THE KOREAN ECONOMIC REVIEW
Volume 8, Number 1, Summer, 1992.

ON THE PREDICTIVE POWER OF THE SPREAD BETWEEN SPOT AND FORWARD EXCHANGE RATES FOR VOLATILITY

TAE HWY LEE*

This paper is concerned with modeling the conditional heteroskedasticity of the prediction error of foreign exchange rates with a function of the spread between spot and forward rates. As spot and forward rates are cointegrated we use a system of error correction models (ECM) for mean prediction. To predict variance we use an extended bivariate generalized autoregressive conditional heteroskedasticity (GARCH) as a function of the spread. The model will be referred to as GARCH-X in ECM. Using daily series for seven currencies, we find that unmodeled conditional heteroskedasticity by GARCH can generally be explained by the squared spread. This indicates that as the spread is bigger the exchange rates are more volatile.

I. INTRODUCTION

The characterizations of exchange rate movements have important implications for many issues in international finance and macroeconomics. It is therefore important to carefully model any temporal variations in the volatility process as well as in the process of change in the exchange rate series. A standard way to specify volatility is autoregressive conditional heteroskedasticity (ARCH) due to Engle (1982) and generalized ARCH (GARCH) of Bollersiev (1986). For the last decade various modified and extended parameterizations of the ARCH model have been suggested in the literature and many have proven effective to summarize observed time series dependence in data with emphasizing particular recognized empirical features and statistical merits. See Bollerslev, Chou and Kroner (1992) for a survey.

In this paper, we consider a new interesting extension of the GARCH model for the error correction models (ECM) of cointegarted series. The extension is multivariate by nature. We use a system of ECMs for the conditional mean, and the extended bivariate GARCH model for the conditional variance. The main virtue of the model lies in its capability of pointing to a paticular feature of

---

*Assistant Professor, Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA. I would like to thank Richard Baillie for providing the data used in this paper, Ji-Chai Lin, Tung Liu, Doug McMillin and Hector Zapata for useful conversations, and College of Business Administration at LSU for financial support.

cointegrated series, which is the potential relationship between disequilibrium and uncertainty.

If the short-run deviation from the long-run cointegrated relationship is an important variable in the conditional mean, so may it be in the conditional variance. This may imply that if the series deviate further from each other it is harder to predict. Examining the behavior of the variances over time as a function of disequilibrium is reasonable when one expects increased volatility due to shocks to the system to propagate on both of the first and the second moments. If disequilibrium is responsible for the conditional heteroskedasticity, it could possibly be modeled with a function of several lagged short-run errors. The model thus seems appropriate for testing for causality in variance through the error correction term.

Using this model we investigate the predictive power of the spread between daily spot and forward exchange rates for seven currencies in predicting volatility of the exchange rate changes. We find that unmodeled conditional heteroskedasticity by the standard GARCH model can be explained by a function of the spread (computed from the residuals of a cointegrating regression between the spot and forward rates). The conditional variance of the prediction error is positively related with the squared spread. This indicates that as the spread is bigger the exchange rates are more volatile and harder to predict. The short-run error from the cointegrating long-run relationship is therefore a useful variable in modeling conditional variance as well as conditional mean.

If the squared short-run error has predictive power for the time varying conditional variances of the spot and forward exchange rate changes, this may be exploited to obtain more precise time varying confidence intervals for point forecasts of exchange rate changes, and hence to get better estimates of time varying risk premia. In other words, the spread between the spot and forward rates has explanatory power for the time varying risk premia. This interpretation is not standard as most studies in literature have examined the opposite causal direction from the conditional variance (or conditional standard deviation) to the conditional mean. While the spread has been considered to reflect the premia for risk or uncertainty so that time varying conditional variances Granger-cause the spread, our study exhibits the presence of the converse causal relationship from the spread to conditional variances as well.

Section II discusses econometric methodology, Section III conducts empirical studies. Section IV concludes.

I. THE GARCH-X IN ERROR CORRECTION MODELS

It is widely believed that many economic time series are integrated. An integrated series contains a unit root and has a long memory. If there are several integrated variables whose linear combination can be stationary, they are called cointegarted.

They move together in the long run, although they may deviate from each other in the short run due to disturbances. The dynamic models for the short run behavior of cointegrated series are called error correction models (ECM).

An ECM may be considered as a prediction equation of the cointegrated series as it is a conditional expectation of the first difference of the series given an information set. A conditional expectation is a function of the variables in the information set. Hence conditional expectations for cointegrated system are functions of the error correction term, which should help predict the conditioned variables.

Consider a 2×1 vector of cointegrated series $X_t = (s_t \ f_t)'$. They may be considered to be generated by the following ECM

$$(1) \quad \Delta X_t = \mu + \gamma z_{t-1} + \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_k \Delta X_{t-k} + \varepsilon_t$$

where $\Delta = 1\text{-}B$, B is the backshift operator, $z_t \equiv \alpha' X_t$ is the magnitude of the short run deviation at time t, and μ, γ, α are all 2×1 vectors of parameters and all Γs are 2×2 matrices of parameters. α is the cointegrating vector which determines the long run equilibrium relationship and γ is the parameter measuring the speed of error corrections. The error vector $\varepsilon_t = (e_{it} \ e_{2t})'$ is assumed to follow a bivariate conditional normal distribution with mean zero and conditional variance-covariance matrix $H_t \equiv E(\varepsilon_t \ \varepsilon_t' | \mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is the σ-field generated by all information available at time t-1.

