completely possible that the 1-step-ahead forecasts will be systematically preferred, in accordance with the criteria of lowest absolute error, for time greater than a burn-in threshold T. Unfortunately, we do not possess the data to test this theory. Nonetheless, within our data framework, our 77-step-ahead forecast model appears to produce plausible predictions for at least some of the time series.

Embedded Multi-step Forecast Method

In our hypothesis, we assume that any additional intra-day information would help in producing better daily estimates for the data. Since the information of a desired forecasting frequency is contained within the higher frequency dataset (i.e. the 1-day only data is contained within the 5-minute dataset), forecasting using the higher frequency data should perform at least as well as forecasting using only the lower frequency dataset; in fact, the successful incorporation of data trends at these higher frequencies is expected to help generate better result. However, we see from our analysis thus far that directly forecasting 77-steps into the future with 5-minute data produces mixed and, often times, questionable results that might not be more accurate than forecasting using daily data. Since our DLM model only considered one-step forecasts, this occurrence is most likely attributed to the failure of our k-step forecasting model to fully capture the information in the intraday data. In fact, when looking at the graph of the 77-step forecasts in a particular day, only the first few forecasts are still able to capture the true pattern of the data. As a possible solution to this problem of multi-step forecast, we propose an alternate approach to better utilize at least a portion of the higher-frequency data in our DLM forecasting which we will call embedded forecasting. The intuition behind this theory is simple and will be presented in the following illustration. The effectiveness of the method is largely dependent on the accuracy of the previously defined k-step-ahead forecasting model, where k is no larger than the error threshold τ .

Assume you have intraday data for every 5-minutes of every day, adding up to be 76 data points per day including opening and closing values, but you initially decide to build a DLM model by only using daily data in order to obtain daily one-step-ahead forecasts. However, you find that by using the model of k-step ahead forecasting, it is reasonable to assume that two-step ahead forecasting at any given intra-day frequency is still fairly accurate and does not deviant too far from the true price. Therefore, you decide to construct a DLM model using half-day data and forecast two-steps ahead, obtaining a prediction that is more precise than the full-day one-step-ahead value. Afterward, you re-run the DLM model using quarter-day data and forecast two-steps ahead to attain a prediction of the first-half day price. If this produces a better estimate for the half-day price than the half-day data one-step forecast, the information in the quarter-day two-step forecast model should be somehow integrated in the half-day two-step forecast model to ultimately provide a more accurate one-day-ahead prediction.

With this approach, the model is able to utilize the useful portion of price trends of higher frequency data forecasts into more macro predictions. Alternatively, trends of lower frequency intra-day data can be used as guiding points for higher-frequency k-step forecasts.

However, the particular updating equations to integrate the information from the quarter-data forecasts in the half-day data forecasts are not obvious. Nevertheless, we will attempt to test the validity of this method using our dataset in a manner similar to the one described in the illustration.

Example

For this exploration, we will use the assumption that the two-step-ahead forecasting model provides an adequate estimate of the price. Also, we will remove the first points of everyday to end up with 76 intra-day points, a sum which is divisible by 4. Using similar DLM specifications as before and keeping the discount constant at 0.9993, the graph below compares the half-day 2-step forecast to the daily one-step forecast to the true price. Here we focus on two currency-pairs in which the 77-step-ahead forecast using 5-minute data was clearly inadequate to 1-step-ahead day-by-day forecasts: JPYUSD and gold.



Figure 8: Daily price forecasts



Figure 9: Daily price forecast errors

We find that these two-step-ahead estimates of the prices for JPYUSD and gold are far superior than those provide by the 1-step-ahead one-day forecasts, implying that this is also better than the 77-step ahead 5-minute forecasts for these particular datasets.

In accordance with our method, we now consider the next level of embedded forecasts: the quarter-day two-step estimate versus the half-day one-step estimate in determining the next half-day price. Again as before, similar DLM specifications are used in modeling the quarter-day data. We are again interested in finding out which prediction is more accurate for the next half-day price in terms of absolute errors.



