Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model

Fabio Canova*

Brown University, Providence, RI 02912, USA
European University Institute, I-50016 San Domenico di Fiesole, Italy

This paper employs a multivariate Bayesian time-varying coefficients (TVC) approach to model and forecast exchange rate data. It is shown that, if used as a data-generating mechanism, a TVC model induces nonlinearities in the conditional moments and leptokurtosis in the unconditional distribution of the series. It is also shown that leptokurtic behavior disappears under time aggregation. As a forecasting device, a Bayesian TVC model improves over a random walk model. The improvements are robust to several changes in the forecasting environment.

1. Introduction

It is widely recognized that short-term changes in a variety of asset prices are well described by linearly unpredictable stochastic processes which are unconditionally leptokurtic and conditionally heteroskedastic. Studies by Meese and Rogoff (1983a, b), Domowitz and Hakkio (1985), Boothe and Glassman (1987), Hsieh (1988), and Diebold (1988) have documented that for the case of exchange rate changes (1) the conditional mean is approximately linear and nonlinearities, if they exist, arise in the even conditional moments of the process and (2) the current value of the process summarizes all the information that is relevant to forecast future values. One implication of

these findings was that a random walk (RW) model appeared to be the best tool for out-of-sample point forecasts.

The lack of predictability in asset price changes, however, is increasingly being questioned both at a theoretical and an empirical level. Theoretically, Sims (1984) argues that any price which is strongly anticipatory of future events will behave like a martingale only as the sampling interval goes to zero. Empirically, Engle, Lilien, and Robbins (1987), Pagan and Yong (1990), and Engel and Hamilton (1990), among others, have demonstrated the existence of nonlinear dependencies in the conditional mean of certain asset prices using data at various frequencies.

In an attempt to exploit these nonlinear dependencies in forecasting exchange rate changes, Diebold and Nason (1990) examine the performance of a class of univariate models where the conditional mean is allowed to be a nonlinear function of the available information set. They employ nonparametric methods to construct an estimate of the conditional mean which is, under a quadratic loss function, the best predictor of future exchange rate changes. Their results are mixed: the particular form of nonlinear dependence they employ does not necessary improve the out-of-sample forecasting performance of linear models (and of a RW model in particular). Mizrach (1990), using an analogous but multivariate technique and a different sampling interval, improves over a RW model in only one exchange market. Similarly, Meese and Rose (1989) show that the same class of models has only a slightly better in-sample fit than linear models over a variety of samples and situations.

This paper contributes to this emerging literature in two ways. I provide a parametric model which, if used as a data-generating mechanism, produces complex nonlinear dependencies in the conditional moments of the generated time series and can replicate several important empirical features of exchange rate data [see Gallant, Hsieh, and Tauchen (1990) for a similar undertaking using Clark's (1973) 'mixture' model]. I then demonstrate that the forecasts of this model substantially improve upon the forecasts of a RW model.

The approach of the paper differs from previous efforts in several respects. First, a Bayesian time-varying coefficient (TVC) autoregressive model of the variety pioneered by Litterman (1982), Doan, Litterman, and Sims (1984), and Sims (1989) is employed. Second, a number of foreign exchange and interest rate markets are simultaneously considered in the forecasting exercise. This approach has the advantage of employing a richer information structure and exploiting common features existing in a variety of financial markets to construct forecasts [see Diebold and Nerlove (1989) for a similar argument].

I show that the TVC methodology has several appealing features for modelling financial data. First, a TVC model is flexible enough to include as

special cases several nonlinear time series specifications recently employed to model the first two conditional moments of exchange rates and of other financial data [see Tsay (1987) for some of these relationships]. Second, if used as a data-generating mechanism (DGM), the model produces nonnormalities in the endogenous time series by means of time variation in the coefficients. Time variation was found to be a relevant ingredient in explaining the leptokurtic behavior of exchange rate data [Friedman and Vandersteel (1982), Boothe and Glassman (1987)] and in forecasting exchange rates with structural models [Wolff (1987), Schinasi and Swamy (1989)]. Third, the generated time series have the property that time aggregation reduces nonnormalities [see Diebold (1988) for this feature of exchange rate data]. Fourth, since the model has a linear recursive conditionally Gaussian structure [Lipster and Shiryayev (1978)], simple estimation and forecasting procedures are applicable to a vector of time series. Finally, as emphasized by Engle and Watson (1985), TVC models have a Bayesian interpretation. Therefore the information contained in various foreign exchange and interest rate markets can be pooled with a simple Bayesian procedure.

I also show that as a forecasting tool, a Bayesian TVC model is useful and improves upon RW forecasts at all horizons and for all the exchange rates considered using three different loss functions. The improvement is statistically significant, essentially independent of the time period chosen to compute forecasts and robust to various alterations in the model specification.

The rest of the paper is organized as follows: the next section describes a TVC model and examines the statistical properties of time series generated with such a model. Section 3 presents details on the forecasting exercise. Section 4 presents the forecasting results and a robustness analysis. Section 5 contains the conclusions.

2. A statistical model for exchange rate data

Consider the following data-generating mechanism (DGM):

$$y_t = \beta_t' x_t + u_t, \tag{1}$$

where u_t, conditional on x_t, is white noise with variance V. For the sake of presentation, I assume that y_t is a scalar stochastic process,[1] β_t' is a $1 \times (l+1)$ vector of time-varying coefficients of the model, and x_t is a $(l+1) \times 1$ vector of lagged dependent or exogenous variables. I also assume that β_t evolves according to

$$\beta_t = G\beta_{t-1} + F\beta_0 + Ae_t, \tag{2}$$

---

[1]None of the results presented in this section depend on the scalar specification assumption.

where e_t is a vector of disturbances and G, F, and A are conformable, positive definite matrices.

2.1. Nonlinearities in the conditional moment structure: Some special cases

To demonstrate the flexibility of (1)–(2) as a DGM, substitute (2) into (1) to obtain

$$y_t = \hat{\beta}'_t x_t + v_t, \tag{3}$$

where $\hat{\beta}_t = G\beta_{t-1} + F\beta_0$ and $v_t = (Ae_t)'x_t + u_t$. Given auxiliary assumptions regarding the nature of x_t and the relationship between x_t and e_t, the present framework encompasses several parametric nonlinear time series specifications often used to model exchange rate data. For example, if x_t and e_t are conditionally independent and e_t has variance Σ, then y_t is a conditionally heteroskedastic process with mean $\hat{\beta}'_t x_t$ and variance $\Omega_t = V + x'_t A' \Sigma A x_t$. In addition, if x_t includes lagged dependent variables and a constant and $(e_t | x_t, \mathcal{F}_t) \sim \mathcal{N}(0, \Sigma)$, where $\mathcal{F}_t$ is the σ-algebra generated by past values of y_t, then the TVC model generates a conditionally normal ARMA-ARCH model [see Weiss (1984) for the theory of ARMA-ARCH models and Hsieh (1989) for an application to exchange rates]. However, if x_t includes latent variables or variables which are not perfectly predictable given $\mathcal{F}_t$, then, conditional on $\mathcal{F}_t$ alone, y_t will be a non-Gaussian heteroskedastic process. Therefore, Clark's (1973) mixture model, recently employed by Gallant, Hsieh, and Tauchen (1990) to model the daily British pound/US dollar exchange rate, is a special case of a TVC model. On the other hand, if some components of x_t and e_t are conditionally dependent (say, x_{it} and e_{jt}, for $i = j$) with $E_{t-1}(e'_t A' x_t) = \Sigma_i \alpha_{ii} \overline{e_{it} x_{it}}$, then y_t has conditional mean equal to $\hat{\beta}'_t x_t + \Sigma_i a_{ii} \overline{e_{it} x_{it}}$ and conditional variance equal to $V + (\Sigma_{j \neq i} \Sigma_i a_{ji} \overline{e_{jt} x_{it}})^2$. Therefore, a version of a TVC model where x_t includes, e.g., stochastic exogenous variables correlated with the innovations in β_t, produces a version of the bilinear model of Granger and Anderson (1978) as a special case.

A TVC model can also generate something similar to an ARCH-M model [see Engle, Lilien, and Robbins (1987) and Domowitz and Hakkio (1985) for an application to exchange rates]. If $e_t = e_{1t} + e_{2t}$, where e_{1t} is independent of x_t and e_{2t} and has covariance matrix Σ, and where e_{2t} is perfectly dependent on x_t, then the conditional mean of y_t is $\hat{\beta} x_t + x'_t A x_t$ and conditional variance is $V + x'_t A' \Sigma A x_t$.

Finally, a version of Hamilton's (1989) two-state discrete shift model, recently employed by Engel and Hamilton (1990), can be cast into this

framework by choosing

$$x_t' = [a_0, a_1, 1, 0, 0, \ldots, 0],$$

$$\beta_t' = [1, S_t, z_t, z_{t-1}, \ldots, z_{t-r}],$$

$$\beta_0' = [1, E_0 S_0, E_0 z_0, \ldots, E_0 z_0],$$

$$e_t' = [0, v_t, \varepsilon_t, 0, \ldots, 0].$$

F is a zero matrix except for the element $f_{22} = (1-q)/\pi_0$ and G has the following structure:

$$G = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \mu & 0 & \cdots & 0 \\ 0 & 0 & 0 & & \\ \vdots & \vdots & \vdots & R & \\ 0 & 0 & 0 & & \end{bmatrix},$$

where R is an unrestricted matrix, $A = I$, and $u_t = 0$ $\forall t$. Here S_t is an unobservable two-state Markov process with AR representation

$$S_t = (1-q) + \mu S_{t-1} + v_t, \qquad \mu = p + q - 1, \tag{4}$$

where p and q are the diagonal elements of the transition matrix, $\pi_0 = E_0 S_0$, and z_t is an AR(r) process with innovations ε_t.

