

CAN THE MARKOV SWITCHING MODEL WITH TIME VARYING TRANSITION PROBABILITIES FORECAST EXCHANGE RATES?

BONG-HAN KIM · JOON-HAENG LEE*

We use Lee's (1991) Markov switching model with time varying transition probabilities (the TVTP Markov model) to analyze the behavior of the U.S. dollar/ British Pound exchange rate. We employ the magnitude of the deviation of the exchange rate from a monetary equilibrium value as the economic fundamental with which the transition probabilities vary.

The empirical results we obtained from this paper are the following: first, the deviation of the exchange rate from a monetary equilibrium value can affect the transition probabilities of the exchange rate from one state to another state. When the exchange rate is overvalued (undervalued) relative to the monetary equilibrium value, the probability of the exchange rate to appreciating (depreciating) will be low (high). Second, the TVTP Markov model can identify both the appreciation state and the depreciation state of the exchange rate better than the Markov model with fixed transition probabilities (the FTP Markov model). Third, the forecasts of the TVTP Markov model are superior at predicting the direction of change of the exchange rate to those of the FTP Markov model.

JEL Classification: F31, F34

Keywords: TVTP Markov Switching Model, Pound-dollar Exchange Rate,
Out-of-sample Forecasts

I. INTRODUCTION

Engel and Hamilton (1990) characterized the behavior of major foreign exchange rates under the recent floating system as long swings, which drift upward for a considerable period of time, and then switch to a long period with

Received for publication: March 30, 2000. Revision accepted: Nov. 12, 2000.

* Department of International Trade and Commerce, Kongju University, Department of International Trade, Suwon University.

The authors would like to thank G.S. Maddala, James Hamilton, and Prasad V. Bidarkota for helpful comments on an earlier draft.

downward drift. Long swings of exchange rates can be explained well by Hamilton's (1990) Markov switching model. For instance, Engel (1994) found that the Markov switching model fits well in-sample many exchange rates, and is good at predicting the direction of change.

However, the Markov model cannot generate superior forecasts to a random walk model, or the forward rate, by the root-mean-squared-error criterion. The poor forecasting ability may be due to the fact that Hamilton's (1990) Markov switching model does not use any fundamental economic variables, except the exchange rate itself. Hence, one major defect of the Markov switching model is that it is a simple trend change model, with no explanatory variables. Because of this defect, although the Markov switching model fits well in-sample, it yields bad out-of-sample forecasts of many exchange rates.

There have been some attempts¹ to extend Hamilton's (1990) Markov switching model. Lee (1991) is one of fruitful extensions of the Markov model. The transition probability of Hamilton's (1990) model is restricted to being a constant. The probability of an exchange rate shifting from one state to another is fixed, irrespective of economic and foreign exchange market situations. However, the probability of an exchange rate continuing to drift upward (downward) can vary with economic fundamentals such as differences in money supply, income, or the short-run interest rates of relevant countries. Lee (1991) extended the Hamilton model to incorporate time-varying transition probabilities between states, and developed an EM algorithm to estimate his extended Markov model. We expect Lee's (1991) Markov switching model with time-varying transition probabilities to fit major exchange rates better than Hamilton's (1990) model can both within and out-of-sample.

When using Lee's (1991) Markov model, it is important to carefully select economic fundamentals which can affect the transition probabilities of exchange rates. Economic intuition may suggest that the possibility of regime shifts of an exchange rate increases under more severe overvaluation (undervaluation) on the basis of economic fundamentals. Interesting contributions supporting this idea have been recently made. Mark (1995) and Chinn and Meese (1995) found that the magnitude of the deviation of an exchange rate from its equilibrium value, determined by a monetary exchange model, can be a useful explanatory variable for its movement. This implies that the farther the exchange rate is from the monetary equilibrium value, the more likely it is for a regime shift of the exchange rate to take place. Hence, in order to analyze the behavior of the U.S. Dollar/British Pound exchange rate under the recent floating exchange rate system, we will use the Markov switching model, with the transition probabilities depending on the amount by which the exchange rate deviates from a monetary

¹ Garcia and Perron(1996) and Raymond and Rich(1992) enlarged the number of state from 2 to 3 or 4. Lee(1991) and Filardo(1994) includes explanatory variables with the transition probabilities depending on economic fundamentals.

equilibrium value.

The rest of the paper is organized as follows. In the next section, we set out the Markov switching model of exchange rates characterized by time-varying transition probabilities. Section 2 provides the basic framework of the EM algorithm and constructs the complete data likelihood function to execute the EM method and then suggests an estimation procedure for maximum likelihood (ML). Section 4 shows how to calculate the magnitude of the deviation from a monetary equilibrium value, using Park and Hahn's (1995) cointegration regression with time-varying coefficients (TVC cointegration). In section 5, we present our empirical results. Concluding remarks are contained in section 6.

II .THE MARKOV SWITCHING MODEL WITH TIME-VARYING TRANSITION PROBABILITIES

The exchange rate (e_t) used in this paper is the U.S. dollar price of one unit of British Pound. We denote by y_t the first difference of the natural logarithm of the exchange rate. Our model introduces an unobserved scalar variable s_t which takes on the value one or two. s_t characterizes the "state" or "regime" that the process was in at date t . When $s_t=1$, the observed y_t is presumed to have been drawn from $N(\mu_1, \sigma_1^2)$, whereas when $s_t=2$, the observed y_t is assumed to be distributed as $N(\mu_2, \sigma_2^2)$. Thus, when $s_t=1$, the trend in the exchange rate is μ_1 which will be expected to be positive, so this state represents depreciation. When $s_t=2$, μ_2 will be expected to be negative, representing appreciation. So, the distribution of the data (y_t), given a particular value for the state variable ($s_t=1$ or $s_t=2$), is

$$p(y_t|s_t, \theta) = \frac{1}{\sqrt{2\pi\sigma_{s_t}^2}} \exp \left[-\frac{(y_t - \mu_{s_t})^2}{2\sigma_{s_t}^2} \right] \quad (1)$$

We assume that $\{S_t\}_{t=1}^T$ is the sample path of a first order, two-state Markov process with the transition probability function as follows:

$$P_t = \begin{bmatrix} p_{11}^t & p_{12}^t \\ p_{21}^t & p_{22}^t \end{bmatrix} \quad t=1, 2, \dots, T \quad (2)$$

where $p_{ij}^t = \text{prob}(s_t = j | s_{t-1} = i, \mathcal{Q}_{t-1})$ and \mathcal{Q}_{t-1} is the information set available at time $t-1$.

Engel and Hamilton (1990) assumed stationary first order Markov process, in the sense that p_t is constant over time and independent of \mathcal{Q}_{t-1} . They found that

estimates of p_{ii}^t , probabilities of remaining a state, are fairly high². Thus, if a currency is appreciating (depreciating) at date t , appreciation (depreciation) will usually continue at the next period. In other words, no matter what situation the foreign exchange market is in, switching from one state to other is very unlikely. This means that market agents are always surprised when any regime switch takes place. But the likelihood of a government intervention in the foreign exchange markets, which can cause a regime switch may not be independent of the market status.