Figure 10: Half-day price forecasts



Figure 11: Half-day price forecast errors

As expected, the two-step quarterly model is shown to produce more precise estimate than the one-step half-day model. However, the question still remains regarding the correct way to include the two-step-ahead quarterly model information in the two-step-ahead half-day model to result in a hybrid model that provides better daily estimates. To complete our exploration, we simply replaced the forecasted price and theta estimates of the 1-step-ahead half-day model with the one obtained from the 2-step quarterly model and proceeded to make the next step prediction of the half-day model. The resulting half-day forecasting equations at the 2nd step are as follows.

$$\Theta_{t+2,h} = \Theta_{t+2,q} + \Omega_{t+2,h} \tag{26}$$

$$Y_{t+2,h} = Y'_{t+2,q}\Theta_{t+2,h} + \nu_{t+2,h}$$
(27)

The following graph shows the estimates of this hybrid model compared to the simple two-step ahead half-day and the actual daily data for JPYUSD and gold.



Figure 12: Daily price forecast



Figure 13: Daily price forecast errors

With our particular specifications for the embedded model, we are not able to obtain conclusive evidence in favor of the hybrid model and against the 2-step forecast of the half-day data. Allowing the first 29 days for burn-in, only 50 percent of the remaining days preferred the embedded model in terms of absolute error, 55 percent for JPYUSD. This topic of integration of high-frequency information into macro forecasts is worth further exploration in other analyses with datasets spanning a longer time period.

Discussion: Integration Models

The embedded multi-step forecasting method, as illustrated above, represents a rough attempt to answer the bigger question of how to adequately combine the information obtained from different forecasting models to create one which produces consistently superior estimates. Here we discuss two other methods of information integration: Bayesian model averaging and Bayesian aggregation. Note, however, that the main difficulty with our scenario is that the difference between DLM models is not model specification; rather it is the amount of data that each model utilizes. Thus, the direct application of traditional combinatory methods may not be statistically cogent. The following brief discussion hopes to provide a starting point from which to conduct further research into this topic.

Multi-Process (Mixture) Models

Section 12.2 in West and Harrison (1997) elaborates on the general framework of a multi-process, class I model which is defined to be a model $M_t(\alpha)$ which holds for some $\alpha \in A$ for all t with parameter space A being a finite, discrete set $A = \{\alpha_1, ..., \alpha_k\}$ for some interger k > 1. The posterior probabilities for each alpha can then be defined and sequentially updated as follows for j = 1, ..., k:

$$p_t(j) = p(\alpha_j | D_t) = Pr[\alpha = \alpha_j | D_t]$$
(28)

$$l_t(j) = p(Y_t | \alpha_j, D_{t-1})$$
(29)

$$p_t(j) = c_t p_{t-1}(j) l_t(j)$$
(30)

where p_{t-1} is the prior probability for the model, l(t) is the likelihood defined to be the n-step ahead forecasting density, and c is a normalizing constant. These posterior probabilities act as weights to the overall multi-process model and the modification of the relationship among these weights can be observed through the use of their Bayes factors, which is simply the ratio between one models likelihood over another.

Although the Bayesian model averaging seems relatively straight-forward and logical, the direct application may prove inadequate in our scenario since the data Dt itself varies with each model. In fact, we are interested in finding the right mixture of models with different datasets which often case have differing model specifications by principle. Furthermore, the number of step ahead forecasting is not uniform across models as assumed by the general framework since they are a function of data frequency. Unfortunately, ways to reconcile this detail are not immediately known from traditional perspectives.

Bayesian Aggregation

One possibility is to view the results obtained from the models as the opinions of a group of people. This analogy is appropriate since each person has access to different amounts of information and uses a different thinking process in formulating their final opinion. The notion of a group belief distribution which attempts to build a conglomerate model superior to its individual parts is discussed in West (1984). In our case, the daily estimates generated by each of the forecasts using different time-steps (i.e. half-day, quarter-day, minute, etc.) can be viewed as three unique opinions and we are interested in finding an overall group distribution with weights a_1 , a_2 , a_3 representing the expertise or credibility of the opinion. Following section 3.3 of West (1984), the aggregation rule can then be applied to obtain a group belief density function by simply taking the product of the individual distributions raised to the a power, having initially defined a monotonic unity function. However, the exact form of the group density may not be easily identified especially when dealing

complex multi-dimensional forecast distributions; a unique maximum might not even exist in the final model. Also, adding to the difficulty of interpreting results from the group model, disagreement in opinion leads to a discrepancy factor which makes the area under the distribution sum to a number less than one.