In conclusion, the TVC model described by eqs. (1) and (2) can accommodate several parametric models currently used to characterize nonlinearities in the first two conditional moments of exchange rate data. Obviously, if e_t has a degenerate distribution for all t, the model reduces to the simple homoskedastic linear model with constant coefficients.

2.2. Nonnormalities in the generated data

Since the TVC model includes Clark's (1973) mixture model as a special case, one could follow the steps of Gallant, Hsieh, and Tauchen (1990) and show that the model generates nonnormal heteroskedastic time series if x_t is serially dependent and unobservable. However, these features emerge from a TVC model even when x_t is observed at t. To see this let $x_t = [y_{t-1}, y_{t-2}, \ldots, y_{t-L}]$, and u_t and e_t be normally and independently distributed random variables, and let e_t and x_t be conditionally independent.[2]

---

[2]This last assumption is not crucial for the results. If we assume that x_t and e_t are conditionally dependent, then, independent of k, y_t is non-Gaussian.

Define

$$\hat{u}_{t+k} = \left(G^{k+1}\beta_{t-1} + F\beta_0 \sum_{j=0}^{k} G^j \right)' (x_{t+k} - \mathrm{E}_{t-1}x_{t+k})$$

$$+ \left(\sum_{j=0}^{k-1} G^{k-j}Ae_{t+j} \right)' x_{t+k} - \mathrm{E}_{t-1}\left(\sum_{j=0}^{k-1} G^{k-j}Ae_{t+j} \right)' x_{t+k}$$

$$+ e'_{t+k}A'x_{t+k} + u_{t+k}.$$

It follows that for fixed t and all k:

$$\mathrm{E}_{t-1}y_{t+k} = \left(G^{k+1}\beta_{t-1} + F\beta_0 \sum_{j=0}^{k} G^j \right)' \mathrm{E}_{t-1}x_{t+k}$$

$$+ \mathrm{E}_{t-1}\left(\sum_{j=0}^{k-1} G^{k-j}Ae_{t+j} \right)' x_{t+k}, \tag{5}$$

$$\mathrm{var}_{t-1}y_{t+k} = \mathrm{E}_{t-1}(\hat{u}_{t+k})^2, \tag{6}$$

$$sk_{t-1}y_{t+k} = \mathrm{E}_{t-1}(\hat{u}_{t+k})^3 \big/ (\mathrm{var}_{t-1}y_{t+k})^{3/2}, \tag{7}$$

$$ku_{t-1}y_{t+k} = \mathrm{E}_{t-1}(\hat{u}_{t+k})^4 \big/ (\mathrm{var}_{t-1}y_{t+k})^2, \tag{8}$$

where sk and ku are the skewness and the kurtosis coefficients. For $k = 0$, it is easy to verify that $sk_{t-1}y_t = 0$, $ku_{t-1}y_t = 3$, so that, conditional on $\mathscr{F}_t$, y_t is normal with time-varying conditional mean $\mathrm{E}_{t-1}y_t = \hat{\beta}'x_t$ and conditional variance Ω_t. For $k = 1$ the conditional mean of y_{t+1} is nonlinear and equal to

$$\mathrm{E}_{t-1}(\beta'_{t+1}x_{t+1}) = \left(G^2\beta_{t-1} + F(I+G)\beta_0 \right)'\mathrm{E}_{t-1}x_{t+1}$$

$$+ \mathrm{E}_{t-1}e'_t A'Gx_{t+1},$$

where $\mathrm{E}_{t-1}x_{t+1} = [\mathrm{E}_{t-1}y_t, y_{t-1}, \ldots, y_{t-l+1}]$, while its conditional variance is given by

$$\mathrm{E}_{t-1}(\hat{u}_{t+1})^2 = \mathrm{E}_{t-1}\big((G^2\beta_{t-1} + F\beta_0(1+G))'(x_{t+1} - \mathrm{E}_{t-1}x_{t+1})$$

$$+ (e'_t A'G'x_{t+1} - \mathrm{E}_{t-1}e'_t A'G'x_{t+1})$$

$$+ e'_{t+1}A'x_{t+1} + u_{t+1} \big)^2.$$

Note that $(x_{t+1} - E_{t-1}x_{t+1})' = [e_t'A'x_t + u_t, 0, 0, \ldots, 0]$ and $((e_t'A'G'x_{t+1}) - E_{t-1}(e_t'A'G'x_{t+1}))$, involve, among other things, terms of the form $e_t'A'G'u_t$. Therefore, even though e_t and u_t are normal and independent, y_{t+1} is nonnormal because the prediction error $\hat{u}_{t+1}$ involves the product of normal random variables, none of which is predictable given $\mathcal{F}_t$.

Using (5)–(8) for any $k \geq 1$, one can show that the above argument holds in general. Hence, this DGM has several potential channels to induce complex nonlinearities and nonnormalities in the data.

2.3. The unconditional moment structure and time aggregation

Boothe and Glassman (1987), Diebold (1988), Hsieh (1988), and Gallant, Hsieh, and Tauchen (1990) showed that the unconditional distribution of exchange rate changes is leptokurtic, slightly skewed to the left for certain currencies, and that it converges to a normal as data is aggregated over time. To show that a TVC model induces these properties in the time series for y_t, one should compute the third and fourth moments of the unconditional distribution of y_t. For general specifications of the model, the computation of these higher unconditional moments has proven to be too complex to be undertaken analytically.[3] Under regularity conditions on the characteristic roots of G, F, and A, the unconditional variance of y_t will be bounded. Therefore, the unconditional distribution of y_t will not belong to the family of stable Paretian distributions which have been used in the past to characterize exchange rate dynamics [see Boothe and Glassman (1987) for references].[4] Also, since the prediction errors generated by this DGM are conditionally heteroskedastic and nonnormal for any $k \geq 1$, one can guess that the unconditional distribution of y_t is leptokurtic. To illustrate the features of the unconditional distribution of y_t and the convergence to the normal under time aggregation I use a Monte Carlo simulation.

The results of the simulations are presented in figs. 1–3. Fig. 1 plots the empirical density, estimated using the procedure described in Rao (1984, p. 102), of the scalar $w_t = \log(y_t) - \log(y_{t-1})$, where y_t is generated with a fourth-order AR model with a constant [i.e., $x_t = (y_{t-1}, \ldots, y_{t-4}, c_t)$], with $\beta_0 = [0.97, 0.00, 0.00, 0.03, 0.00]$, $G = 0.999*I$, $F = 0.001*I$, $A = I$, $\Sigma = \text{diag}(\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{44}, \sigma_{cc})$, $\sigma_{ll} = (0.002/l^2, l = 1, \ldots, 4)$, $\sigma_{cc} = 0.002$, and $V = 0.002$. Both u and ε are drawn from a normal distribution. 500 points were

---

[3]Analytic results for special cases are available, e.g., in Diebold (1988) and Gallant, Hsieh, and Tauchen (1990).

[4]These distributions are conditionally heteroskedastic, unconditionally leptokurtic, possess a domain of attraction, but are invariant under time aggregation, and have infinite variance except in the normal case [see Blattberg and Sargent (1971), Fama and Roll (1968)].
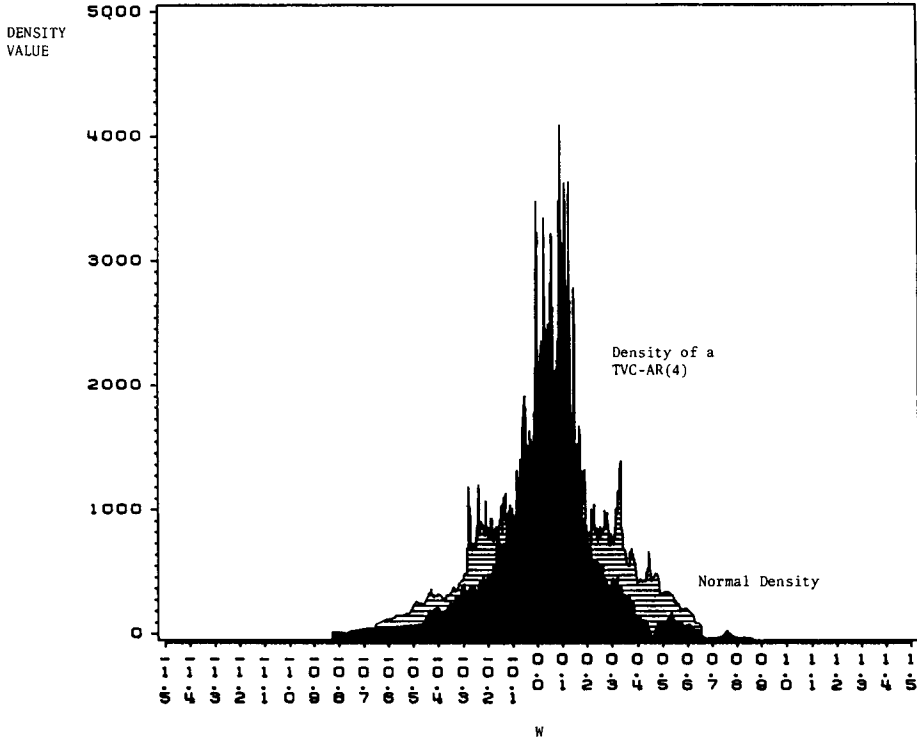
Fig. 1. Data set 1.

generated for the specific realization reported. Plotted with w_t is a normal distribution with the same mean and the same variance of w_t (dashed area). Fig. 2 plots the empirical density of a w_t series whose coefficients are subject to a structural break. To be specific I generate 250 points with the same parameters as in fig. 1 and 250 points with the same G, β_0, A, F, V, and σ_{cc} parameters, while $\sigma_{ll} = \mathrm{diag}(0.0002/l^2)$. This simulation mimics a situation where the uncertainty in the coefficients drops after a certain date. A normal distribution with the same mean and the same variance of w_t is also plotted in fig. 2. In both cases the empirical distribution of w_t is more peaked around zero and has fatter tails relative to the normal, and in the second case it is slightly skewed to the left.