In general, information about economic fundamentals may help investors to improve their forecasts of future changes in regime of the exchange rate. They can be used as a signal of what the state will be. Mark (1995) and Chinn and Meese (1995) found that the deviation of the exchange rate from a monetary equilibrium relationship can be useful for predicting future changes in the exchange rate. This implies that there may exist a cointegration relationship between the exchange rate and the explanatory variables of the monetary exchange rate determination model, and so the exchange rate tends to converge to a monetary equilibrium value in the long run. We conjecture that the more the exchange rate deviates from a monetary equilibrium value, the higher is the probability that a regime switch will take place in the exchange rate. Hence, we expect that the transition probability depends on the magnitude of the deviation of the exchange rate from a monetary equilibrium value. We denote the magnitude of the deviation at $t-1$ as f_{t-1} . We now construct the transition probabilities as follows:

$$\begin{aligned} p_{11}^t &= p(s_t = 1 | s_{t-1} = 1, f_{t-1}; \Psi) = h(f_{t-1}, \alpha) \\ p_{12}^t &= p(s_t = 2 | s_{t-1} = 1, f_{t-1}; \Psi) = 1 - h(f_{t-1}, \alpha) \\ p_{21}^t &= p(s_t = 1 | s_{t-1} = 2, f_{t-1}; \Psi) = 1 - k(f_{t-1}, \beta) \\ p_{22}^t &= p(s_t = 2 | s_{t-1} = 2, f_{t-1}; \Psi) = k(f_{t-1}, \beta) \end{aligned} \quad (3)$$

The number of parameters Ψ depends on how we construct the probability function. The functions $h(\cdot)$ and $k(\cdot)$ will be specified. We parameterize the unconditional distribution of the initial state as independent of Ψ . Engel and Hamilton (1990) assumed that the initial state is drawn from the ergodic Markov process for s_t , thus depending upon Ψ . But this assumption is not valid for our model of the initial state since the transition probabilities are not stationary over time. We treat the probability of the initial state as a free parameter, $p(s_1 = 1) = \rho$, to be estimated by maximum likelihood. Of course, $p(s_1 = 2) = 1 - \rho$.

² Their estimates of p_{11} 's are 0.848, 0.822 and 0.927 for Germany, France and U.K. respectively, and the corresponding p_{22} 's are 0.928, 0.908, and 0.913, respectively.

III. ESTIMATION OF THE MODEL

3.1 EM algorithm for incomplete data.

The Markov switching model is regarded as an incomplete data problem due to the lack of observations on states. The expectation maximization (EM) algorithm of Dempster, Laird and Rubin (1977) is convenient to deal with maximum likelihood estimation of incomplete data models. The EM algorithm is a robust procedure for maximizing the incomplete data likelihood via iterative maximization of the expected complete-data likelihood, conditional upon the observed data. This algorithm starts with an initial guess of the unknown parameters of the model, and then generates a sequence of better guesses, in the sense that they increase the value of the likelihood function. Finally, the sequence of guesses of the parameters converges to a local maximum of the likelihood function.

A more specific description of the EM algorithm is as follows: To get the MLE of $\phi = (\theta, \Psi, \rho)$, we would like to use the joint density function of the complete data $\underline{\chi}_T = (\underline{y}_T, \underline{s}_T, \underline{f}_T)$, where $\underline{y}_T = (y_1, y_2, \dots, y_T)'$, $\underline{s}_T = (s_1, s_2, \dots, s_T)'$, $\underline{f}_T = (f_0, f_1, \dots, f_{T-1})'$. However, the complete data is not observable, since we have no observations on \underline{s}_T . Hence, we maximize the expected value of the joint density function of the complete data, conditional on the incomplete but observable data $(\underline{y}_T, \underline{f}_T)$. In the expectation step, we construct expected log-likelihood $Q(\phi; \phi^{(p)}) = E[\ln f(\underline{\chi}_T; \phi) | \underline{y}_T, \underline{f}_T, \phi^{(p)}]$, where f is the density function of the complete data, $\underline{\chi}_T$, and $\phi^{(p)}$ is the estimate of ϕ at the p -th iteration. In the Maximization step, we select $\phi^{(p+1)}$ to be the value of ϕ which maximizes $Q(\phi; \phi^{(p)})$. These two steps are repeated, until ϕ converges.

The EM algorithm is attractive when the incomplete data $(\underline{y}_T, \underline{f}_T)$ follows a model that is complicated to analyze directly, but there is a model of the complete data $\underline{\chi}_T$ whose likelihood function is easy to maximize. That is, we estimate the unknown parameters in the maximization step by maximizing the function Q as though the complete data were observed.

3.2 Likelihood function for the model.

Suppose we were able to observe all the states (\underline{s}_T) directly. The joint likelihood for the states (\underline{s}_T) and the observable data $(\underline{y}_T, \underline{f}_T)$ - complete data specification-would be given by

$$\begin{aligned} f(\underline{y}_T, \underline{s}_T | \underline{f}_T; \phi) &= p(y_1, s_1 | \underline{f}_T; \phi) \cdot \prod_{t=2}^T p(y_t, s_t | y_{t-1}, s_{t-1}, \underline{f}_T; \phi) \\ &= p(y_1 | s_1; \theta) \cdot p(s_1) \cdot \prod_{t=2}^T p(y_t | s_t; \theta) \cdot p(s_t | s_{t-1}, f_{t-1}, \Psi) \end{aligned} \quad (4)$$

where the unknown parameters ϕ consist of $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, $\Psi = (\alpha, \beta)$ and ρ . From the relationship between the complete and the incomplete data specifications, the likelihood function for the incomplete data can be obtained by summing (4) over all possible values of \underline{s}_T :

$$g(y_T | f_T; \phi) = \sum_{s_T=1}^2 \sum_{s_{T-1}=1}^2 \cdots \sum_{s_1=1}^2 f(y_T, s_T | f_T; \phi) \quad (5)$$

One can attempt to maximize the incomplete data likelihood directly by using traditional numerical maximization methods, such as, the Newton-Raphson algorithm or the scoring method. However, such a procedure becomes computationally complex and, in some cases, convergence may not be achieved. The estimates of parameters α and β are very sensitive to the likelihood function surface, since our model assumes complicated transition processes which are affected by the values of the economic fundamentals (\underline{f}_t). It is possible that the numerical maximization algorithm may either fail or stop at the boundary of the parameter space due to poor initial values.

Therefore, we specify the complete data log-likelihood function whose expectation will be maximized :

$$\begin{aligned} \ln f(y_T, s_T) = & \sum_{t=1}^T \{ I(s_t=1) \cdot \ln p(y_t | s_t=1) + I(s_t=2) \cdot \ln p(y_t | s_t=2) \} \\ & + \sum_{t=2}^T \{ I(s_t=1, s_{t-1}=1) \cdot \ln p_{11}^t + I(s_t=2, s_{t-1}=1) \cdot \ln p_{12}^t \quad (6) \\ & + I(s_t=1, s_{t-1}=2) \cdot \ln p_{21}^t + I(s_t=2, s_{t-1}=2) \cdot \ln p_{22}^t \} \\ & + \{ I(s_1=1) \cdot \ln \rho + I(s_1=2) \cdot \ln(1-\rho) \} \end{aligned}$$

where $I(A)$ denotes the indicator function of the set A and P_{ij}^t for $i, j=1, 2$ is defined in (3). Notice that the complete-data log-likelihood function can be decomposed into three components, where the parameters of each component are not related. Suppose that we have direct observation of \underline{s}_T .

The first component involves only the parameters θ which describe the characteristics of densities of $\{y_t\}$. The second component has the parameters Ψ specifying the transition probability function. The third one involves the initial condition parameter ρ . Thus, they can be maximized separately.