If the ECM (1) is correctly specified for the conditional mean $E(\Delta X_t | \mathcal{F}_{t-1})$, it should be that $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ with probability one by construction. Then the one step ahead prediction error ε_t must not be predicted by any means using the same information set. Thus the lag length k should be chosen so that ε_t exhibits the martingale difference property. When dealing with empirical examples we shall choose k using the Schwarz information criterion (SIC). The SIC ensures that correct k can be chosen with probability approaching one (Sawa (1978)), if the model is linear in conditional mean. If there is neglected nonlinearity in conditional mean it may lead to dynamic misspecification with incorrect choice of k, which may then invalidate inferences on higher order conditional moments.

The error ε_t is a martingale difference process but is not necessarily independent. The squared error may be predictable using the information set (Granger (1983)), and thus the conditional variance may be a function of the variables in the information set. In the absence of any structural model for the conditional variance the GARCH model became a standard way to model this dependency.

The generalization of the univariate GARCH model to a bivariate GARCH model requires allowing the covariance to change with time. We use Bollerslev's (1990) model and assume the conditional correlation to be constant so that all the variations over time in conditional covariance are due to changes in two conditional variances. Let $h_{ijt} \equiv E(e_{it} \ e_{jt} \ | \mathcal{F}_{t-1})$. Then the standard GARCH(p, q) specification with the assumption of constant conditional correlation can be written as

$$(2) \quad h_{iit} = w_i + \sum_{s=1}^{q} \alpha_{is} e^2_{i,\, t-s} + \sum_{s=1}^{p} \beta_{is} h_{ii,\, t-s}, \quad i=1, 2$$
$$p = h_{12t} (h_{11t} h_{22t})^{-1/2}$$

The ECM specification in <1> is based on the assumption that the regression function for the conditional mean is linear and measurable with respect to the σ-field generated by the variables in the right hand side of the equation <1>. By the same reason, the conditional variance may be a function of those variables. Each element of H_t can be written as a measurable function of the variables in $\mathcal{F}_{t-1}$. See White (1984, p.52). We may consider a function of ΔX_{t-s} such as selective elements of $\Delta X_{t-s} \Delta X'_{t-s}$ for some $s \geqslant 1$. Empirical evidence for the significance of some economic variables in the conditional variance equation has been obtained by Giovannini and Jorion (1989) and Attanasio (1991), among others.

However, the short run disequilibrium error z_{t-1} may seem an especially interesting variable. We consider the squared z_{t-1} so that the above specification <2> can be extended as follows with possibly longer lags of z^2_{t-s}, $s=1, \ldots, r$.

$$(3) \quad h_{iit} = \omega_1 + \sum_{s=1}^{q} \alpha_{is} e^2_{i,\, t-s} + \sum_{s=1}^{p} \beta_{is} h_{ii,\, t-s} + \sum_{s=1}^{r} \delta_{is} z^2_{t-s}, \quad i=1, 2$$
$$\rho = h_{12t} (h_{11t} h_{22t})^{-1/2}.$$

This may be referred to as 'GARCH (p, q)-X(r) in ECM'.

To emphasize the relationship between h_{iit} and z^2_{t-1} we may use GARCH(0, 0)-X(1). Also GARCH(1, 0)-X(1) or GARCH(0, 1)-X(1) may be considered. It seems, however, that GARCH(1, 1)-X(1) may be useful in practice, especially for estimation. Previous studies such as Hsieh (1989, p.317) and Baillie and Bollerslev (1989b, p.303) use $p=q=1$ to model GARCH for daily foreign exchange rate changes.

If the constant conditional correlation assumption seems too restrictive, other types of bivariate ARCH models may also be used. The alternatives are the model used in Bollerslev, Engle and Wooldridge (1988); the factor ARCH specification in Engle, Ng and Rothschild (1990); and the multivariate exponential GARCH model of Nelson (1991), etc. These extensions are left for future work.

III. SPREAD AND VOLATILITY IN SPOT AND FORWARD EXCHANGE RATES

III.A. Data

In the following analysis we use daily spot and thirty-day forward exchange rate data from the New York Foreign Exchange Market. The data begin on March

1, 1980 and end on January 28, 1985, which constitute a total of 1245 observations. The data are those used in Baillie and Bollerslev (1989a, 1989b), which were provided by Data Resources Incorporated (DRI). They are opening bid prices for British Pound (BP), German Deutchmark (DM), Japanese Yen (JY), Canadian Dollar (CD), French Franc (FF), Italian Lira (IL) and the Swiss Franc (SF). Apart from the BP all the series are in terms of the number of US Dollars for one unit of foreign currency. The notation s_t and f_t is used for the spot and forward series, respectively, so that $x_t = (s_t\ f_t)'$.