Fig. 3 plots the 90% confidence band for the skewness and the kurtosis of the two specifications above when the generated series are time aggregated over various intervals. Together with the band, fig. 3 plots the skewness and the kurtosis of a normal distribution (these are equal to 0 and 3, respectively). To construct the bands I draw 1000 samples of 500 data points each for each
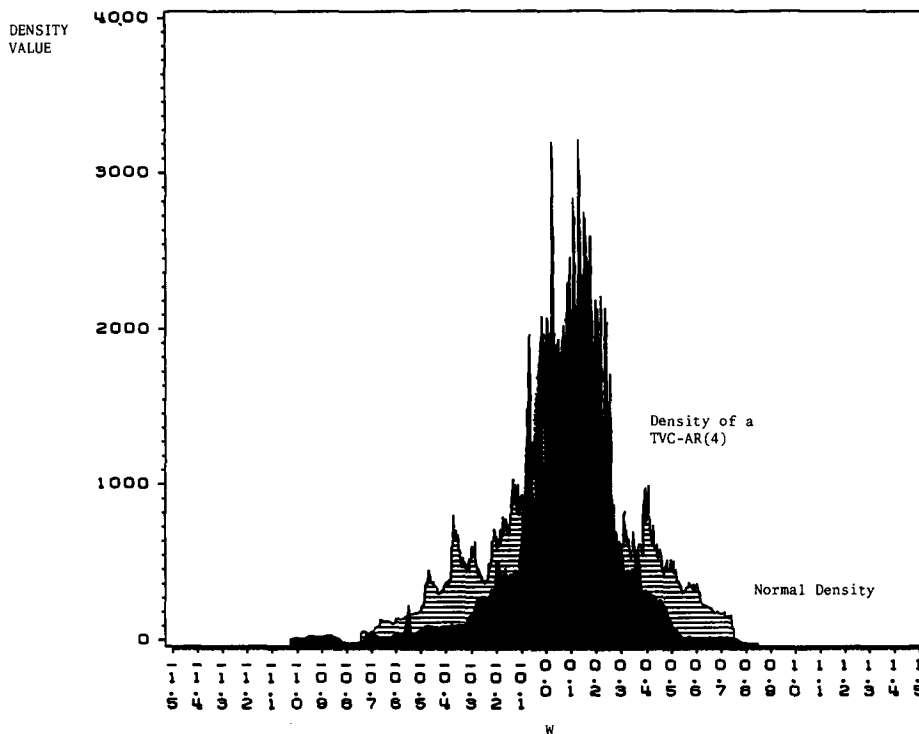
Fig. 2. Data set 2.

of the two specifications. For each sample and each specification I then aggregate every $2, 4, 8, 12, 16, 20$ observations to construct samples of $250, 125, 63, 42, 31, 25$ data points, respectively. For each of these seven samples and for each draw, I then estimate the skewness and the kurtosis, order the estimated values, and compute the 90% confidence bands for each of the two model specifications. Let $sk_{T,i}$ and $ku_{T,i}$ be the ordered estimated skewness and kurtosis for a sample size T and for draw i. Then, for each T, the 90% confidence bands for the skewness ranges from $sk_{T,50}$ to $sk_{T,950}$ and the 90% confidence bands for the kurtosis ranges from $ku_{T,50}$ to $ku_{T,950}$. As fig. 3 indicates, in both specifications the size of the band and the median value of the kurtosis decrease as T decreases. Therefore, with time aggregation, convergence to normality is likely to occur. From fig. 3 one can also see that the band for the estimated skewness for the second model specification is centered below zero and does not change very much with time aggregation.

In conclusion, the simulations show that the distribution of w_t is leptokurtic, is skewed to the left in certain cases, and converges to the normal when data is time aggregated. These features were all found to be present in
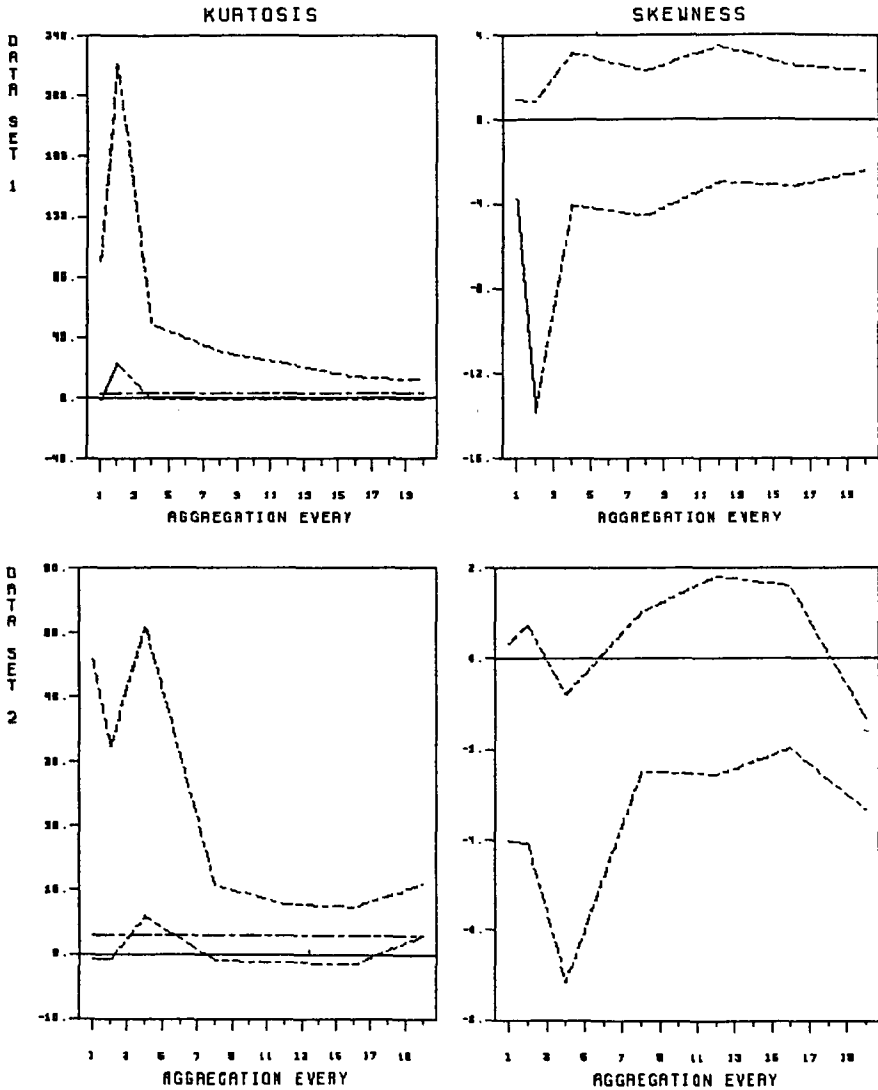
Fig. 3. 90% confidence bands.

exchange rate changes as well as in other asset price changes [see, e.g., Fama (1976)]. Earlier applications of the TVC methodology to the problem of modelling data from other asset markets also confirm the usefulness of the approach [see Shaefer et al. (1975), Rosenberg (1973)]. To check whether a TVC model captures the out-of-sample features of the exchange rate data as well, I turn next to a forecasting exercise.

3. The specification of the forecasting exercise

3.1. The data

In this exercise I employ a panel of spot exchange rates and spot interest rates. Interest rates are used because they provide information for exchange rate dynamics which is not contained in past values of the exchange rates. The data set spans a nine-year interval from the first week of 1979 to the last week of 1987. For all variables weekly sampling (at Wednesday) of daily values are constructed. The exchange rates are the quoted closing values at the New York market for five different currencies [French franc (FF), Swiss franc (SF), German mark (DM), English pound (L), and Japanese yen (YEN)] in terms of the US dollar. The six interest rates considered are short-term (13-week) Eurodeposit rates on $, FF, SF, DM, L, and YEN dominated securities computed as the average of the bid–ask spread.

3.2. A Bayesian version of the model

Throughout the rest of the paper y_t is an 11×1 vector (including five spot exchange rates and six interest rates) with the following representation:

$$y_t = \beta_t' x_t + u, \qquad u_t | \mathscr{F}_t \sim \mathscr{N}(0, V), \qquad (9)$$

$$B_t = G B_{t-1} + F\tilde{B} + A e_t, \quad e_t \sim (0, \Sigma_t), \qquad (10)$$

where β_t is an $11 \times 11(l + 1)$ matrix of coefficients and $B_t = \text{vec}(\beta_t)$ is its stacked version, $\tilde{B} = \text{vec}(\beta_0)$, x_t is an $11(l + 1) \times 1$ vector of lagged dependent and exogeneous variables, and (u_t, e_t) are white noise processes, independent of x_t, with covariance matrices V and Σ_t, respectively. For computational reasons, I restrict the lag length to 4. The covariance of e_t is specified to be time-dependent in order to capture slow but continuous changes in the distribution of the coefficients. However, for the sake of simplicity, discrete regime shifts of the type described in Sims (1989) are not considered.