All that we need in the E-step of the EM algorithm is to take expectation of the following complete-data log likelihood, conditional upon the observed data.

$$Q(\phi^{(p+1)} | \phi^{(p)}) = E[\ln f(y_T, \underline{s}_T) | \underline{f}_T, \phi^{(p)}] \quad (7)$$

This is the objective function to be maximized in the M-step of the EM algorithm. To construct $Q(\phi^{(p+1)} | \phi^{(p)})$, we need the expected values of the

indicator functions, given the observed data and parameters $\phi^{(p)}$. This amounts to substitution of the smoothed state probabilities³ for indicator functions in the complete-data log likelihood function. For example,

$$I(s_t=1, s_{t-1}=1) = \text{prob}(s_t=1, s_{t-1}=1 \mid \underline{y}_T, \underline{f}_T; \phi^{(p)})$$

3.3 Parameter estimation

Once we get smoothed probabilities, the maximization problem can be regarded as weighted maximum likelihood estimation problem in which we use as posterior probabilities that sample observations belong to an appropriate state, given the current approximate estimate $\phi^{(p)}$.

From the first order conditions, we have closed form solutions for $\mu_1, \mu_2, \sigma_2^2, \rho$ as follows. For notational simplicity, we suppress dependence of the smoothed probabilities on $(\underline{y}_T, \underline{f}_T)$ and $\phi^{(p)}$ throughout this section. For example, the smoothed probability for state 1 is expressed as $p(s_t=1)$.

$$\hat{\mu}_i = \frac{\sum_{t=1}^T y_t \cdot p(s_t=i)}{\sum_{t=1}^T p(s_t=i)} \quad i=1,2 \quad (8-a)$$

$$\hat{\sigma}_i^2 = \frac{\sum_{t=1}^T (y_t - \hat{\mu}_i)^2 p(s_t=i)}{\sum_{t=1}^T p(s_t=i)} \quad i=1,2 \quad (8-b)$$

$$\hat{\rho} = p(s_1=1) \quad (8-c)$$

We can obtain closed form solutions for the parameters, α and β , in the transition probability functions for some special cases. When the transition probability function is constant over time - $h_t=\alpha$, $k_t=\beta$ for all t - we can get the following closed form solutions.

$$\hat{\alpha} = \frac{\sum_{t=2}^T p(s_t=1, s_{t-1}=1)}{\sum_{t=2}^T p(s_{t-1}=1)}, \quad \hat{\beta} = \frac{\sum_{t=2}^T p(s_t=2, s_{t-1}=2)}{\sum_{t=2}^T p(s_{t-1}=2)}$$

In general, we have nonlinear equations for the parameters α and β from the first order conditions. Our task is to get closed form solutions for these

³ The smoothed probabilities can be calculated by the method suggested by Hamilton(1990)

parameters. For general transition probability functions, these parameters can be expressed in closed forms only when the first order conditions are modified.

When the transition probability is a function of economic fundamentals (f_t), we need to select an appropriate functional form for the transition probability. Even though any specification that maps economic fundamentals into the unit interval $[0,1]$ is a valid candidate, the maximum likelihood estimation will constrain the set of valid candidates. The logistic function, probit function, Cauchy integral and piecewise continuously differentiable functions are all valid candidates. We employ the logistic functional form for transition probability. The logistic transition probability function is expressed as:

$$h_t = \frac{\exp(\alpha_1 + \alpha_2 \cdot f_{t-1})}{1 + \exp(\alpha_1 + \alpha_2 \cdot f_{t-1})}, \quad k_t = \frac{\exp(\beta_1 + \beta_2 \cdot f_{t-1})}{1 + \exp(\beta_1 + \beta_2 \cdot f_{t-1})} \quad (9)$$

If we rearrange the first order conditions using the derivative of the logistic function, we have the following:

$$\sum_{t=2}^T \{p(s_t = 1, s_{t-1} = 1) - h_t \cdot p(s_{t-1} = 1)\} = 0 \quad (10)$$

$$\sum_{t=2}^T \{p(s_t = 1, s_{t-1} = 1) - h_t \cdot p(s_{t-1} = 1)\} \cdot f_{t-1} = 0 \quad (11)$$

To get closed form expressions for $\alpha = (\alpha_1, \alpha_2)$, the linear approximation of h_t is derived from the Taylor expansion around $\alpha^0 = (\alpha_1^0, \alpha_2^0)$ up to first order.

$$h_t(\alpha) \cong h_t(\alpha^0) + \frac{\partial h_t(\alpha)}{\partial \alpha} \Big|_{\alpha=\alpha^0} \cdot (\alpha - \alpha^0) \quad (12)$$

This approximation is quite reliable because the evaluation points α^0 are close enough to α in the later stages of the EM algorithm. Let us denote

$$h_{1t}'(\alpha^0) = \frac{\partial h_t(\alpha^0)}{\partial \alpha_1}, \quad h_{2t}'(\alpha^0) = \frac{\partial h_t(\alpha^0)}{\partial \alpha_2}$$

Substituting equation (12) into equations (10) and (11) and rewriting them, we have two linear equations for the parameters α_1 and α_2 and their closed form solutions are given by:

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \begin{bmatrix} \sum_{t=2}^T p(s_{t-1} = 1) h_{1t}'(\alpha^0) & \sum_{t=2}^T p(s_{t-1} = 1) h_{2t}'(\alpha^0) \\ \sum_{t=2}^T p(s_{t-1} = 1) h_{1t}'(\alpha^0) & \sum_{t=2}^T p(s_{t-1} = 1) h_{2t}'(\alpha^0) \cdot f_{t-1} \end{bmatrix}^{-1}$$

$$\times \left[\begin{array}{c} \sum_{t=2}^T [p(s_t=1, s_{t-1}=1) - p(s_{t-1}=1) \{h_t(\alpha^0) + h_{1t}(\alpha^0) \cdot \alpha_1^0 + h_{2t}(\alpha^0) \cdot \alpha_2^0\}] \\ \sum_{t=2}^T [p(s_t=1, s_{t-1}=1) - p(s_{t-1}=1) \{h_t(\alpha^0) + h_{1t}(\alpha^0) \cdot \alpha_1^0 + h_{2t}(\alpha^0) \cdot \alpha_2^0\}] f_{t-1} \end{array} \right]$$

In the same way, we can derive closed form solutions for β_1 and β_2 as:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \left[\begin{array}{cc} \sum_{t=2}^T p(s_{t-1}=2) k_{1t}(\beta^0) & \sum_{t=2}^T p(s_{t-1}=2) k_{2t}(\beta^0) \\ \sum_{t=2}^T p(s_{t-1}=2) k_{1t}(\beta^0) & \sum_{t=2}^T p(s_{t-1}=2) k_{2t}(\beta^0) \cdot f_{t-1} \end{array} \right]^{-1} \times \left[\begin{array}{c} \sum_{t=2}^T [p(s_t=2, s_{t-1}=2) - p(s_{t-1}=2) \{k_t(\beta^0) + k_{1t}(\beta^0) \cdot \beta_1^0 + k_{2t}(\beta^0) \cdot \beta_2^0\}] \\ \sum_{t=2}^T [p(s_t=2, s_{t-1}=2) - p(s_{t-1}=2) \{k_t(\beta^0) + k_{1t}(\beta^0) \cdot \beta_1^0 + k_{2t}(\beta^0) \cdot \beta_2^0\}] f_{t-1} \end{array} \right]$$

The EM algorithm employed in this paper is outlined by the following steps.