III.B. Unit Root, Cointegration, and ECM

Baillie and Bollerslev (1989a, p.171) present the results that the unit root hypothesis cannot be rejected for all the series . They also report the results of the cointegration tests of Phillips (1987) between s_{t+22} and f_t, which are found strongly cointegrated. However, if s_{t+22} and f_t are used in the cointegrating regression, the ECM <1> includes variables not available at time *t*. As we are interested in prediction using the conditional expectations given information at time *t*, we instead use s_t and f_t in the cointegrating regression so that the right-hand side of equation <1> should not include the variables unknown at time *t*. Using the tests of Phillips (1987) and Johansen (1988), s_t and f_t are also found strongly cointegrated in all currencies at less than 1% level. To save space the estimated statistics are not reported.

The lag lengths (k) in ECM (1) are chosen by the use of the SIC, which are 2 (BP), 3 (DM), 3 (JY), 5 (CD), 0 (FF), 1 (IL), and 3 (SF).

III.C. Initial Diagnostics

In Table 1, the asymptotic *p*-values for various specification tests for the ECM <1> estimated by least squares assuming conditional homoskedasticity are presented. These tests are: (a) the Ljung-Box portmanteau test for up to the 20th order serial correlation in the residuals (Q(20)), and in the squared residuals ($Q^2(20)$), (b) Wooldridge's (1990) robust regression based LM test for autocorrelations (AR), (C) the LM test for ARCH of orders of 1, 2, 5, 10, (d) White's (1989) neural network test for neglected nonlinearity in conditional mean, and (e) the LM test for GARCH-X. The coefficients of unconditional skewness (m_3) and excess kurtosis (m_4) of the residuals are also reported. The LM tests for serial correlation, ARCH, and functional form are denoted by AR, ARCH, and Neural, respectively.

As noted by Domowitz and Hakkio (1986), Diebold (1987), Cumby and Huizinga (1992), and Wooldridge (1990), among others, the presence of time varying higher moments generally induces an incorrect (usually too large) size of specification tests. Thus we use the robust LM test for autocorrelations. Using the lag lengths (*k*) in ECM <1> chosen by the SIC, the residuals are not serially correlated.

The Ljung-Box statistics for the squared residuals and the LM test statistics

[**Table 1**] Initial Diagnostics for ECM (1) Estimated by OLS

	BP		DM		JY		CD	
	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t
Q(20)	0.329	0.312	0.636	0.820	0.230	0.158	0.105	0.105
$Q^2(20)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Neural(3)	0.989	0.680	0.271	0.381	0.000	0.000	0.168	0.687
AR(1)	0.130	0.584	0.132	0.521	0.774	0.939	0.463	0.275
AR(2)	0.948	0.972	0.970	0.930	0.612	0.812	0.508	0.580
AR(5)	0.998	0.988	0.324	0.380	0.500	0.802	1.000	0.999
AR(10)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ARCH(1)	0.016	0.007	0.002	0.003	0.000	0.000	0.000	0.000
ARCH(2)	0.015	0.006	0.000	0.000	0.000	0.000	0.000	0.000
ARCH(5)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ARCH(10)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ARCH[0]-X(1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ARCH[1]-X(1)	0.795	0.609	0.020	0.027	0.000	0.000	0.007	0.019
ARCH[2]-X(1)	0.804	0.595	0.022	0.027	0.000	0.000	0.015	0.034
ARCH[5]-X(1)	0.769	0.532	0.058	0.052	0.000	0.000	0.024	0.052
ARCH[10]-X(1)	0.912	0.572	0.113	0.091	0.000	0.000	0.043	0.077
GARCH[1, 1]-X(1)	0.699	0.527	0.007	0.001	0.000	0.000	0.855	0.911
GARCH[1, 1]-X(2)	0.796	0.522	0.011	0.000	0.000	0.000	0.702	0.704
m_3	−0.248	−0.292	0.381	0.365	0.359	0.664	0.279	0.279
m_4	1.800	1.976	1.313	1.284	2.593	3.750	3.304	3.430

Note: The number in [ ] is the degree of freedom. The number in [ ] is the order of GARCH fitted under the null hypothesis to test for GARCH-X. All the values are asymptotic *p*-values except for the coefficients of skewness and excess kurtosis. For the neural network test we use 10 phantom hidden units, 3 principal components of them, and 5 draws of the test in computing the Hochberg Bonferroni bound (see Lee, White, and Granger (1992)).

for ARCH are very significant. McLeod and Li (1983) show that Q^2 (20) is asymptotically distributed as χ^2 (20) under the null hypothesis that e_{it} is independent. While the null hypothesis of no ARCH can not be decisively rejected for the spot series of FF, it may be noted that the extreme degree of kurtosis or time varying conditional kurtosis may possibly have affected the power of the tests.