Under a standard Bayesian interpretation of the model [see, e.g., Engle and Watson (1985), Ljung and Soderstrom (1987)], (10) is treated as part of the prior and $\tilde{B}$ is assumed to have a prior distribution with mean $E(\tilde{B})$ and covariance $\Sigma_0 = k * I$, where I is the identity matrix and k is either a given large number or a parameter with a prior distribution.

Litterman (1982), Doan, Litterman, and Sims (DLS) (1984), and Sims (1989) refined the Bayesian version of the TVC model in several directions.

First, since the variability of the forecasts is proportional to $n^2(l + 1)$, where n is the dimension of the y_t vector and l is the lag length of the model, point forecasts of y_{t+k}, $k = 1, 2, \ldots$, obtained from a highly parameterized or a large scale version of the model (9)–(10) tend to be very poor. DLS proposed making the matrices G, F, A, and Σ_t (describing the law of motions of the B's), $E(\tilde{B})$ and Σ_0 (describing the initial conditions) dependent on a small set of unknown parameters θ. This refinement introduces a new layer of uncertainty in the model which is of smaller dimension than the B vector, decreases the number of parameters to be estimated and possibly reduces the variability of the forecasts. Second, they assumed Σ_0 to be of the form $k * \text{diag}[h(l)f(i, j)]$, where $h(l)$ is a function describing how the prior covariance matrix becomes concentrated around the median as $l = 1, 2, \ldots, L$ increases and $f(i, j)$ describes the relative importance of variable j in equation i.[5] Finally, they specified $E(\tilde{B})$ to be zero except for the elements $E(\tilde{B}_{i,1})$, $i = 1, 2, \ldots, n$, which are assumed to be equal to one.

3.3. The specification of the prior

In this paper I follow the DLS approach but modify the specification of the dependence of Σ_t, G, F, A, $E(\tilde{B})$, and Σ_0 on the θ vector. I assume:

$$G = \theta_0 * I, \tag{11}$$

$$F = I - G,$$

$$A = I, \tag{12}$$

$$E(\tilde{B}_{il}) = \begin{cases} \theta_{i1} & \text{if } l = 1, \\ \theta_{i2} & \text{if } l = 4, \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, 2, \ldots, 11, \tag{13}$$

$$\Sigma_t = \Gamma_1 \Sigma_0 + \Gamma_2 \Sigma_{t-1}, \tag{14}$$

$$\Gamma_1 = \theta_3 * I,$$

$$\Gamma_2 = \theta_4 * I, \tag{15}$$

$$\Sigma_0 = \text{diag}(\sigma_{ijl}), \tag{16}$$

$$\sigma_{ijl} = \left(\theta_5/l^{\theta_6}\right) f(i, j)(s_i/s_j), \tag{17}$$

where i stands for equations, j for variable, and l for lag.

---

[5]Canova (1992) extends the approach of DLS to include frequency domain restrictions which make Σ_0 a nondiagonal matrix. Sims (1989) describes a 'dummy observation' procedure that tmakes Σ_0 nondiagonal as well.

The motivation for these choices is the following. With (11)–(12) I assume that the law of motion of the coefficients has a simple first-order Markov specification with decay toward the time zero value. θ_0 controls the speed of the decay. For $\theta_0 = 0$, the coefficients are random around $\bar{B}$. For $\theta_0 = 1$, they are random walks. One possible alternative, which can capture higher-dimensional Markov process, is $G = \theta_0 * C$, where C is a the companion matrix of the Markov process.

Since no information is available on the contemporaneous correlation of the elements of e_t, I choose $A = I$.

I use a mean of zero for the prior on all coefficients except the first own and fourth lag in each equation. Canova (1989) found significant four-week cycles in exchange rate data. By allowing the fourth lag to have mean different from zero, I hope to capture these patterns. However, contrary to DLS, I do not restrict $\theta_1 + \theta_2$ to be equal to one.

The dispersion matrix of the coefficients is assumed to evolve linearly over time. The idea behind this specification is that there is some form of persistence in the uncertainty surrounding the true value of the coefficients. Since the matrices Γ_1 and Γ_2 are fixed, the specification prevents discrete shifts in the coefficient vector even though it allows for a specific form of heteroskedasticity to appear. To see this, one can solve (14) and (15) backward to obtain

$$\Sigma_t = \delta_t * \Sigma_0, \qquad \delta_t = \theta_4^t + \theta_3 * \frac{1 - \theta_4^{2t-1}}{1 - \theta_4}. \tag{18}$$

For $\theta_4 = 0$, the coefficients are time-varying but no heteroskedasticity is present; for $\theta_3 = 0$, the variance of the coefficients is geometrically related to Σ_0. Finally, if both parameters are equal to zero, neither time variation nor heteroskedasticity is present.

The structure of Σ_0 is standard: θ_5 is the general tightness on the variance of the coefficients (corresponding to the parameter k in the previous subsection) and θ_6 controls the speed of the lag decay in the variance. Lower values of θ_5 imply higher concentration of the prior density of the entire coefficient vector around the prior median. Higher values of θ_6 imply smaller prior variance for the coefficients as l increases. s_i and s_j are the standard errors of variable i and j, and $f(i, j)$ is a matrix of weights, $f(i, i) = 1.0$. Here $f(i, j)$ controls the pooling of information among variables and markets. If all $f(i, j) = 0.0$, $i \neq j$, the model reduces to a panel of univariate TVC models. If all $f(i, j) = 0.0$ and θ_6 is large, the model approximates a panel of RW models with time-varying coefficients. In addition, if θ_5 is zero, the model for each variable is exactly a RW model with constant coefficients.

With the above choices the number of free parameters is still large. Thisted and Wecker (1981), Zellner and Hong (1989) showed that the forecasting performance of models like our's may improve if certain parameters are restricted to be common to all equations. For this reason I standardize $\theta_0, \theta_3, \theta_4$ to be the same in all equations, while $\theta_1, \theta_5, \theta_6$ and the elements of the $f(i, j)$ matrix are restricted to be the same in the block of exchange rates and in the block of interest rates. In particular, $f(i, j)$ will feature two tightness parameters: θ_7, which measures the relative importance of exchange rates in all equations, and θ_8, which measures the relative importance of interest rates in all equations.[6] With these restrictions the model has 11 equations and a total of 15 parameters to be estimated.

3.4. Estimation and forecasting

The model (9)–(10) with the prior (11)–(17) has a probabilistic structure of the form:

$$\mathcal{H}(y_t, \beta_t | x_t, \theta) = P(y_t | \beta_t, x_t) Q(\beta_t | \theta). \tag{19}$$

To close the model, one could specify a prior distribution for the parameter vector θ, say $R(\theta)$, apply Bayes rule to construct its posterior distribution, and use this distribution to build confidence intervals for y_{t+k} [see, e.g., Sarris (1973), Geweke (1989)]. This approach is straightforward but may be analytically intractable. As an alternative, one could integrate β_t out of (19) and obtain a function, say $\mathcal{L}(y_t | x_t, \theta)$, which, for given y_t and x_t, plays the role of the likelihood function of the θ (see DLS for this interpretation). If $R(\theta)$ is diffuse over a bounded hypercube, the posterior of θ will then be propositional to $\mathcal{L}(y_t | x_t, \theta)$ and the mode of the posterior will be proportional to the peak of the likelihood. Therefore, as a simplification, one could use the θ vector generating the model of the posterior both for inference and for constructing point forecasts of y_{t+k}. In this case, the estimation of the θ vector by maximum likelihood approximates the Bayesian procedure of choosing the vector of θ that makes the remaining portion of the prior, $Q(\beta_t | \theta)$, best agree with the data.

Since the analytic maximization of the likelihood for the problem is complicated, I employ numerical methods to obtain an estimate of θ. DLS show that the likelihood function can be written as a weighted average of one-step-ahead forecast errors with weights given by a weighted average of the recursive variance of the forecasts. Under the assumption of normality of e_t and u_t, one-step-ahead forecast errors are easily obtained with the

---

[6]For the German and the Japanese interest rates equations, θ_7 and θ_8 are allowed to be different from the standardized values.

Kalman filter algorithm once a value of θ is specified. Therefore, to numerically maximize the likelihood function, I run the Kalman filter through the given data set using a grid of θ vectors all having equal *a priori* probability.[7]

Once the 'optimal' θ vector is found, out-of-sample point forecasts for steps ranging from 1 to 52 are computed recursively and estimates of the β's are updates using the Kalman filter algorithm. For the exercises reported in section 4 forecasting statistics are collected over a two-year period. Therefore, there are 103 one-step-ahead forecasts and 52 52-step-ahead forecasts.

3.5. Comparing the forecasts of the TVC and of the RW models

To compare the forecasting performance of the TVC model to that of a RW model, I employ three different loss functions: the Schwarz (1978) criterion, the Theil U-statistics, and the ratio of the Mean Absolute Deviations (MAD) of the two models.

The Schwarz criterion compares the discounted (by the number of parameters) maximized value of the likelihood of the two specifications and can be used here because an approximate RW model is nested in the TVC model. Also since the Theil U-statistic is the ratio of the Mean Square Error (MSE) of the two models, if it obtains values greater than one the RW model dominates, while if it obtains values less than one the TVC model dominates. Although the Theil U is routinely used in these type of comparisons, it is not necessarily the most appropriate criterion when the unconditional distribution of the data has fat tails [see Blattberg and Sargent (1971), Koenker and Basset (1982)]. In this case the ratio of MAD's provides a better criterion to assess the forecasting ability of the two models.