- Step 1: We begin with an initial guess of the unknown parameters, say $(\phi^{(0)})$,
Then for the p -th iteration.
- Step 2: Compute the smoothed probabilities, based on the whole data y_T and $\phi^{(p-1)}$.
- Step 3: Compute an improved estimate of ϕ , denoted $\phi^{(p)}$, from (8) and closed form solutions of α and β .
- Step 4: Check for convergence of the parameter values.
If convergence is not reached, set $\phi^{(p)}$ to $\phi^{(p+1)}$, $p = p + 1$ and return to step 2.

In summary, we do not estimate the incomplete-data likelihood function directly. Rather, starting with an initial guess, we convert the incomplete-data problem to a complete-data one by replacing each of the unobserved state variables with their smoothed probabilities. In other words, we use the smoothed probabilities to reweigh the observed data y_t . Then we calculate new estimates by using explicit solutions. These new estimates are then used to recalculate the smoothed probabilities, and the data are reweighed with the new probabilities. Each such calculation of probabilities and reweighing the data can be shown to increase the value of the likelihood function. The process is repeated until a fixed point for ϕ is found.

IV. THE DEVIATION FROM THE EXCHANGE RATE FROM AN EQUILIBRIUM VALUE.

Mark (1995) and Chinn and Meese (1995) recently showed that long horizon exchange rate movements can be explained using predetermined values of the

differences between the actual and monetary equilibrium values of the exchange rate. Hence, we assume that the transition probabilities of the exchange rate between states vary with the deviation of the actual exchange rate from a monetary equilibrium value.

The (flexible-price) monetary exchange rate determination model relies on the twin assumptions of purchasing power parity and the existence of stable money demand functions for the domestic and foreign economies. If the constraint that the income and the interest rate elasticities of the money demand function are equal across countries is imposed, a monetary equilibrium value of the log-exchange rate can be determined by the following equation:

$$e_t = \lambda_0 + \lambda_1(m_t - m_t^*) + \lambda_2(y_t - y_t^*) + \lambda_3(i_t - i_t^*) + \varepsilon_t \quad (13)$$

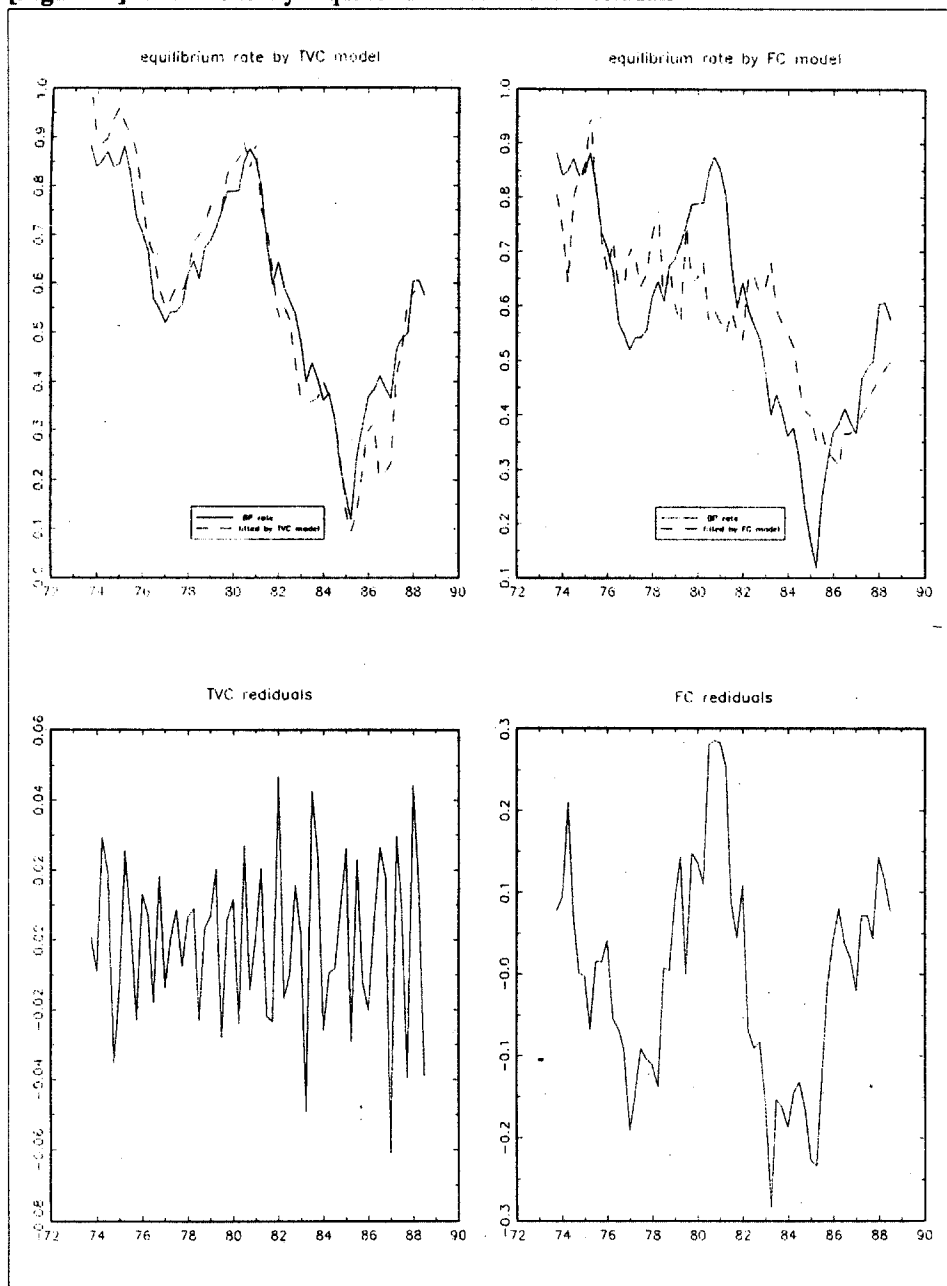
where e_t , m_t , y_t and i_t are the logarithms of the spot exchange rate (domestic currency price of foreign currency), domestic money supply, domestic real income, and the level of the domestic nominal interest rate. The corresponding foreign variables are denoted by an asterisk.

The residuals ($e_t - \hat{e}_t$) of the regression (13) can measure the amount by which the exchange rate deviates from the monetary equilibrium value. However, there does not exist a consensus on the evidence for cointegration of the exchange rate with the explanatory variables of the monetary model we consider⁴. Kim and Yu(1996) demonstrated that when the coefficients of equation (13) are fixed over time, the regression fits considerably poorly in-sample for major exchange rates and the residuals of the regression seemed to be highly nonstationary. They used Park and Hahn's (1995) TVC cointegration regression method⁵ in which any regression coefficients can be time-varying. When the regression is estimated with a varying coefficient (λ_1) of the difference of the money supplies, the fitness of the regression model improved significantly and the residuals became much more stationary. We shall work with both fixed coefficient (FC) residuals and time-varying coefficient (TVC) residuals of regression (13) and compare these residuals in terms of their explanatory power for movements of exchange rates and the degree of stationarity of the residuals.