It is possible that the significant ARCH effects are due to neglected nonlinearity in the conditional mean resulting in dependence in higher order conditonal moments. Thus we test linearity in the ECM <1>. The neural network test for linearity in conditional mean is not significant for all currencies but JY and FF. As the conditional mean may be misspecified for JY and FF, neglected nonlinear dependence in conditional mean may be exploitable for better prediction. The presence of ARCH may also affect the power of the neural network test for conditional mean. Thus we also compute the Wooldridge's robust version of the neural

[**Table 1**] Continued

	FF		IL		SF	
	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t
Q(20)	0.238	0.109	0.231	0.431	0.376	0.361
$Q^2(20)$	0.083	0.000	0.000	0.000	0.000	0.000
Neural(3)	0.083	0.005	0.999	0.829	0.427	0.249
AR(1)	0.038	0.040	0.929	0.384	0.781	0.722
AR(2)	0.999	0.987	0.998	0.885	0.932	0.938
AR(5)	0.999	1.000	1.000	1.000	0.152	0.231
AR(10)	0.921	0.955	1.000	1.000	1.000	1.000
ARCH(1)	0.159	0.026	0.000	0.001	0.080	0.145
ARCH(2)	0.228	0.008	0.000	0.000	0.000	0.000
ARCH(5)	0.001	0.000	0.000	0.000	0.000	0.000
ARCH(10)	0.013	0.000	0.000	0.000	0.000	0.000
ARCH[0]-X(1)	0.000	0.000	0.000	0.000	0.000	0.000
ARCH[1]-X(1)	0.000	0.000	0.127	0.017	0.040	0.038
ARCH[2]-X(1)	0.000	0.001	0.144	0.024	0.052	0.048
ARCH[5]-X(1)	0.000	0.007	0.228	0.062	0.095	0.072
ARCH[10]-X(1)	0.000	0.008	0.230	0.062	0.121	0.098
GARCH[1, 1]-X(1)	0.776	0.000	0.062	0.172	0.000	0.000
GARCH[1, 1]-X(2)	0.000	0.000	0.118	0.129	0.000	0.000
m_3	−0.191	−0.148	0.291	0.181	0.372	0.360
m_4	7.926	3.746	1.926	2.083	1.232	1.221

network test, which is not significant for JY and FF.

III.D. Testing for GARCH-X

We test the GARCH-X using the Lagrange multiplier (LM) test, the Wald test, and thd likelihood ratio (LR) test. Our LM test is based on a consequence of the null hypothesis. When h_{iit} is obtained under the null hypothesis, it should be that $E(e_{it}^2 - h_{iit} | \mathcal{F}_{t-1}) = 0$, i = 1,2 with probability one if the null hypothesis is true. Consequently, $(e_{it}^2 - h_{iit})$ is uncorrelated with any measurable function of the variables in $\mathcal{F}_{t-1}$. We construct an LM test statistic for the null hypothesis that $(e_{it}^2 - h_{iit})$ is uncorrelated with z_{t-1}^2.

In computing the test statistic, we replace e_{it} with estimated residuals. From the standard asymptotic arguments the TR^2 statistic has the $\chi^2(1)$ distribution asymptotically when the null hypothesis is true, where T is the number of observations and R^2 is the uncentered squared multiple correlation from a standard linear regression of $(e_{it}^2 - h_{iit})$ on z_{t-1}^2.

We consider the null hypotheses that h_{iit} is ARCH(q), $q = 0, 1, 2, 5, 10$. In computing the test statistics we estimate h_{iit} by OLS using the least squares residual e_{it} from the equation <1>. The test with $q=0$ provides information on the correlation structure between e_{it}^2 and z_{t-1}^2.

We also consider the null hypothesis that h_{iit} is GARCH(1, 1). We estimate h_{iit} of the null hypothesis by maximum likelihood estimation. If a higher order GARCH-X(r) is to be tested, the LM test statistic will have the $\chi^2(r)$ distribution asymptotically under the null hypothesis. Table 1 reports the statistics for $r=1$, 2. For the spot series of FF the test is significant for $r=2$ while it is not for $r=1$. The LM test statistics for GARCH-X are significant for all currencies but BP and CD.

We also use $|z_{t-1}|$ instead of z_{t-1}^2 as a 'test function'. For our data the power of the test with the test function $|z_{t-1}|$ is generally weaker (except for FF and IL) than with z_{t-1}^2, and the results using $|z_{t-1}|$ will not be reported. Interestingly, however, we find that when $p=q=0$ under the null hypothesis the power of the test with the test function $|z_{t-1}|$ is generally stronger than with the test function z_{t-1}^2. This shows that e_{it}^2 is generally more correlated with $|z_{t-1}|$ than with z_{t-1}^2. This may be similar to the previous evidences in stock returns. Taylor (1988), Kariya, Tsukuda and Maru (1990), and Granger (1991) find that while the autocorrelations of stock returns are very small, the autocorrelations for the squared returns are higher and those of absolute value of returns are even more so. The choice of a test fuction and the GARCH-X specification <3> for an empirical study may be based on the power property of the test statistic.