I consider the Theil U and the ratio of MAD statistics at 1-, 13-, 52-step-ahead horizons. For each of the two statistics I also report an average over 52 steps for each variable, an average over variables at 1, 13, 52 steps and an overall average over all variables and all steps. I chose to present statistics at the 13-step-ahead (three-month) horizon because the interest rates chosen have a 13-week maturity and therefore three months is an appropriate horizon to judge the forecasting performance of the model. I also consider the 52-step-ahead (one-year) horizon to check the long-run performance of the model. Finally, averages over variables and over steps are

---

[7]The Bayesmth algorithm, written by Sims (1986), is used to locate the peak of the likelihood. The algorithm takes as inputs the value of the likelihood at a finite number of grid points and uses a spline interpolation to fit a surface to the inputted likelihood values and to provide a guess for the location of the peak(s) of the function and for the vector of θ corresponding to the peak. The algorithm then proceeds iteratively, adding the value of the likelihood generated with the θ vector corresponding to the guessed peak to the old ones, repeating the interpolation and providing a new guess for the location of the peak(s) of the function and for the vector of parameters corresponding to the guessed peak. The procedure stops when the increase in the likelihood function from the previous iteration is less than a specified amount.

employed to eliminate series or horizon-specific errors in the evaluation of the general performance of the specification [see Schinasi and Swamy (1989) for additional motivations on the use of averages over steps].

4. The results

The results of the forecasting exercise are presented in table 1. The model (9)–(10) is estimated over the 1979.1–1985.52 sample and forecasts are generated from 1986.1 up to 1987.51. The 'optimal' θ vector is presented in table 2, together with the maximized value of the likelihood and the value of the likelihood for the approximated RW model.

It is easy to see that the TVC model improves upon a RW model according to all criteria. When the Schwarz criterion is used the difference between the maximized values of the likelihood of the two models is significant. When the Theil U-statistic is used, the TVC model dominates a RW model for all

Table 1

Estimation sample 1979.1–1985.52/Forecasting sample 1986.1–1987.51.

Variable	Steps			Average at steps			Average over steps	Overall average
	1	13	52	1	13	52		
				Theil U statistics				
$/DM	0.9724	0.7574	0.5096	0.9817	0.8726	0.8314	0.7187	0.8581
$/L	0.9829	0.8172	0.7422				0.8738	
$/SF	0.9780	0.7804	0.6379				0.7060	
$/FF	0.9740	0.8295	0.5591				0.7697	
YEN/$	0.9758	0.8840	0.6890				0.7940	
IR$_{WG}$	0.9984	1.1182	1.9381				1.3418	
IR$_{UK}$	0.9851	0.7905	0.5743				0.7269	
IR$_{SW}$	0.9817	1.0569	0.6270				0.8647	
IR$_{FR}$	0.9890	0.8577	0.5744				0.7446	
IR$_{US}$	1.0130	1.1772	1.9680				1.4071	
IR$_{JAP}$	0.9479	0.5599	0.3309				0.4918	
				Ratio of MAD's				
$/DM	0.9427	0.7633	0.4050	1.0914	1.0244	0.8843	0.6650	0.9555
$/L	0.7612	0.6885	0.3359				0.5694	
$/SF	1.1812	1.1236	0.7358				0.9312	
$/FF	1.0000	0.6822	0.6626				0.7823	
YEN/$	0.9885	0.8009	0.5889				0.7545	
IR$_{WG}$	1.0521	1.1967	2.0369				1.4436	
IR$_{UK}$	1.1436	2.4739	1.3888				1.9063	
IR$_{SW}$	0.4091	0.2572	0.1320				0.2161	
IR$_{FR}$	2.4501	1.5066	1.1875				1.4219	
IR$_{US}$	1.0224	1.1832	1.9335				1.3574	
IR$_{JAP}$	0.9605	0.5919	0.3209				0.4744	

Table 2

Optimal parameter estimates and likelihood value.[a]

	TVC model	RW model	TVC model for Monte Carlo and bootstrap
θ_0	1.0	1.0	1.0
θ_1	1.0 if $i = IR_{US}, IR_{FR}$ 0.9 otherwise	1.0	1.02 if $i = ER$ 0.96 if $i = IR$
θ_2	0.1 if $i = ER$ 0.0 if $i = IR$	0.0	0.0
θ_3	0.99	0.99	0.99
θ_4	0.01	0.01	0.01
θ_5	0.01 if $i = ER$ 0.05 if $i = IR$	0.000001	0.01
θ_6	3.0 if $i = IR_{FR}, IR_{US},$ $IR_{JAP}, \$/YEN$ 0 otherwise	10.0	1.0
θ_7	0.4 if $i = IR_{WG}$ 0.05 otherwise	0.0	0.5
θ_8	0.2 if $i = IR_{JAP}$ 1.0 otherwise	0.0	0.2
Likelihood values	10694.42	10526.97	

[a]IR = interest rate, ER = exchange rate.

variables and all horizons (except for the US dollar and German mark denominated deposit rates). The improvements are of the order of 2–3% at the one-step horizon, of 10–20% at the 13-step horizon, and of 25–50% at the 52-step horizon. According to the average Theil U over steps and variables the improvement is substantial, ranging from 13 to 30%. When the MAD criterion is used, the TVC model still outperforms a RW model for all exchange rates, while for interest rates the results are mixed.

Although these results are encouraging, they do not demonstrate the statistical superiority of the TVC model in forecasting exchange rates. Two objections can be raised. First, since estimates of the standard error for the Theil U and the ratio of MAD's are not available, differences between the two models may be statistically insignificant. Second, since the forecasting ability of the TVC model may be period-dependent, slight alterations of the conditions existing in the sample may reverse the results.

To examine the statistical significance of the forecasting improvements obtained with a TVC model, I conducted a simple Monte Carlo exercise using two different DGM's: a RW and a TVC model. In the first case I generated 11 random walks 100 times, estimated a TVC model on the synthetically generated data replications using the same θ vector which

produced table 1, and constructed the frequency distribution of the resulting Theil U and ratio of MAD statistics. To make the experiment credible, I imposed that the drift and the conditional variances of the simulated random walks must be the same as those of the five exchange rates and the six interest rates over the 1979.1–1987.51 period. In the second case, I generated 11 time series 100 times, using a TVC model and the parameters presented in the last column of table 2, and estimated a TVC model, using the 'optimal' θ vector. The reason for estimating the model with a θ vector which differs from the one used to generate the data is to check whether the forecasting ability of the model is independent of the exact choice of parameters.

Besides providing useful standard errors for the two statistics under a well-specified null hypothesis, these exercises may also tell us whether a naive rule (Theil U or ratio of MAD's greater than one) will give the correct rejection criterion with high probability.

Table 3 contains the mean value and the standard error (over replications) of the Theil U and of the ratio of MAD statistics at $1, 13, 52$ steps and its standard error for each of the 11 variables of the system when the conditional variances are estimated using nonparametric methods as in Pagan and Ullah (1988). Table 4 contains the same statistics when the DGM is a TVC model. For each step and each statistic I also report in each table the probability that a value greater than one is observed.[8]

The results indicate that when a RW is used as DGM, the mean Theil U over replications at each of the four horizons considered is greater than one in 9 out of the 11 cases. The standard errors tend to be large. However, the values presented in table 1 for the 1- and 13-steps-ahead fall in the lower 5% of the simulated distribution for four out of five exchange rates. The standard errors at the 52-step horizon are huge and the values reported in table 1 for this horizon are insignificantly different from one. For the ratio of MAD's the results are stronger. The mean value is always substantially above one and the standard errors are larger than in the case of the Theil U. However, the values presented in table 1 for this statistic for all five exchange rates at 1- and 13-steps horizon fall in the lower 5% of the simulated distribution. Again for the 52-step horizon, none of the entries of table 1 is significantly different from one. Finally, if one uses a naive rule of preferring a RW model when the Theil U (ratio of MAD's) is above one, one would reject the TVC model when a RW is the true DGM, on average, in about 78% (99%) of the cases.[9]

---

[8]Given the computational complexity of this exercise, I was forced to limit the size of the Monte Carlo to 100 draws only. On a mainframe VAX machine using RATS programs it took approximately five hours of CPU time to construct each table.

[9]I also computed forecasting statistics when the conditional variance of the DGM is estimated with a GARCH(1, 1) model. The results, which are available on request, are similar to those reported in table 3. The only difference is that the standard errors for the two statistics are somewhat smaller.