In Figure 1, the fitted exchange rates and the residuals from regression (13) are plotted. When regression (13) is estimated with a time-varying coefficient of the monetary aggregate difference, the regression fits well and the residuals obtained seem to be considerably stationary, as Kim and Yu (1996) showed.

⁴ The existence of a cointegration relation of the monetary exchange determination equation has been generally rejected in studies using residual-based tests. However it has been not rejected in studies which used the Johansen method.

⁵ Park and Hahn (1995) showed that the coefficients of a regression can be approximated by time and trigonometric polynomial functions. The maximum orders of time and trigonometric polynomial functions can be selected with the GCV or the AIC criteria.

[Figure 1] The Monetary Equilibrium Rates and Residuals

We can now check whether FC and TVC residuals can be useful for predicting changes in the exchange rate, by making use of regression (14). Mark (1995) called (14) the long-horizon regression:

$$e_t - e_{t-1} = \gamma_0 + \gamma_1 f_{t-1} + \eta_t \quad (14)$$

where $(e_t - e_{t-1})$ is the change in the exchange rate at date t and f_{t-1} denotes the information available at date $t-1$. Because we assume that the deviation of the exchange rate from a monetary equilibrium value can explain future movements of the exchange rate, the magnitude of the deviation at date $t-1$ which can be represented by the residuals of regression (13) at date $t-1$, can be one appropriate candidate for f_{t-1} . When the adjustment coefficient (γ_1) of regression (14) is estimated to be significantly less than one, we can say that the deviation from the equilibrium relationship seems to adjust in more than one period. We therefore use as f_{t-1} various sums of residuals from date $t-1$ to the near past.

Table 1 reports estimation results of regression (14). In all the cases considered,

[Table 1] Estimates of Regression $e_t - e_{t-1} = \gamma_0 + \gamma_1 f_{t-1} + \varepsilon_t$

e_t is 100 times the log of U.S. Dollar/British Pound exchange rate and f_{t-1} the deviation of the actual value from a monetary equilibrium value of the exchange rate.

f_{t-1}	$\hat{\gamma}_1$	$s.e(\hat{\gamma}_1)$	R^2
FC residual \hat{v}_{t-1}	-0.0616	0.0015	0.0017
$\sum_{i=1}^2 \hat{v}_{t-i}$	-0.0267	0.0009	0.0103
$\sum_{i=1}^3 \hat{v}_{t-i}$	-0.0086	0.0151	0.0000
$\sum_{i=1}^4 \hat{v}_{t-i}$	-0.0167	0.0006	0.0140
$\sum_{i=1}^5 \hat{v}_{t-i}$	-0.0157	0.0002	0.0304
TVC residual \hat{w}_{t-1}	-0.1213	0.0171	0.2310
$\sum_{i=1}^2 \hat{w}_{t-i}$	-0.1181	0.0082	0.2630
$\sum_{i=1}^3 \hat{w}_{t-i}$	-0.1265	0.0288	0.0022
$\sum_{i=1}^4 \hat{w}_{t-i}$	-0.1017	0.0560	0.0801
$\sum_{i=1}^5 \hat{w}_{t-i}$	-0.1351	0.0288	0.2391

Note: $s.e(\hat{\gamma}_1)$ reports New-West standard error using New and West's (1994) lag selection method. TVC residuals are estimated with Park and Hahn (1995)'s regression method when coefficients of monetary index are time-varying. \hat{v}_i and \hat{w}_i are FC and TVC residuals from the monetary exchange determination regression, respectively.

the estimates of γ_1 are highly significant and negative. γ_1 can be thought of as the adjustment coefficient by which we can measure the effect of the deviation from the monetary equilibrium relationship on the movement of the exchange rate. This can imply that even though the exchange rate can deviate from the monetary equilibrium value in the short-run, the rate has the tendency to converge to the monetary equilibrium rate in the long-run. The absolute values of γ_1 and the values of R^2 from the regression with the TVC residuals used as the independent variable are larger than those from the regression with FC residuals⁶. Hence, we can say that the TVC residuals explain movements in the exchange rate better than the FC residuals.

V. EMPIRICAL RESULTS

5.1 Estimation and specification test results

The data used in the paper are quarterly. U.S. dollar/ British Pound exchange rates are drawn from the Citibank data base, measured at the end of each quarter. Other macroeconomic variables are taken from the OECD Main Economic Indicators. The chosen monetary aggregate is M1, the income measure is GDP, and the short-term interest rate is the 3-month Treasury bill rate. The monetary aggregate and the GDP series are seasonally adjusted. The estimation period is from the fourth quarter of 1973 to the second quarter of 1988.

When the transition probabilities depend on the residuals from the fixed coefficient (FC) monetary exchange model, we encountered difficulty in convergence⁷. As we can see in figure 1, the residuals from the FC monetary model seem to be quite nonstationary, so they can not explain well the stationary exchange rate changes. We do not have any convergence problems when the transition probabilities vary with the stationary residuals from the TVC monetary model.

Table 2 summarizes the estimation results of the Markov switching model when the transition probabilities depend on the residuals from the TVC monetary model. Most of the parameters are estimated quite accurately because standard errors of estimates are relatively small⁸. For comparison, we report estimates of Hamilton's (1990) model whose transition probabilities are constant. The quarterly

⁶ Estimates of R^2 in regression (14) may be larger than their true values due to the high degree of autocorrelation of the error terms.

⁷ Our convergence criterion is to stop when the maximal element of $\phi^{(s+1)} - \phi^{(s)}$ is less than 1×10^{-8} in absolute value.

⁸ Engel and Hamilton (1990) reported standard errors based on the inverse of the matrix of the second derivatives of the objective function. Following Lee's (1991) suggestion, we calculated the standard errors of $\hat{\phi}$ by $A(\hat{\phi})^{-1}B(\hat{\phi})A(\hat{\phi})^{-1}$, where $A(\hat{\phi})$ and $B(\hat{\phi})$ are the hessian matrix and the average outer product of the score vector of the incomplete log-likelihood function

estimated rates of depreciation (μ_1) and appreciation (μ_2) for the Markov model with constant transition probabilities (the FTP model) are 2.7285 and 3.6007, respectively. For the time-varying transition probabilities models (the TVTP model), the appreciation and depreciation rates are larger and smaller than those

[Table 2] Parameter Estimates for Markov Switching Model

Estimation period: 1973. 4/4 - 1988. 2/4

$y_t = e_t - e_{t-1}$ where e_t is 100 times the log of the U.S. dollar / British Pound exchange rate