The test may have power in some other directions and other misspecification can possibly be detected. If the results indicate the presence of the GARCH-X effect they might not provide definitive evidence of neglected GARCH-X effect. The possible presence of higher order GARCH effects or the potential effects of other neglected variables in the information set may lead to the observed results. Much of the model selection literature in which model choice is based on the SIC or the Akaike information criterion (AIC), etc. is concerned with selecting parsimoniously undominated models usually in terms of likelihood and subject to a restriction of correct specification of a model in that the residual processes from the (first and higher) conditional moments be martingale difference sequences. If a model is correctly specified in conditional mean and conditional variance, then we may use the GARCH-X specification as a useful description of the salient aspects of the chosen phenomena.

The GARCH-X effect is also tested by the Wald tests (Table 4) and by the LR tests (Table 6) for five currencies. The Wald tests suggest the presence of the GARCH-X effect for all currencies while the LR tests show no GARCH-X effect in CD. The model selection criteria (Table 6) also indicate that the model with GARCH(1, 1)-X has a better fit than the model with GARCH(1, 1), except for CD.

III.E. Estimation

We estimate the model with formulations of the conditional mean $E(\Delta X_t \mid \mathcal{F}_{t-1}$ as in <1> and the conditional variance $H_t \equiv E(\varepsilon_t\, \varepsilon_t' \mid \mathcal{F}_{t-1})$ as in <2> or <3>. Let θ be the vector of all the parameters in the conditional mean and conditional variance. In order to implement the maximum likelihood estimation, we assume that the error vecter ε_t is distributed as conditionally normal. The parameter estimates are obtained by maximizing the log-likelihood function over θ using the Berndt, Hall, and Hausman (BHHH, 1974) algorithm with numerical derivatives being used. The parameter ω_1 is restricted to be positive. Each of α_1 and β_1 is constrained on a unit interval between zero and unity, but no restriction is imposed on δ_1. See Harvey (1990, p.129) for constrained optimization. At the maximizing θ, an estimate of the asymptotic covariance matrix of the estimated parameters is obtained from the inverse of the outer product of the score vectors.

Table 2 reports the results of the estimated conditional variances in <2>. To save space the results for the codnditional mean in <1> are not reported. The asymptotic *t*-values are in parentheses. For all seven currencies, $\alpha_1 + \beta_1$, $i=1$, 2 is very close to unity in the GARCH(1, 1) model.

In Table 4 the estimated parameters of the conditional variances in <3> are reported for five currencies. BP and JY are not included as we can not find a step of increasing likelihood from many different initial values of θ including those obtained in Table 2. The estimates of β become substantially lower than those in Table 2. For all five currencies the estimates of δ are very significant.

Since e^2_{it} is highly correlated with z^2_{t-1} as shown from the LM test for ARCH(0)-

[**Table 2**] Estimates of Bivariate GARCH(1, 1) parameters

	BP	DM	JY	CD	FF	IT	SW
ω_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.193)	(9.209)	(3.029)	(0.213)	(4.881)	(8.772)	(3.129)
α_1	0.025	0.168	0.053	0.056	0.377	0.153	0.102
	(9.479)	(12.453)	(7.304)	(11.120)	(18.274)	(11.256)	(13.764)
β_1	0.976	0.722	0.942	0.946	0.682	0.739	0.897
	(320.066)	(38.229)	(137.189)	(227.887)	(57.768)	(39.350)	(140.033)
ω_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(10.123)	(3.260)	(0.781)	(4.403)	(11.471)	(3.230)
α_1	0.027	0.169	0.054	0.054	0.378	0.162	0.097
	(10.881)	(12.778)	(8.946)	(11.110)	(17.436)	(13.725)	(14.203)
β_2	0.975	0.713	0.941	0.948	0.684	0.715	0.902
	(414.219)	(38.633)	(143.670)	(237.553)	(65.857)	(45.332)	(147.061)
ρ	0.981	0.996	0.947	0.993	0.983	0.977	0.996
	(1467.690)	(5635.899)	(411.990)	(2621.888)	(1599.355)	(808.852)	(5299.937)

Note: Asymptotic *t*-values are in (   ).

X in Table 1, so is e^2_{it-1} with z^2_{t-1}. Thus in order for the GARCH(1, 1)-X model to be useful in practice, not only should δ be significant (as shown by the Wald tests in Table 4) but the squared spread z^2_{t-1} should also have *additional* predictive power for the variances of the exchange rate changes (as shown by the LM tests in Table 1 and the LR tests in Table 6). Although the estimates of δ are significant in CD, there seems no significant additional contribution of the squared spread in predicting the second moments of the exchange rate changes.

III.F. Specification Tests

In order to test the validity of the GARCH(1, 1) model reported in Table 2 a series of specification tests for the model standardized by $h^{1/2}_{iit}$ are presented in Table 3, where h_{iit} is estimated by GARCH(1, 1). The LM tests suggest that some conditional heteroskedasticity is captured by the model. The Ljung-Box tests for the squared standardized residuals, $e^2_{it}h^{-1}_{iit}$, become insignificant for most currencies. If the conditional mean in <1> and the conditional variance in <2> are correctly specified, the standardized error $e_{it}h^{-1/2}_{iit}$ should have zero unconditional mean and unit unconditional variance, and be iid. All of the standardized residuals have mean (m_1) close to zero and variance (m_2) close to unity. The LM tests indicate no serial correlation in the standardized residuals. Hsieh (1989, p.310) shows that with correctly specified conditional mean and variance, the skewness (m_3) and the excess kurtosis (m_4) of the standardized residuals ought to be smaller in modulus than those of e_{it}. Thus a large increase in the absolute size of the coefficient of skewness and/or kurtosis in the standardized residuals over that of the residuals is evidence against the model. If the coefficients of skewness and excess kurtosis in Table 3 are compared with those in Table 1, the skewness is generally lower except for the BP and JY spot series and the kurtosis is generally lower except for SF.