Table 3

DGM: random walk; Theil U and ratio of MAD statistics; mean, standard errors, and Pr($H > 1$).[a]

Variable		Theil U at steps			Ratio of MAD's at steps		
		1	13	52	1	13	52
RW$_1$		1.004(0.012)	1.098(0.137)	1.239(1.15)	1.481(0.08)	1.302(0.23)	1.451(0.94)
	Pr($H > 1$)	0.80	0.66	0.58	1.00	0.95	0.54
RW$_2$		1.003(0.012)	1.094(0.140)	1.264(0.95)	1.441(0.08)	1.360(0.22)	1.410(0.91)
	Pr($H > 1$)	0.81	0.64	0.55	1.00	0.89	0.50
RW$_3$		1.003(0.011)	1.099(0.148)	1.247(0.99)	1.470(0.08)	1.314(0.25)	1.427(0.82)
	Pr($H > 1$)	0.83	0.65	0.51	1.00	0.94	0.55
RW$_4$		1.005(0.012)	1.099(0.132)	1.260(1.10)	1.432(0.08)	1.281(0.26)	1.478(0.98)
	Pr($H > 1$)	0.80	0.65	0.50	1.00	0.92	0.44
RW$_5$		1.005(0.011)	1.104(0.102)	1.156(0.99)	1.421(0.08)	1.282(0.22)	1.34(1.03)
	Pr($H > 1$)	0.84	0.78	0.56	1.00	0.92	0.60
RW$_6$		1.002(0.010)	1.090(0.089)	0.906(0.82)	1.432(0.12)	1.205(0.18)	1.026(0.97)
	Pr($H > 1$)	0.79	0.69	0.47	1.00	0.91	0.36
RW$_7$		1.002(0.011)	1.083(0.094)	1.029(0.89)	1.415(0.08)	01.256(0.18)	1.179(0.94)
	Pr($H > 1$)	0.81	0.70	0.58	1.00	0.93	0.48
RW$_8$		0.999(0.009)	1.090(0.057)	0.968(0.96)	1.795(0.11)	1.512(0.18)	1.262(0.91)
	Pr($H > 1$)	0.81	0.69	0.51	1.00	1.00	0.59
RW$_9$		1.001(0.011)	1.093(0.106)	0.993(0.77)	1.152(0.09)	1.035(0.22)	1.065(0.93)
	Pr($H > 1$)	0.79	0.70	0.52	1.00	0.79	0.36
RW$_{10}$		0.999(0.010)	0.975(0.127)	1.094(0.98)	1.150(0.19)	1.03(0.22)	1.064(0.93)
	Pr($H > 1$)	0.80	0.74	0.56	0.98	0.56	0.45
RW$_{11}$		0.999(0.010)	0.975(0.127)	1.093(0.98)	1.382(0.11)	1.16(0.21)	1.127(0.84)
	Pr($H > 1$)	0.84	0.86	0.58	1.00	0.89	0.40

[a]Pr($H > 1$) is the probability that the statistic is greater than one. RW$_1$ through RW$_{11}$ are generated using the same drift and the same conditional variance of the five exchange rates and the six interest rates. In parentheses is given the standard error for the statistic (computed over replications).

Table 4

DGM: TVC model; Theil U and ratio of MAD statistics; mean, standard errors, and $\Pr(H > 1)$.[a]

Variable		Theil U at steps			Ratio of MAD's at steps		
		1	13	52	1	13	52
TVC$_1$		0.981(0.005)	0.801(0.048)	0.731(0.140)	0.940(0.004)	0.773(0.045)	0.662(0.147)
	$\Pr(H > 1)$	0.04	0.05	0.00	0.04	0.04	0.02
TVC$_2$		0.980(0.005)	0.811(0.051)	0.756(0.132)	0.934(0.004)	0.778(0.049)	0.698(0.172)
	$\Pr(H > 1)$	0.06	0.04	0.01	0.04	0.05	0.01
TVC$_3$		0.977(0.004)	0.776(0.052)	0.705(0.135)	0.988(0.005)	0.792(0.051)	0.700(0.128)
	$\Pr(H > 1)$	0.06	0.05	0.02	0.04	0.06	0.02
TVC$_4$		0.980(0.005)	0.806(0.049)	0.688(0.123)	0.951(0.005)	0.777(0.046)	0.687(0.181)
	$\Pr(H > 1)$	0.05	0.05	0.01	0.02	0.03	0.03
TVC$_5$		0.979(0.004)	0.860(0.050)	0.700(0.130)	0.947(0.004)	0.742(0.050)	0.671(0.137)
	$\Pr(H > 1)$	0.06	0.04	0.02	0.06	0.04	0.03
TVC$_6$		0.989(0.007)	0.881(0.105)	0.788(0.263)	0.921(0.009)	0.788(0.046)	0.701(0.169)
	$\Pr(H > 1)$	0.07	0.06	0.06	0.03	0.04	0.01
TVC$_7$		0.984(0.007)	0.862(0.086)	0.712(0.298)	0.926(0.010)	0.764(0.041)	0.681(0.189)
	$\Pr(H > 1)$	0.05	0.06	0.06	0.05	0.03	0.02
TVC$_8$		0.988(0.006)	0.899(0.075)	0.799(0.267)	0.935(0.011)	0.782(0.043)	0.700(0.145)
	$\Pr(H > 1)$	0.06	0.06	0.03	0.04	0.04	0.01
TVC$_9$		0.986(0.006)	0.804(0.060)	0.722(0.300)	0.931(0.009)	0.757(0.049)	0.656(0.195)
	$\Pr(H > 1)$	0.04	0.04	0.00	0.07	0.04	0.01
TVC$_{10}$		0.990(0.007)	0.892(0.108)	0.737(0.266)	0.935(0.010)	0.791(0.054)	0.692(0.203)
	$\Pr(H > 1)$	0.08	0.04	0.06	0.06	0.03	0.02
TVC$_{11}$		0.987(0.006)	0.775(0.125)	0.651(0.224)	0.940(0.008)	0.708(0.047)	0.595(0.233)
	$\Pr(H > 1)$	0.04	0.05	0.04	0.05	0.02	0.00

[a]$\Pr(H > 1)$ is the probability that the statistic is greater than one. TVC$_1$ through TVC$_{11}$ are generated using the parameters displayed in the third column of table 2. In parentheses is given the standard error for the statistic (computed over replications).

Table 5

Theil U-statistics; 100 replications over 1986.1–1987.52 sample; 90% confidence interval; L = lower bound, U = upper bound.

Variable		Steps			Average at steps			Average over steps	Overall average
		1	13	52	1	13	52		
$/DM	L	0.982	0.677	0.486	0.966	0.772	0.667	0.618	0.816
	U	0.995	0.716	0.531	0.974	0.826	0.857	0.733	0.878
$/L	L	0.999	0.766	0.682				0.807	
	U	1.006	0.798	0.727				0.865	
$/SF	L	0.984	0.712	0.519				0.642	
	U	0.997	0.744	0.646				0.728	
$/FF	L	0.994	0.774	0.499				0.687	
	U	1.003	0.849	0.849				0.784	
YEN/$	L	0.997	0.940	0.758				0.701	
	U	0.998	0.958	0.817				0.813	
IR$_{WG}$	L	0.947	0.895	1.376				1.003	
	U	0.956	1.016	1.593				1.196	
IR$_{UK}$	L	0.907	0.691	0.514				0.646	
	U	0.974	0.702	0.581				0.738	
IR$_{SW}$	L	0.937	0.759	0.584				0.772	
	U	0.943	0.884	0.634				0.895	
IR$_{FR}$	L	0.980	0.660	0.502				0.670	
	U	0.986	0.705	0.583				0.764	
IR$_{US}$	L	1.002	0.962	1.125				1.120	
	U	1.011	0.996	1.246				1.196	
IR$_{JAP}$	L	0.948	0.746	0.281				0.411	
	U	0.957	0.785	0.377				0.538	

When the true DGM is a TVC model, both the Theil U and the ratio of MAD's have mean values below one and the value of one is in the upper 5% of the simulated distribution, except for the one-step-ahead for some interest rates. Note also the standard errors under this DGM are smaller than in the previous case and that, although the θ are misspecified, the estimated model still captures the nonlinear dependence present in the mean of the DGM.

To examine the robustness of the results to changes in the forecasting environment I conducted several experiments. First, I computed replications of the forecasting statistics using a suboptimal θ vector (listed in the last column of table 2) to forecast the eleven variables of the model in 100 replications of the sample 1986.1–1987.51. Replications are constructed using an algorithm which bootstraps the normalized recursive residuals of the TVC model. Table 5 presents a 90% confidence band for the ordered Theil U over replications and indicates that, even when the parameters of the prior are

arbitrarily chosen, the TVC model is superior to the RW model in forecast-
ing all the variables at all horizons (except for the Eurodollar rate). While the
performance of the TVC model in forecasting exchange rates at the one-step
horizon is unimpressive, at longer horizons the 90% confidence band for the
Theil U-statistic is still significant below one.[10]

To further examine the robustness of the results to various alterations of
the forecasting environment, first I sequentially reduced the estimation
period dropping one year of data at a time and keeping the end date fixed at
1985.52. In this way one can check the extent to which the availability of data
influences the forecasts of the TVC model. Second, I changed the forecasting
sample. Engel and Hamilton (1990) have recently demonstrated that ex-
change rates exhibit long swings. Therefore, the forecasting performance of
the model could be exaggerated if the forecasting sample coincides with one
of these swings. To check for this possibility, I estimated the model from
1979.1 up to five different dates (1982.52–1983.26–1983.52–1984.26–1984.52)
and then computed forecasting statistics for the following two years in each
of the five cases. Table A.1 in the appendix provides the Theil U-statistic and
the ratio of MAD's when the parameters of the TVC model are estimated
with one year of data (1985.1–1985.52) and forecasts are computed over the
1986.1–1987.51 period. Table A.2 reports the forecasting statistics for the
TVC model when the parameters are estimated over the 1979.1–1984.52
period and forecasts are computed over the 1985.1–1986.51 sample.[11] I find
that the availability of data influences the MSE of the TVC model at all
horizons and the performance of the model deteriorates when the forecasting
sample is changed. However, even in these cases, the MSE and MAD of the
TVC model are still smaller than those of the RW model for all the exchange
rates considered.

Finally, to assess how sensitive the results are to changes in the model
specification I conducted a few experiments by altering features of the prior.
There are at least three features of the specification which can, in principle,
account for the results obtained – a seasonal prior mean, the presence of
interest rates in the exchange rate equations, and the presence of nonlineari-
ties. It is therefore worthwhile to try to disentangle the contribution of each
of these features to the forecasting performance of the model. Table A.3
presents the forecasting statistics obtained when the prior means of the
exchange rates are assumed to be random walks (i.e., $\theta_2 = 0$ in all equations),
while table A.4 presents the forecasting statistics when interest rates are

---

[10]Since the results for Theil U and for the ratio of MAD's were qualitatively similar, I present
results for the Theil U only. On a VAX mainframe, the exercise required about three hours of
CPU time.