Transition probability function for TVTP model

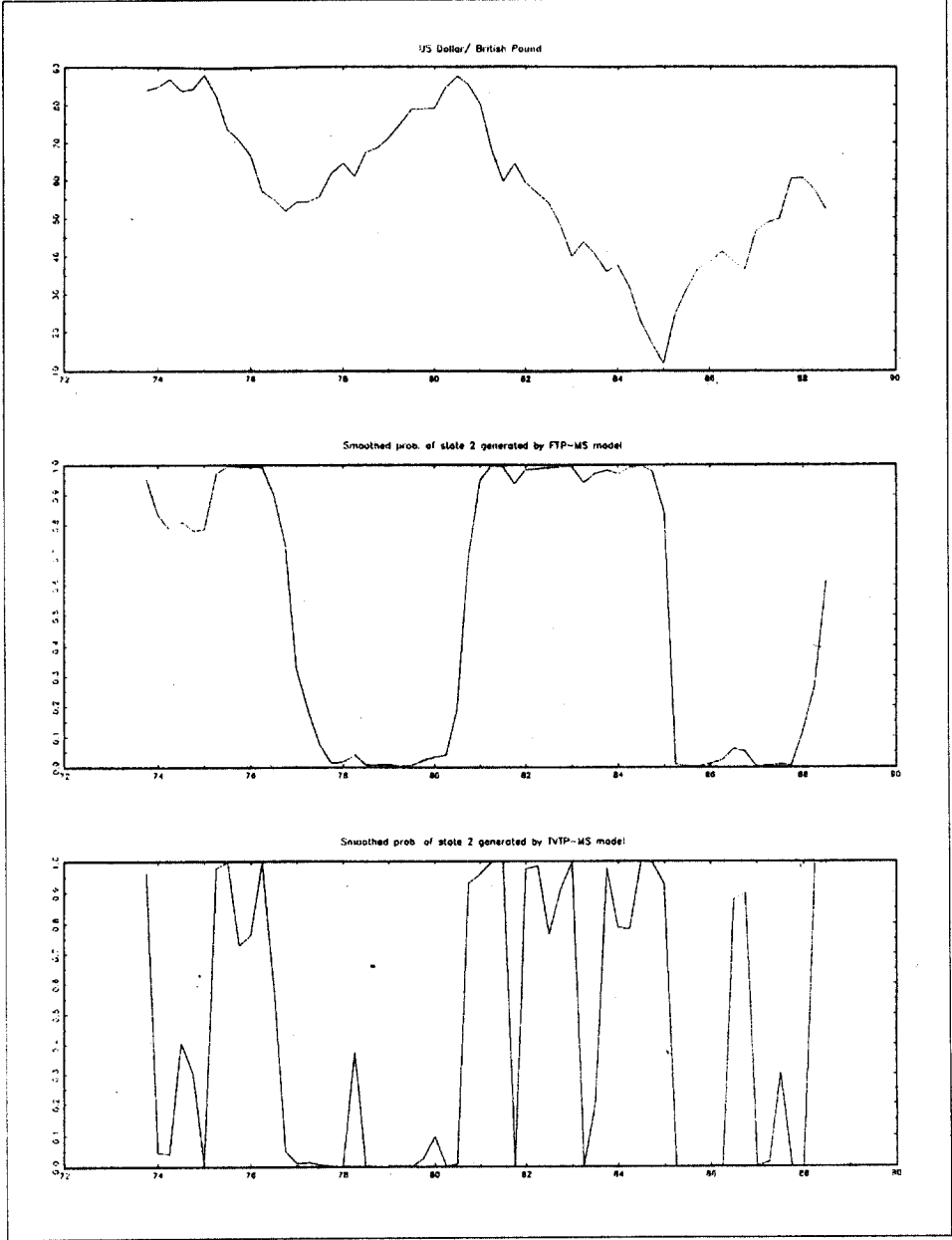
$$p_{11}^t = \frac{\exp(\alpha_1 + \alpha_2 f_{t-1})}{1 + \exp(\alpha_1 + \alpha_2 f_{t-1})} \quad p_{22}^t = \frac{\exp(\beta_1 + \beta_2 f_{t-1})}{1 + \exp(\beta_1 + \beta_2 f_{t-1})}$$

Parameter	FTP-MS model	TVTP-MS model		
		$f_{t-1} = \sum_{i=1}^2 \hat{w}_{t-i}$	$f_{t-1} = \sum_{i=1}^4 \hat{w}_{t-i}$	$f_{t-1} = \sum_{i=1}^5 \hat{w}_{t-i}$
μ_1	2.7285 (1.110)	3.9201 (1.019)	1.6220 (0.905)	2.6378 (1.013)
μ_2	-3.6007 (1.164)	-3.8853 (1.006)	-6.5685 (1.088)	-4.8632 (0.805)
ρ	0.0467 (1.958)	0.0057 (0.511)	0.8779 (3.462)	0.0184 (1.864)
σ_1^2	16.5735 (4.866)	12.0400 (4.132)	17.3961 (4.245)	15.3450 (4.557)
σ_2^2	17.4433 (4.873)	12.1238 (4.132)	7.3687 (4.693)	10.5137 (2.999)
p_{11}	0.9390 (0.085)	-	-	-
p_{22}	0.9127 (0.9127)	-	-	-
α_1	-	1.6262 (1.294)	2.9915 (0.591)	3.4199 (1.861)
α_2	-	-1.5279 (0.847)	-0.9670 (0.249)	-1.7911 (1.054)
β_1	-	2.5596 (1.843)	3.8680 (0.832)	3.7076 (2.245)
β_2	-	1.1370 (0.726)	2.9749 (0.484)	1.9086 (1.374)
log-likelihood	-121.41	-116.22	-117.02	-115.71

Note: Standard errors are in parentheses. \hat{w}_t is the TVC residual from the monetary exchange determination regression.

for the FTP model, respectively. The main reason for these results is that the smoothed probabilities calculated by the FTP model are quite different from those of the TVTP model. As we can see in equation (8a), the estimates of μ_1 and μ_2 can be considered as weighted sums of the observations where the weights are

[Figure 2] The Smoothed Probabilities of State 2



proportional to the smoothed probabilities. As figure 2 shows, there are two shorter periods of depreciation during dollar appreciation period from 1980 to 1985. To be specific, the value of the U.S. dollar against the British Pound declined around the 4th quarter of 1981 and the 2nd quarter of 1985. The FTP model calculates the smoothed probability of dollar depreciation at the 4th quarter of 1981 as 0.06 whereas the TVTP model in which the transition probability depends on the sum of the residuals from $t-1$ to $t-5$ produces the probability as 0.98. Hence, actual data (5.71 % depreciation) of the 4th quarter of 1981 is given a nearly zero weight in constructing the estimate of μ_2 in the FTP model, while the data is given a weight of 0.94 for getting μ_2 in the TVTP model. As a result, the estimate of the depreciation rate in the TVTP model is bigger than that in the FTP model.

Estimates of p_{11} and p_{22} , the transition probabilities, in the FTP model are 0.9390 and 0.9127, respectively. These probabilities can replicate the phenomenon of long-swings, which is one of the most important empirical characteristics of the foreign exchange rate. These estimates also imply that if the exchange rate appreciates (depreciates), it is very likely to continue to appreciate (depreciate), irrespective of the foreign exchange market situations or what agents expect of the future path of the economic fundamentals driving the exchange rate. However, we expect that the likelihood of the exchange rate to depreciate (appreciate) will increase, if the rate were overvalued (undervalued) relative to fundamental-based equilibrium value. Our expectation will be correct only when the probability from depreciation to depreciation (p_{11}) is a decreasing function of f_{t-1} , and the probability from appreciation to appreciation (p_{22}) is an increasing function of f_{t-1} . The estimation results of the TVTP model shows that $\frac{\partial p_{11}^t}{\partial f_{t-1}} < 0$ and $\frac{\partial p_{22}^t}{\partial f_{t-1}} > 0$, since $\alpha_2 < 0$ and $\beta_2 > 0$. This confirms our expectation: when the exchange rate is undervalued and has depreciated at date $t-1$, then the probability of a regime shift at t will be high. When the exchange is undervalued and has appreciated at date $t-1$, then the probability of appreciation at t will be high. This is also consistent with the empirical results of Mark (1994) and Chinn and Meese (1995), where the amount by which exchange rates deviate from these monetary equilibrium values can explain future movements of exchange rates.