In order to test the validity of the GARCH-X model reported in Table 4, a series of specification tests for the model standardized by $h^{1/2}_{iit}$ are also presented in Table 5, where h_{iit} is estimated by GARCH(1, 1)-X. The LM tests suggest that some coditional heteroskedasticity is captured by the model. The Ljung-Box tests for the squared standardized residuals, $e^2_{it}h^{-1}_{iit}$, become insignificant for some currencies. All of the standardized residuals have mean close to zero and variance close to unity. The LM tests indicate no serial correlation in the standardized residuals. The coefficients of the skewness and the excess kurtosis in Table 5 are smaller in modulus than those in Table 1 (except the kurtosis for SF). The conditional variance may be misspecified for SF.

As is well recognized in Baillie and Bollerslev (1989b) and Hsieh (1989) among many others, daily exchange rate series show a considerable amount of leptokurtosis even after accounting for GARCH(1, 1) or GARCH(1, 1)-X. Although the leptokurtosis is somewhat reduced in the standardized residuals, it is nonetheless significant. Hence robust inference procedures may be called for as is used in Baillie

[**Table 3**] Diagnostics for ECM (1) Standardized by GARCH(1, 1)

	BP		DM		JY		CD	
	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t
$Q^2(20)$	0.286	0.262	0.531	0.450	0.112	0.053	0.643	0.443
AR(1)	0.440	0.644	0.408	0.529	0.403	0.423	0.297	0.428
AR(2)	0.841	0.821	0.897	0.832	0.948	0.988	0.200	0.196
AR(5)	0.998	0.993	0.534	0.569	0.713	0.874	1.000	0.998
AR(10)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ARCH(1)	0.011	0.003	0.188	0.256	0.001	0.004	0.231	0.337
ARCH(2)	0.038	0.012	0.370	0.491	0.003	0.008	0.229	0.282
ARCH(5)	0.099	0.058	0.621	0.605	0.021	0.066	0.371	0.287
ARCH(10)	0.155	0.117	0.770	0.679	0.078	0.177	0.495	0.417
$Q_\rho(20)$	0.275		0.480		0.268		0.554	
$\rho(1)$	0.030		0.376		0.296		0.218	
$\rho(2)$	0.093		0.619		0.431		0.372	
$\rho(5)$	0.322		0.245		0.742		0.461	
$\rho(10)$	0.360		0.293		0.654		0.453	
m_1	-0.038	-0.040	-0.005	-0.005	-0.048	-0.044	-0.024	-0.027
m_2	0.988	0.975	1.000	0.999	0.984	1.009	1.011	1.011
m_3	-0.311	-0.335	0.219	0.210	0.392	0.509	-0.204	-0.173
m_4	1.658	1.849	0.875	0.872	2.017	2.351	1.519	1.649

Note: The number in ( ) is the degree of freedom of each test statistics. All the values are asymptotic p-values except for the coefficients of the four unconditional moments.

and Bollerslev (19990). With the robust LM tests for ARCH due to Wooldridge (1990) (not reported here but available upon request), the leptokurtosis in the standardized residual remains significant for JY and for the spot series of FF. It may be noted that the model for the spot series of FF seems misspecified in conditional variance as the LM tests for ARCH for the standardized residuals are more significant than those for the raw residual.

Under the assumption of constant conditional correlation, the cross product of the standardized residuals, $e_{it}\, e_{jt}(h_{iit}\, h_{jjt})^{-1/2}$, $i \neq j$, should also be serially uncorrelated. The Ljung-Box test is used as in Bollerslev (1990) to test the 20th order serial correlation in the cross product of the standardized residuals and is denoted as Q_ρ (20) in Tables 3 and 5. Also, the LM test for the serial correlation in the cross-product of the standardized residuals of orders of 1, 2, 5, 10 are reported with the notation $\rho(1)$, $\rho(2)$, $\rho(5)$, $\rho(10)$. There is strong evidence against the assumption of constant conditional correlation in FF, while there is marginal evidence in IL and SF. We also use Wooldridge's robust LM tests (not reported but available upon request) which do not show any definitive evidence against the assumption for all currencies. The significant test statistics may rather be resulted from the adversely affected size of the tests due to distributional misspecification or due to changing conditional higher moments over time. It may well be necessary to