[11]I chose to present statistics for the 1985.1–1986.51 sample because the trend in all exchange
rates in terms of the dollar changes during the sample. If the model exploited a local trend to
forecast, it is bound to perform very poorly in this situation.

excluded from the exchange rate equations (i.e., $\theta_8 = 0$ in all equations). The tables indicate that, although both a seasonal prior mean and the presence of interest rates in the exchange rate equations improve the general forecasting performance of the model, they are not primarily responsible for the substantial superiority of the TVC model over the RW model in the long run. The 'good' long-run performance of the TVC model is therefore due the nonlinearities generated by time variation in the coefficients.

4.1. A comparison with the existing literature

It is not easy to compare the results of this paper with those presented in the existing literature for several reasons. First, the models recently employed by Diebold and Nason (1990), Meese and Rose (1989), and Mizrach (1990) are not nested in the framework of analysis of this paper. In their approach nonlinearities emerge in the functional specification of the DGM, i.e., their DGM is $y_t = f(x_t, \beta)$, where f is nonlinear, while here the DGM is conditionally linear and nonlinearities appear because of time variation in the β's. In addition, since they all estimate the coefficients of the model using a fixed window width, they do not allow for the type of time variation examined here. Second, in some cases the sampling frequency and the exchange rates used are not the same as ours. For example, Mizrach (1990) uses daily data from 1980 to 1985 on three EMS currencies (DM, FF, and Italian lira) in terms of the US dollar, while Engel and Hamilton (1990) use quarterly data for the 1973.1–1988.1 period for three currencies (DM, FF, L) in terms of the US dollar. Third, the samples used for estimation and forecasting are different from ours. This prevents, for example, a comparison with the work of Schinasi and Swamy (1989), who estimate a univariate frequentist version of the TVC model employed in this paper, using monthly data for the period 1973.1–1981.4, and forecast over the next 15 months. Fourth, some authors [e.g., Meese and Rose (1990) and Cheung and Pauly (1990)] examine only the in-sample fit of their model and provide no out-of-sample experiments.

The only meaningful comparison that can be made here is with the work of Diebold and Nason (1990). Diebold and Nason use weekly data for ten different currencies (Canadian dollar, FF, DM, Italian lira, YEN, SF, L, Belgium franc, Danish kroener, Dutch guilder) in terms of the US dollar for the period 1973.1–1987.38. Of this sample the first thirteen and a half years are used for estimation and about one year of data (1986.24–1987.38) is used to forecast from one to twelve weeks ahead. Table 6 presents the Theil U-statistic for the best nonparametric specification employed by Diebold and Nason and the Theil U of the TVC model when the parameters of the prior are chosen optimally over the period 1979.1–1986.23. For comparability, the Theil U-statistic for the TVC model is computed over the 1986.24–1987.38

Table 6

Comparison with Diebold and Nason (1990); estimation sample 1979.1–1986.38/forecasting sample 1986.39–1987.38; Theil U-statistics.[a]

Variable		Steps		
		1	4	12
$/DM	TVC	0.9826	0.9289	0.8593
	DN	0.9709	1.0128	1.0353
$/L	TVC	0.9885	0.9307	0.8514
	DN	0.9447	1.0610	0.9759
$/SF	TVC	0.9923	0.9511	0.8778
	DN	0.9939	1.0088	0.9853
$/FF	TVC	0.9712	0.9085	0.8357
	DN	1.0130	1.0473	1.0920
YEN/$	TVC	0.9895	0.9850	0.9797
	DN	0.9271	1.0457	1.0116

[a]DN indicates the values of the statistics computed with the best model found by Diebold and Nason. For $/DM, $/FF, and $/SF, the best model required a value of 10 for the smoothing parameter, for $/L a smoothing parameter value of 0.2, and for YEN/$ a smoothing parameter value of 0.7. For all five currencies the best model has lag length of three.

period. One can see that the performance of the two specifications are approximately equivalent at the one-step horizon, but that the model of this paper substantially outperforms their best specification at the twelve-step horizon.

5. Conclusions

This paper shows that a TVC model produces complex nonlinear dependencies in the moments of generated time series, replicates features of the conditional and the unconditional distributions of actual exchange rates, and improves upon the point forecasts of a RW model. It also demonstrates that the improvements obtain at various forecasting horizons and persist when parameters are chosen arbitrarily and when the forecasts are computed over repeated samples.

Two ingredients of the methodology are primarily responsible for the results: the use of time variation in the coefficients, which produces a specific form of nonlinearity and nonnormality in the estimated process, and the use of a Bayesian 'shrinkage' procedure, which exploits common features of several exchange markets. Neither of these features was exploited in previous work. Time variation in the coefficients was not, in general, allowed in the estimation procedure. Nor were interdependences across markets explicitly modelled and accounted for in the estimation. The paper shows that the

incorporation of these two forms of nonlinearities can appreciably improve the forecasting performance of a RW model.

The results obtained in this paper suggest two directions for future research. First, since the technique employed in the paper is complementary to that of Diebold and Nason, Meese and Rose, and Mizrach, it is of interest to examine whether the use of both forms of nonlinearities further improve our understanding of the DGM of exchange rate data. Also, since at shorter forecasting horizons nonlinearities of the type examined in this paper do not play a prominent role, further work is needed to assess the usefulness of other types of nonlinearities in forecasting at short horizons.

Second, since a TVC model is an appropriate statistical framework to model a situation where agents learn over time about the structure of the economy, the results suggest that work along the lines of Lewis (1989) or Kaminsky (1989) may be necessary to bring current theoretical models of exchange rate determination in line with the properties of the actual data.

Appendix

Table A.1

Estimation period 1985.1–1985.52/Forecasting sample 1986.1–1987.51.

Variable	Steps			Average at steps			Average over steps	Overall average
	1	13	52	1	13	52		
Theil U-statistics								
$/DM	1.031	1.088	1.576	1.021	1.075	1.189	1.252	1.382
$/L	1.008	1.077	2.174				1.492	
$/SF	1.028	1.093	1.849				1.351	
$/FF	1.034	1.238	1.782				1.421	
YEN/$	0.994	1.180	1.088				1.184	
IR_{WG}	1.017	1.137	1.447				1.193	
IR_{UK}	1.014	1.071	1.548				1.148	
IR_{SW}	0.9881	1.191	1.450				1.410	
IR_{FR}	1.005	1.148	1.985				1.419	
IR_{US}	1.005	1.074	2.288				1.480	
IR_{JAP}	1.157	1.441	1.599				1.848	
Ratio of MAD's								
$/DM	0.988	1.057	1.272	1.176	1.279	1.839	1.103	1.371
$/L	0.777	0.906	1.145				0.957	
$/SF	1.219	1.442	2.073				1.688	
$/FF	1.051	0.951	1.017				1.380	
YEN/$	1.034	0.993	1.085				1.051	
IR_{WG}	1.085	1.244	1.317				1.150	
IR_{UK}	1.210	2.104	3.352				2.873	
IR_{SW}	0.413	0.266	0.343				0.275	
IR_{FR}	1.105	1.518	1.900				1.787	
IR_{US}	1.016	1.087	1.372				1.412	
IR_{JAP}	1.175	1.456	1.440				1.413	

Table A.2

Estimation period 1979.1–1984.52/Forecasting period 1985.1–1986.52.

Variable	Steps			Average at steps			Average over steps	Overall average
	1	13	52	1	13	52		
Theil U-statistics								
$/DM	0.988	0.884	1.140	1.021	1.111	1.121	0.981	1.019
$/L	0.990	1.058	1.402				1.206	
$/SF	0.999	1.036	1.383				1.152	
$/FF	0.988	0.999	1.315				1.122	
YEN/$	0.992	0.968	1.145				1.023	
IR_{WG}	1.044	1.164	1.400				1.179	
IR_{UK}	1.034	1.005	1.089				0.968	
IR_{SW}	0.996	0.966	0.935				0.897	
IR_{FR}	1.016	0.886	1.105				0.934	
IR_{US}	0.991	0.929	0.668				0.780	
IR_{JAP}	1.019	1.092	0.726				0.988	
Ratio of MAD's								
$/DM	0.987	0.833	1.018	1.051	1.109	1.039	0.867	1.012
$/L	0.891	0.700	0.595				0.677	
$/SF	0.973	0.861	0.887				0.920	
$/FF	0.950	1.003	1.152				1.048	
YEN/$	1.001	1.025	0.952				0.990	
IR_{WG}	1.091	1.346	1.522				1.361	
IR_{UK}	1.060	1.033	0.947				1.016	
IR_{SW}	1.001	1.021	0.865				0.972	
IR_{FR}	1.103	1.171	1.108				1.145	
IR_{US}	1.003	1.020	0.990				1.039	
IR_{JAP}	1.089	1.134	1.062				1.098	

Table A.3

Estimation period 1979.1–1985.52/Forecasting sample 1986.1–1987.51; no seasonal prior mean.