The calculation of the weights is also ambiguous. On the whole, the group belief density is a measure of what the group agrees the forecasted value should be rather than what the true value actual is. The weights, then, should be time-varying reflecting the historical accuracy of each opinion to the true value. One rudimentary way we can do this is to allow for some initial data to see what proportion of the time window a certain model produced the closest prediction and create proportions from there. However, this is not adequate in our case since, as evidenced above, the higher frequency models explored in the paper generally outperform relatively lower frequency ones and, thus, essentially no weight will be given to lower frequency models. Instead, one could formulate some sort of weighting that accounts for accuracy not only in point predictions, but also in other aspects of pattern such as correct direction and favorable volatility conditions.

Optimal Portfolio Allocation

In this section, we approach the problem of optimal portfolio weighting from the viewpoint of an average trader who can only update his portfolio on a daily basis. From our analysis, we find the one-day forecasts using the 5-minute model to be sub-par in comparison to the forecasts just using daily closing prices. Furthermore, we discover that the 2-step-ahead estimates using half-day data provide more precise predictions than the 1-step-ahead daily model. Therefore, we decide to include the 2-step-ahead forecasts of half-day data in this discussion.

We define optimality as the portfolio weights that minimize risk given a specified daily return threshold. Since we are building a model for the average trader, we cannot short-sell and, thus, cannot have negative weights for any series in the portfolio. For simplicity, we also assume no interaction costs associated with re-weighting. Re-writing these considerations as equations, we have the following.

$$\min \Sigma (Z_{t+1} - P_t)^2 \tag{31}$$

$$\Sigma(\log Z_{t+1} - \log P_t) = r \tag{32}$$

$$\Sigma w_i = 1 \tag{33}$$

$$\mathbf{w} \ge 0 \tag{34}$$

where equation (28) is the optimized given equations (29),(30), and (31). Z_{t+1} represents the forecasted price, P_t is the last observed price, and w is the weighting matrix of the portfolio. Equation (29) states that the total forecasted daily return should be equal to a preset value.

The first 29 days of our daily data is burned to allow the forecasting model to adequately adapt. Lagrange multipliers were used to solve for the optimal weightings given the constraints. In determining the specified daily return threshold we considered multiples of the historical average daily market return, which is derived from the historical S&P500 annual return of 10%. For this particular iteration we specified a desired annual return of 40% or, in other words, a portfolio return of 0.13% per day.

Applying the weightings indicated by the Lagrange multipliers and calculating the actual daily returns for our portfolio, we end up with an annualized total return of -19.65%. Given the financial crisis and the volatility of global markets as a whole during the first half of the 2009, this negative portfolio return is not too surprising. Unfortunately, we do not have information regarding the average performance of other foreign exchange portfolios at that time period as a direct basis for comparison.

Now we calculate competing portfolio values by apply the same optimality conditions to the forecast data obtained from the one-step ahead daily model as well as those taken from the simple autoregressive model. Note that the AR(1) is merely a representation of the traditional returns equation mentioned at the outset of this paper. Here each the time series is modeled solely based on this own historical data; in other words, they do not account for possible correlation among the different series. We find the annualized returns for the one-step daily portfolio and the base autoregressive portfolio to be -26.96% and -31.09%, respectively.

Through a cursory comparison of the autoregressive and one-step daily DLM returns, we find evidence supporting our initial assertion that we can acquire superior forecasts by implementing models that take into account co-movements among variables within related time-series rather than by just relying on the traditional returns model. Likewise, by looking at the returns of the one-step daily versus the two-step half-day DLM returns at face value, we see slight evidence in favor of the use of higher frequency intraday data. However, we have only used the portfolio value at the end of the period to arrive at our conclusions. The following is a graph of the cumulative portfolio return percentage of each of the three models as each day progresses.



Figure 14: Culmulative Portfolio Values

As seen above, we only reach our expected model hierarchy near the end of the time period. Although this can be attributed to model-specific details such as not allowing enough burn-in for the DLM models to properly adapt or insufficient discount factors, the graph could highlight the inadequacy of the simple portfolio weighting model. More accurate price forecasts is not enough to guarantee universally superior forecast returns; rather, we must account for the overall size and variance of the movement in relation to the other time series forecasts. Our portfolio specifications do not include the variances of the forecasts as calculated by the DLM model into the risk equation we seek to minimize. Nonetheless, despite these concerns, we are still able to observe the DLM model returns surpass the AR(1) return and the half-day DLM results steadily rising over the daily DLM ones over time. Moreover, the autoregressive returns show slightly more volatile daily movements than the other two, which could illustrate the weaker hedging strategy due to ignorance of time series co-movements.