[**Table 3**] continued

	FF		IL		SF	
	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t
$Q^2(20)$	0.095	0.101	0.634	0.595	0.001	0.001
AR(1)	0.528	0.548	0.769	0.331	0.600	0.839
AR(2)	0.323	0.354	0.986	0.951	0.676	0.661
AR(5)	0.995	0.990	1.000	1.000	0.473	0.583
AR(10)	0.997	1.000	1.000	1.000	1.000	1.000
ARCH(1)	0.002	0.004	0.340	0.392	0.000	0.000
ARCH(2)	0.000	0.000	0.628	0.687	0.000	0.000
ARCH(5)	0.005	0.002	0.321	0.213	0.000	0.000
ARCH(10)	0.012	0.012	0.428	0.360	0.000	0.001
Q_ρ	0.080		0.549		0.002	
$\rho(1)$	0.000		0.299		0.267	
$\rho(2)$	0.000		0.299		0.267	
$\rho(5)$	0.000		0.013		0.044	
$\rho(10)$	0.000		0.031		0.026	
m_1	0.106	0.101	−0.075	−0.078	0.002	0.000
m_2	0.987	0.988	0.991	0.991	1.018	1.015
m_3	−0.044	−0.138	0.282	0.159	0.232	0.282
m_4	1.176	1.156	1.585	1.537	1.756	1.915

[**Table 4**] Estimates of Bivariate GARCH(1, 1)-X parameters

	DM	CD	FF	IT	SW
ω_1	0.000	0.000	0.000	0.000	0.000
	(8.990)	(0.000)	(8.073)	(6.780)	(8.118)
α_1	0.185	0.092	0.230	0.119	0.044
	(10.536)	(13.660)	(13.911)	(7.153)	(5.003)
β_1	0.666	0.902	0.626	0.452	0.734
	(25.279)	(168.996)	(31.663)	(7.370)	(27.785)
δ_1	0.422	0.049	0.415	0.434	2.764
	(4.505)	(6.427)	(15.505)	(7.237)	(8.996)
ω_2	0.000	0.000	0.000	0.000	0.000
	(9.972)	(0.139)	(10.436)	(7.625)	(7.868)
α_2	0.189	0.096	0.197	0.129	0.046
	(10.767)	(14.678)	(14.786)	(8.173)	(4.924)
β_2	0.652	0.898	0.643	0.396	0.715
	(25.427)	(174.775)	(51.915)	(6.745)	(24.655)
δ_2	0.397	0.049	0.419	0.501	2.903
	(4.412)	(6.805)	(17.760)	(7.955)	(9.134)
ρ	0.996	0.993	0.983	0.977	0.996
	(6332.998)	(2800.988)	(1445.637)	(1038.427)	(5869.457)

Note: Asymptotic t-values are in ( ). BP and JY are not included as we can not find a step of increasing likelihood from many different initial values of θ.

[**Table 5**] Diagnostics for ECM (1) Standardized by GARCH(1, 1)-X

	DM		CD		FF		IL		SF	
	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t	Δs_t	Δf_t
$Q^2(20)$	0.213	0.164	0.990	0.986	0.038	0.049	0.000	0.000	0.000	0.000
AR(1)	0.561	0.664	0.820	0.769	0.332	0.298	0.760	0.840	0.259	0.364
AR(2)	0.997	0.972	0.192	0.185	0.702	0.745	0.978	0.959	0.603	0.611
AR(5)	0.538	0.564	1.000	1.000	1.000	1.000	1.000	1.000	0.164	0.249
AR(10)	1.000	1.000	1.000	1.000	0.997	0.995	1.000	1.000	1.000	1.000
ARCH(1)	0.087	0.115	0.835	0.924	0.045	0.110	0.843	0.691	0.090	0.129
ARCH(2)	0.204	0.270	0.869	0.861	0.119	0.207	0.445	0.524	0.011	0.016
ARCH(5)	0.392	0.364	0.848	0.951	0.169	0.040	0.000	0.000	0.000	0.000
ARCH(10)	0.444	0.350	0.959	0.986	0.009	0.052	0.000	0.000	0.000	0.000
Q_ρ	0.177		0.989		0.035		0.000		0.000	
$\rho(1)$	0.230		0.318		0.272		0.329		0.192	
$\rho(2)$	0.662		0.847		0.005		0.121		0.082	
$\rho(5)$	0.994		0.990		0.060		0.954		0.011	
$\rho(10)$	1.000		1.000		1.000		1.000		0.448	
m_1	−0.005	−0.005	−0.036	−0.038	−0.027	−0.029	−0.004	−0.006	0.016	0.016
m_2	0.999	0.999	0.975	0.985	0.994	0.996	0.999	0.999	0.998	0.998
m_3	0.233	0.230	−0.270	−0.255	0.192	−0.013	0.194	0.122	0.310	0.328
m_4	0.776	0.787	1.294	1.392	2.047	1.613	1.382	1.520	1.416	1.443

Note: The number in ( ) is the degree of freedom of each test statistics. All the values are asymptotic p-values except for the coefficients of the four unconditional moments.

relax the assumption of constant conditional correlation and to consider nonnormal distributions which, however, may be expensive to implement for multivariate cases.