Variable	Steps			Average at steps			Average over steps	Overall average
	1	13	52	1	13	52		
Theil U-statistics								
$/DM	0.9643	0.7657	0.4778	0.9825	0.8757	0.8214	0.7105	0.8556
$/L	0.9762	0.8061	0.7395				0.8704	
$/SF	0.9711	0.8006	0.6177				0.7058	
$/FF	0.9753	0.8347	0.4982				0.7527	
YEN/$	0.9758	0.8404	0.6890				0.7940	
IR_{WG}	0.9984	1.1191	1.9431				1.3437	
IR_{UK}	0.9860	0.7927	0.5741				0.7269	
IR_{SW}	0.9854	1.0340	0.6223				0.8646	
IR_{FR}	0.9890	0.8522	0.5742				0.7446	
IR_{US}	1.0139	1.1809	1.9679				1.4071	
IR_{JAP}	0.9471	0.5640	0.3309				0.4918	

Table A.3 (continued)

Variable	Steps			Average at steps			Average over steps	Overall average
	1	13	52	1	13	52		
Ratio of MAD's								
$/DM	0.9382	0.7698	0.3908	0.9895	0.9206	0.8165	0.6496	0.8851
$/L	0.7570	0.6853	0.3380				0.5694	
$/SF	1.1683	1.1387	0.7256				0.9274	
$/FF	1.0000	0.6837	0.5939				0.7618	
YEN/$	0.9885	0.8009	0.5889				0.7547	
IR$_{WG}$	1.0521	1.1971	2.0437				1.4460	
IR$_{UK}$	1.1436	1.4763	1.3881				1.2345	
IR$_{SW}$	0.4091	0.2572	0.1320				0.2165	
IR$_{FR}$	1.5440	1.5018	1.1871				1.3446	
IR$_{US}$	1.0224	1.1832	1.9335				1.3574	
IR$_{JAP}$	0.9605	0.5918	0.3209				0.4744	

Table A.4

Estimation sample 1979.1–1985.52/Forecasting sample 1986.1–1987.51; system with exchange rates only.

Variable	Steps			Average at steps			Average over steps	Overall average
	1	13	52	1	13	52		
Theil U-statistics								
$/DM	0.9866	0.8063	0.4854	0.9906	0.8832	0.6375	0.7095	0.8033
$/L	0.9984	1.0000	0.8984				0.9483	
$/SF	0.9912	0.8397	0.6223				0.7951	
$/FF	0.9878	0.8857	0.4928				0.7697	
YEN/$	0.9794	0.8840	0.6886				0.7939	
Ratio of MAD's								
$/DM	0.9652	0.8346	0.4116	0.9910	0.8944	0.5537	0.6593	0.7599
$/L	0.7684	0.8769	0.4145				0.6286	
$/SF	1.2000	1.2110	0.7297				0.9861	
$/FF	1.0361	0.7486	0.6243				0.7716	
YEN/$	0.9885	0.8010	0.5885				0.7543	

References

Blattberg, R. and T. Sargent, 1971, Regression with nonstable Gaussian disturbances: Some sampling results, Econometrica 39, 501–510.

Boothe, P. and D. Glassman, 1987, The statistical distribution of exchange rates, Journal of International Economics 22, 297–319.

Canova, F., 1988, Seasonalities in foreign exchange markets, Working paper no. 8901 (Brown University, Providence, RI).

Canova, F., 1992, An alternative approach to modelling and forecasting seasonal time series, Journal of Business and Economic Statistics 10, 97–108.

Cheung, Y. and P. Pauly, 1990, Random coefficients autoregressive modelling of conditionally heteroskedastic processes: Short run exchange rate dynamics, Manuscript (University of Toronto, Toronto).

Clark, P., 1973, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41, 136–156.

Diebold, F., 1988, Empirical modelling of exchange rate dynamics (Springer Verlag, New York, NY).

Diebold, F. and J. Nason, 1990, Nonparametric exchange rate predictions?, Journal of International Economics 28, 315–328.

Diebold, F. and M. Nerlove, 1989, The dynamic exchange rate volatility: A multivariate latent-factor ARCH model, Journal of Applied Econometrics 4, 1–22.

Doan, T., R. Litterman, and C. Sims, 1984, Forecasting and conditional projections using a realistic prior distribution, Econometric Reviews 3, 1–100.

Domowitz, I. and C. Hakkio, 1985, Conditional variance and the risk premium in foreign exchange markets, Journal of International Economics 19, 47–66.

Engel, C. and J. Hamilton, 1990, Long swings in the exchange rate: Are they in the data and do markets know it?, American Economic Review 80, 689–713.

Engle, R., D. Lilien, and R. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, Econometrica 55, 391–408.

Engle, R. and M. Watson, 1985, The Kalman filter: Application to forecasting and rational expectations models, in: T. Bewley, ed., Advances in econometrics, 5th world congress, Vol. 1 (Cambridge University Press, Cambridge, MA) 245–283.

Fama, E., 1976, Foundations of finance (Basic Books, New York, NY).

Fama, E. and R. Roll, 1968, Some properties of symmetric stable distributions, Journal of the American Statistical Association 63, 817–836.

Friedman, D. and S. Vandersteel, 1982, Short run fluctuations in foreign exchange rates: Evidence from the data 1973–1979, Journal of International Economics 13, 171–186.

Gallant, R., D. Hsieh, and G. Tauchen, 1990, On fitting a recalcitrant series: The pound/dollar exchange rate, 1974–1983, in: W. Barnett, J. Powell, and G. Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics, Proceedings of the 5th international symposium in economic theory and econometrics (Cambridge University Press, Cambridge, MA).

Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1317–1340.

Granger, C. and A. Anderson, 1978, An introduction to bilinear time series models (Vandenhoeck and Ruprecht, Göttingen).

Hamilton, J., 1989, A new approach to the economic analysis of nonstationary time series and the business cycles, Econometrica 57, 357–384.

Hsieh, D., 1988, The statistical properties of daily foreign exchange rates: 1974–1983, Journal of International Economics 24, 129–145.

Hsieh, D., 1989, Modelling heteroskedasticity in daily exchange rates, Journal of Business and Economic Statistics 7, 307–318.

Kaminsky, G., 1989, The peso problem and the behavior of the exchange rate: The dollar–pound exchange rate, 1976–1987, Manuscript (University of San Diego, San Diego, CA).

Koenker, R.W. and G.W. Basset, 1982, Tests of linear hypothesis and $L-1$ estimation, Econometrica 50, 1577–1583.

Lewis, K., 1989, Can learning affect exchange rate behavior? The case of the dollar in the early 1980's, Journal of Monetary Economics 23, 79–100.

Lipster, R. and A. Shiryayev, 1978, Statistics of random processes (Springer Verlag, New York, NY).

Litterman, R., 1982, Specifying VAR's for macroeconomic forecasting, Staff report no. 92 (Federal Reserve of Minneapolis, Minneapolis, MN).

Ljung, L. and T. Soderstrom, 1987, Theory and practice of recursive identification (MIT Press, Cambridge, MA).

Meese, R. and K. Rogoff, 1983a, Empirical exchange rate models of the seventies: Do they fit out of sample, Journal of International Economics 14, 3–24.

Meese, R. and K. Rogoff, 1983b, The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification, in: J. Frankel, ed., Exchange rate and international economics (University of Chicago Press, Chicago IL).

Meese, R. and A. Rose, 1989, An empirical assessment of non-linearities in model of exchange rate determination, Discussion paper no. 367 (Board of Governors of the Federal Reserve System, International Finance Division, Washington, DC).

Mizrach, B., 1990, Multivariate nearest neighbor forecasts of EMS exchange rates, Journal of Applied Econometrics, forthcoming.

Pagan, A. and A. Ullah, 1988, The econometric analysis of models with risk terms, Journal of Applied Econometrics 3, 87–105.

Pagan, A. and K. Yong, 1990, Nonparametric estimation and the risk premium, in: W. Barnett, J. Powell, and G. Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics, Proceedings of the 5th international symposium in economic theory and econometrics (Cambridge University Press, Cambridge).

Rao, P., 1984, Nonparametric functional estimation (Academic Press, New York, NY).

Rosenberg, B., 1973, Random coefficients models: The analysis of a cross section of time series by stochastically convergent parameter regression, Annals of Economic and Social Measurement 2, 399–428.

Sarris, A., 1973, A Bayesian approach to estimation of time varying regression coefficients, Annals of Economic and Social Measurement 2, 501–523.

Schinasi, G. and P.A. Swamy, 1989, The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change, Journal of International Money and Finance 8, 375–390.

Schwarz, G., 1978, Estimating the dimension of a model, American Statistics 6, 461–464.

Shafear, S. and al., 1975, Alternative models of systematic risk, in: E. Elton and M. Gruber, eds., International capital markets: An inter and intra country analysis (North-Holland, Amsterdam) 150–161.

Sims, C., 1984, On the martingale behavior of asset prices, Working paper no. 205 (University of Minnesota, Minneapolis, MN).

Sims, C., 1986, Bayesmth: A program for multivariate Bayesian interpolation, Working paper no. 234 (University of Minnesota, Minneapolis, MN).

Sims, C., 1989, A nine variable probabilistic macroeconomic forecasting model, Discussion paper no. 14 (Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, Minneapolis, MN).

Thisted, R. and W. Wecker, 1981, Predicting a multitude of time series, Journal of the American Statistical Association 76, 516–523.

Tsay, R., 1987, ARCH and time varying parameter model, Journal of the American Statistical Association 82, 590–604.

Weiss, A., 1984, ARMA models with ARCH errors, Journal of Time Series Analysis 5, 129–143.

Wolff, C., 1987, Time varying parameters and the out-of-sample forecasting performance of structural exchange rate models, Journal of Business and Economic Statistics 5, 87–97.

Zellner, A. and H.S. Hong, 1989, Forecasting international growth rates using Bayesian shrinkage and other procedures, Journal of Econometrics 40, 183–202.