Further improvements can be made to the specification of our portfolio in hope of obtaining a more appealing end return. Suggested modifications would be to integrate the DLM variances into the overall portfolio risk equation and to allow to specified daily return threshold to vary according to the actualized past portfolio return of a given day. Looking at the portfolio return, a new threshold value of subsequent daily returns can be calculated that allows us to reach our desired annual return. However, although varying thresholds give us some hope in meeting our return goal, actually reaching the mark is not guaranteed and may even result in a poorer portfolio performance.

Concluding Remarks

Throughout this paper, we explore the feasibility of modeling high-frequency data for a basket of foreign exchange currency-pairs within the framework of Bayesian dynamic linear models. Though the data can be fitted with some innovations to the original model specification, the inability to make accurate k-step-ahead forecast for a large value of k seemingly trumps the usefulness in price forecasting at such a high frequency. The embedded forecasting theory is suggested, which proposed the basis of a possible method to incorporate information from increasingly-higher frequency data into the forecasting model; however, the optimal specifications of the embedded model are not readily discernable. Other methods of integration such as Bayesian model averaging and Bayesian aggregation could provide a basis for creating a more coherent and interpretable group model. In terms of optimal portfolio allocation, the tumultuous time period for which we have data from can partially explain the sub-par performance of our portfolio; moreover, since our portfolio only contained foreign exchange pairs and commodities, we must benchmark its performance against the average portfolio return of these markets in 2009 rather than against the average return of equity markets such as the SP500; unfortunately, at this time, we do not possess this information. On the other hand, comparing our portfolio returns to those obtained by different models, we find evidence supporting the use of DLMs and intraday high-frequency data. Also, it is important to note that our focus is mainly from the prospective of the average trader who is not able to effectively engage in high-frequency trades. The one-step forecasts obtained from fitted DLM model of the 5-minute data may still be useful to Wall Street traders who are able to make several interactions on any given day.

As noted on the outset of this paper, several historical research papers, such as Meese & Rogoff (1983), suggest that a nave random walk strategy dominates more sophisticated forecasting techniques in predicting exchange rates as shown by their analysis of a portfolio of foreign currencies. As steps for follow-up analysis, further research could look to compare the performance of our portfolio with that of one obtained from additional random walk models. Additionally, within the DLM framework, a method can be developed to recorded the time periods in which a predictor other than the currencys own lagged price is significant; the can subsequently be used in calculating how much to attribute to that predictor in determining a particular currencys price in percentage-terms.

Acknowledgements

I would like to thank my research professor Dr. Mike West for introducing me to the topic of Bayesian forecasting and for providing expert guidance throughout this exploration; my questions and ideas were always treated with great consideration and I can only hope other advisors are as attentive to their advisees. Also, I would like to acknowledge Haolan Cai and Hao Wang for their assistance in the learning process.

Appendix



Figure 15: Daily estimates of every day. (left) Difference between the absolute error of the 1-step forecast compared to the 77-step forecast.(right)



Figure 15 (continued): Daily estimates of every day. (left) Difference between the absolute error of the 1-step forecast compared to the 77-step forecast.(right)



Figure 15 (continued): Daily estimates of every day. (left) Difference between the absolute error of the 1-step forecast compared to the 77-step forecast.(right)



Figure 15 (continued): Daily estimates of every day. (left) Difference between the absolute error of the 1-step forecast compared to the 77-step forecast.(right)

References

- [1] Cai, H. (2009). In preparation
- [2] Jeffreys, H. (1961). Theory of Probability, 3rd Ed. Oxford University Press.
- [3] Meese, R.A. & Rogoff, K. (1983). Theory of Probability, 3rd Ed. Oxford University Press. Empirical exchange rate models of the seventies: Do they fit out of sample?. *Journal of International Economics* 14, pp. 324
- [4] Prado, R. & West, M. (2010). Time Series: Modelling, Computation and Inference. Chapman & Hall & CRC Press, The Taylor Francis Group.
- [5] Quintana, J. & West, M. (1987). An Analysis of International Exchange Rates Using Multivariate DLM's. *Journal of the Royal Statistical Society. 36, 275-81*
- [6] West, M. (1986). Bayesian Aggregation. Journal of the Royal Statistical Society. 147, 600-607
- [7] West, M. (1986). Bayesian Model Monitoring. Journal of the Royal Statistical Society. 48, 70-78
- [8] West, M. & Harrison, J. (1997). Bayesian Forecasting and Dynamic Linear Models, 2nd Ed. Springer Publications.