Table 3 reports results of specification tests for the TVTP model. An alternative to the two-state Markov model is the random walk. We would like to test the null hypothesis that the exchange rate follows a random walk against the alternative of a two-state Markov switching process. We could represent the null hypothesis as the claim that $\mu_1 = \mu_2$. As Engel and Hamilton (1990) note, the standard regularity conditions for establishing asymptotically valid test of the null hypothesis do not hold. Under the null hypothesis, the parameters p_{11} and p_{22} are unidentified and the information matrix is singular. Engel and Hamilton (1990) and Engel (1994) sidestep the problem by using other possible null

hypotheses.

[Table 3] Specification Tests of the Markov Switching Models

Estimation period: 1973. 4/4-1988. 2/4

Test	FTP model	TVTP model		
		$f_{t-1} = \sum_{i=1}^2 \hat{w}_{t-i}$	$f_{t-1} = \sum_{i=1}^4 \hat{w}_{t-i}$	$f_{t-1} = \sum_{i=1}^5 \hat{w}_{t-i}$
Wald test for $\mu_1 = \mu_2$	27.6117 (0.000)	60.4303 (0.000)	67.8579 (0.000)	60.5379 (0.000)
Wald test for $\sigma_1^2 = \sigma_2^2$	0.0175 (0.894)	0.0002 (0.988)	2.7788 (0.095)	0.8113 (0.367)
LR test for FTP model	-	10.38 (0.005)	8.78 (0.012)	11.40 (0.003)

Note: the two Wald test statistics are asymptotically $\chi^2(1)$ and the LR test statistics are asymptotically $\chi^2(2)$. Asymptotic p -values are in parentheses. \hat{w}_t is the TVC residual from the monetary exchange determination regression.

They use $H_0 : \mu_1 = \mu_2^9$ as the first null. Under this null, the two states can still be identified by their variances, since the variance is a function of the state $\sigma_{s_t}^2$, so the exchange rate follows a random walk with heteroskedastic errors. It is evident from table 3 that we can strongly reject this null hypothesis for all Markov switching models considered.

If the values of p_{11} and p_{22} are large, then the exchange rate movements can be characterized by long swings. Following this idea, they use $H_0 : p_{11} + p_{22} = 1^{10}$ as the second null. Under this null, the distribution of s_t is independent of s_{t-1} . Table 3 show that this null is rejected for the TVP Markov model at the 1% significance level. We therefore, conclude that the Markov switching model with two states explains better the behavior the exchange rate than a model with a single state.

Time-variation of the transition probabilities can be tested by the likelihood ratio test of $H_0 : \alpha_2 = 0, \beta_2 = 0$ versus $H_0 : \alpha_2 \neq 0, \beta_2 \neq 0$. This is a joint test of the appropriateness of the functional form of the transition probability (the logistic function) and the statistical significance of the coefficients on the fundamental variable f_{t-1} . The LR test statistics under the null hypothesis are asymptotically distributed as $\chi^2(2)$. We can reject the null hypothesis (the constant transition

⁹ A Wald statistic for testing the null is given by $\frac{(\hat{\mu}_1 - \hat{\mu}_2)^2}{\hat{var}(\hat{\mu}_1) + \hat{var}(\hat{\mu}_2) - 2 \cdot \hat{cov}(\hat{\mu}_1, \hat{\mu}_2)} \approx \chi(1)$

¹⁰ A Wald statistic for testing the null is given by $\frac{(\hat{p}_{11} - (1 - \hat{p}_{22}))^2}{\hat{var}(\hat{p}_{11}) + \hat{var}(\hat{p}_{22}) - 2 \cdot \hat{cov}(\hat{p}_{11}, \hat{p}_{22})} \approx \chi(1)$

probability) at the 5% significance level. This test result means that the behavior of the dollar/Pound exchange rate can be described well by the Markov switching model, with the transition probabilities varying according to the magnitude of the deviation of the exchange rate from a monetary equilibrium value. This is also supported by a comparison of the smoothed probabilities of the appreciation state calculated by the FTP and the TVTP model, respectively. As we can see in figure 2, both the appreciation and the depreciation periods are identified more accurately in the TVTP model than in the FTP model.

5.2 Forecasting performance

The forecast period is from the third quarter of 1988 to the fourth quarter of 1994. Forecasts are generated at horizons of one, two, three and four quarters¹¹. In order to get these forecasts, we use estimates of the parameters with data up to the second quarter of 1988. The filtered probabilities and the transition probabilities which are based on estimates of the parameters are updated with the addition of each new data point. As a result, only one-step ahead forecast is ex-ante.

The forecasting performance of the relevant models is reported in tables 4 and 5. According to the RMSE (root mean squared error) evaluation metric, the out-of-sample forecasting abilities of the Markov switching models and a random walk model are compared in table 4. *OUT/RW* denotes the ratio of the RMSE of the Markov switching model to that of the random walk model without drift. Diebold-Mariano test statistics are employed to test the equality of the RMSE of a Markov switching model and that of the driftless random walk model. Under the null hypothesis of the equality of the RMSE's, Diebold-Mariano test statistics are asymptotically distributed as the standard normal. The predictive performance of our Markov switching models is not significantly better than the driftless random walk at the 1-quarter horizon. However, our Markov switching models display significant improvement in predictive accuracy at horizons beyond one quarter. For two, three and four quarter horizon, results of the Diebold-Mariano test show that our Markov switching models generate better forecasts than the driftless random walk model at the 10% significance level. The TVTP Markov

¹¹ The Markov switching model can generate forecasts in the following way. Let $q_{ij}^t(k)$ be the probability of going from state i to state j in k steps from t , and $\eta_i^t(k)$ be the probability of being in state i k steps ahead from t . The k -step ahead forecast is given by:

$$\begin{aligned} y_{t+k|t} &= [\eta_1^t(0), \eta_2^t(0)] \begin{bmatrix} p_{11}^{t+1} & p_{12}^{t+1} \\ p_{21}^{t+1} & p_{22}^{t+1} \end{bmatrix} \cdots \begin{bmatrix} p_{11}^{t+k} & p_{12}^{t+k} \\ p_{21}^{t+k} & p_{22}^{t+k} \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} \\ &= [\eta_1^t(0), \eta_2^t(0)] \begin{bmatrix} q_{11}^t(k) & q_{12}^t(k) \\ q_{21}^t(k) & q_{22}^t(k) \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} \\ &= \sum_{i=1}^2 \sum_{j=1}^2 \eta_i^t(k-1) \cdot p_{ij}^{t+k} \cdot \hat{\mu}_j \end{aligned}$$

$$\text{where } \eta_i^t(k) = \sum_{j=1}^2 \eta_j^t(k-1) \cdot p_{ji}^{t+k} \quad k=1, 2, 3, 4 \dots$$

model generates better one-quarter ahead forecast than the FTP Markov model.

[Table 4] Out-of-sample Forecast Performance

Forecast period: 1988:3~1994:4

	forecast horizon			
	1	2	3	4
FTP model				
OUT/RW	1.1077	0.7324	0.6664	0.6005
DM	2.1740 (0.985)	-1.6021 (0.054)	-1.4930 (0.067)	-1.4644 (0.071)
TVTP				
$f_{t-1} = \sum_{i=1}^2 \hat{w}_{t-i}$	0.9816	0.6244	0.7287	0.6157
OUT/RW	-0.2983	-1.8879	-1.3534	-1.4699
DM	(0.382)	(0.029)	(0.087)	(0.070)
$f_{t-1} = \sum_{i=1}^3 \hat{w}_{t-i}$	1.0734	0.7686	0.6039	0.5413
OUT/RW	1.3378	-1.4943	-1.6216	-1.5216
DM	(0.909)	(0.067)	(0.052)	(0.064)
$f_{t-1} = \sum_{i=1}^5 \hat{w}_{t-i}$	0.8995	0.7036	0.7164	0.5261
OUT/RW	-0.9196	-1.7736	-1.2900	-1.5008
DM	(0.178)	(0.038)	(0.098)	(0.066)

Notes: OUT/RW is the ratio of the root mean squared error for Markov-switching models out-of-sample forecasts(OUT) to that for the driftless random walk(RW). DM is the Diebold-Mariano statistics where the long-run variance is calculated using the Parzen window with the truncation lag determined by Andrews(1990) AR(1) rule.

[Table 5] Number of Correct Forecasts of Direction of Change

Forecast period: 1988:3~1994:4

Model	Horizon (number of Forecast periods)			
	1(25)	2(24)	3(23)	4(24)
FTP-model	8	7	6	6
TVTP-model				
$f_{t-1} = \sum_{i=1}^2 \hat{w}_{t-i}$	16	11	8	7
$f_{t-1} = \sum_{i=1}^4 \hat{w}_{t-i}$	11	11	15	16
$f_{t-1} = \sum_{i=1}^5 \hat{w}_{t-i}$	16	10	12	18

In general, the accurate prediction of the exchange rates in terms of the RMSE is not as important as the correct prediction of the direction of the change in the exchange rates to investors in foreign exchange markets. The profit of open positions in foreign exchange markets depends on the direction of the change in the relevant exchange rate. If there are long swings in the data generating process of the exchange rate, the Markov model is able to generate accurate forecasts of the direction of change. Engel (1994) explained the reason for the Markov model to perform well in terms of getting the direction of change correct as follows:

Because of long-swings in the exchange rate, there will be runs in one direction and then the other in changes in the exchange rate. The runs do not have to persist for a long time before the Markov switching model concludes that the state has shifted. Hence, it will miss the direction of change for a short time around the date at which the regime shift, but will tend to get the direction of change correct during long periods of time in which the exchange rate drifts in one direction.

Table 5 reports the count on how many times each model forecasts the correct direction. We do not consider the driftless random walk model in this comparison. For all forecast horizons considered, the TVTP Markov models predict the direction better than the FTP Markov model. This implies that the TVTP Markov model can detect the direction of change for a short period of regime shift better than the FTP Markov model. Hence, the deviation from a monetary equilibrium relationship may be useful information for predicting the regime shift of the exchange rate. When the size of the forecast error in terms of the RMSE is de-emphasized, the TVTP Markov model seems to be more attractive than the FTP Markov model to investors in foreign exchange markets.

VI. CONCLUSION

We use Lee's (1991) Markov switching model with time varying transition probabilities (the TVTP Markov model) to analyze the behavior of the U.S. dollar/British Pound exchange rate under the recent floating exchange rate system. Lee (1991) extended Hamilton's (1990) Markov model to incorporate time-varying transition probabilities between states, and developed an EM algorithm to estimate his extended Markov model. We employ the magnitude of the deviation of the exchange rate from a monetary equilibrium value as the economic fundamental with which the transition probabilities vary.

The empirical results we obtained from this paper are the following: first, the deviation of the exchange rate from a monetary equilibrium value can affect the transition probabilities of the exchange rate from one state to another state. When the exchange rate is overvalued (undervalued) relative to the monetary equilibrium value, the probability of the exchange rate appreciating (depreciating) will be low (high). Second, the TVTP Markov model can identify both the appreciation state and the depreciation state of the exchange rate better than the Markov model with

fixed transition probabilities (the FTP Markov model). Third, the forecasts of the TVTP Markov model are superior at predicting the direction of change of the exchange rate to those of the FTP Markov model.

REFERENCES

- Andrews, D. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, Vol. 59, 817-858.
- Chinn, M. D. and R. Meese (1995), "Banking on Currency Forecasts: How Predictable is Change in Money?," *Journal of International Economics*, Vol. 38, 161-178.
- Dempster, A., Laird, N. and Rubin, D. (1977), "Maximum Likelihood from Incomplete Data Via EM Algorithm," *Journal of Royal Statistical Society, Series B*, 39, 1-38.
- Diebold, F. X., Lee, J. H. and Weinbach, G. C. (1994), "Regime Switching with Time-Varying Transition Probabilities," in *Nonstationary Time Series Analysis and Cointegration*, ed. C.Hargreaves, Oxford University Press.
- Diebold, F. X. and R. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, Vol. 13, 253-263.
- Engel, C. (1994), "Can the Markov Switching Model Forecast Exchange Rates?," *Journal of International Economics*, Vol. 36, 151-165.
- Engel, C. and J. Hamilton (1990), "Long Swing in the Dollar: Are They in the Data and Do Markets Know It?," *American Economic Review*, Vol. 80, 689-713.
- Filardo, A. J. (1994), "Business-Cycle Phases and Their Traditional Dynamics," *Journal of Business & Economic Statistics*, Vol. 12, 299-308.
- Garcia, R. and Perron, P. (1990), "An Analysis of Real Interest Rate Under Regime Shifts," Mimeo, Princeton University.
- Hamilton, J. D. (1990), "Analysis of Time Series Subject to Changes in Regime," *Econometrica*, Vol. 45, 39-70.
- Kaminsky, G. (1993), "Is There a Peso Problem? Evidence from Dollar/Pound Exchange rate," *American Economic Review*, Vol. 83, 450-472.
- Kim, B. H. and B. C. Yu (1996), "Can Time-Varying Coefficient Improve the Empirical Performance of Monetary Exchange Rate Determination Models?," Mimeo, Inje University, Korea.
- Lee, J. H. (1991), "Nonstationary Markov Switching Models of Exchange Rates: the Pound-Dollar Exchange Rate," Ph.D. Dissertation, University of Pennsylvania.
- Maddala, G. S. (1992). "Regime Switching Models," Mimeo, the Ohio State University.
- MacDonald, R. and M. Taylor (1994), "The Monetary Model of the Exchange Rate: Longrun Relationship, Short-run Dynamics and How to Beat a Random Walk," *Journal of International Money and Finance*, Vol. 13, 276-290.
- Mark, C. N. (1995), "Exchange Rates and Fundamentals: Evidence on Long-horizon Predictability," *American Economic Review*, Vol. 85, 201-218.
- Meese, R. and K. Rogoff (1983), "Empirical Exchange Rate Models of the 1970's: Do They Fit Out of Sample?," *Journal of International Economics*, Vol. 14,

3-24.

- Newey, W. K. and K. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Vol. 55, 703-708.
- Newey, W. K. and K. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, Vol. 61, 631-653.
- Park, J. Y. (1992), "Canonical Cointegrating Regressions," *Econometrica*, Vol. 60, 119-143.
- Park, J. Y. and S. Hahn (1995), "Cointegration Regression with Time Varying Coefficients," Working Paper.
- Raymond, J. E. and Rich, R. W. (1992), "Changes in Regime and the Behavior of Inflation," Manuscript, Vanderbilt University.