From the above diagnostics, we think the model using bivariate GARCH(1, 1)-X in the system of the error correction models with the constant conditional correlation assumption are reasonably well specified with some reservations for future research.

III.G. The LR tests and Model Selection

Table 6 contains the likelihood ratio (LR) test statistics, the AIC, and the SIC for the three specifications of conditional variances: Model 1 of equation <1> using least squares assuming the conditional homoskedasticity; Model 2 of equations <1>, <2> using BHHH with a bivariate GARCH(1, 1); and Model 3 of equations <1>, <3> using BHHH with a bivariate GARCH(1, 1)-X. For DM, FF, IL and SF, the LR tests are in favor of the GARCH-X model and the model selection criteria also indicate that Model 3 is undominated by Model 1 and Model 2.

III.H. Implications

In sum, the unmodeled conditional heteroskedasticity by GARCH(1, 1) may

[**Table 6**] Likelihood Ratio Test and Model Selection Criteria

Model	Log-Likelihood	LR statistic (p-value)	AIC	SIC
BP				
Model 1	11191.46	146.92 (0.000)	−22358.93	−22297.40
Model 2	11264.92		−22491.85	−22394.44
DM				
Model 1	11774.93	200.24 (0.000)	−23517.87	−23435.84
Model 2	11875.05	28.66 (0.000)	−23704.11	−23586.19
Model 3	11889.38		−23728.77*	−23600.59*
JY				
Model 1	10321.63	488.30 (0.000)	−20611.27	−20529.24
Model 2	10565.78		−21085.57	−20967.65
CD				
Model 1	13809.93	455.90 (0.000)	−27571.87	−27448.82
Model 2	14037.88	0.44 (0.803)	−28013.77*	−27854.83*
Model 3	14038.10		−28010.21	−27841.02
FF				
Model 1	10017.40	1260.10 (0.000)	−20026.81	−20006.30
Model 2	10647.45	417.50 (0.000)	−21272.91	−21216.51
Model 3	10856.20		−21686.41*	−21619.76*
IL				
Model 1	10776.96	169.26 (0.000)	−21537.93	−21496.91
Model 2	10861.59	258.76 (0.000)	−21693.19	−21616.28
Model 3	10990.97		−21947.95*	−21860.79*
SF				
Model 1	11175.71	435.36 (0.000)	−22319.43	−22237.40
Model 2	11393.39	345.62 (0.000)	−22740.79	−22622.87
Model 3	11566.20		−23082.41*	22954.23*

Note: * denotes the minimum of AIC or SIC. Model 1 is the equation <1> estimated by least squares assuming the conditional homoskedasticity; Model 2 is the equations <1>, <2> estimated by maximizing the likelihood using the BHHH with a bivariate GARCH(1, 1); and Model 3 is the equations <1>, <3> by the BHHH with a bivariate GARCH(1,1)-X

be explained by a function of the spread. The conditional variances of prediction errors of the exchange rate changes are positively related to the short run deviation from the cointegrating long-run relationship. This means that when the spread is bigger, the exchange rates are more volatile and uncertainty increases.

If z_{t-1}^2 has additional predictive power for the changing variances of the spot and forward exchange rate changes, this may be exploited to obtain more precise time varying confidence intervals for point forecasts of exchange rate changes, and hence to get better estimates of time varying risk premia. Thus the spread between the spot and forward rates has explanatory power for the time varying risk premia. This interpretation is not standard as most of studies have examined the opposite causal direction from the conditional variance (or conditional standard deviation) to the spread like in the ARCH-M model of Engle, Lilien and Robins (1987). While the spread has been considered to reflect the premia for risk or uncertainty so that time varying conditional variances Granger-cause the spread, our study exhibits the presence of the converse causal relationship from the spread to conditional variances.

IV. CONCLUSIONS

In this paper we investigate a model which seems useful to see how the short run disequilibrium has an effect on uncertainty in predicting cointegrated series. Examining the behavior of the variances over time as a function of disequilibrium is reasonable when one expects increased volatilty due to shocks to the system to propagate on first and second moments. The model is thus appropriate for testing for causality in variance as well as in mean through the error correction term.

While the model is made of the first two conditional moments it seems that these more general specification would ensure a better fit and would be consistent with empirically significant effects in other possible examples in economics. We have conducted the same analysis on monthly short and long term interest rate series, and on monthly interest rate series of the commercial paper and the Treasury bill. These results indicate strong GARCH-X in ECM for those data.

Hence the model seems useful to study the relationship between the short-run deviation from a long-run relationship (spread or disequilibrium) and uncertainty. Further study concerning the formulation of structural models which could justify the empirical specification for the observed GARCH-X effects remains interesting.

REFERENCES

ATTANASIO, ORAZIO P. (1991), "Risk, Time Varying Second Moments and Market Efficiency," *Review of Economic Studies,* 58, 479-494.

BAILLIE, R.T. and T. BOLLERSLEV (1989a), "Common Stochastic Trends in a System of Exchange Rates," *The Journal of Finance*, 44, 167-181.

BAILLIE, R.T. and T. BOLLERSLEV (1989b), "The Message in Daily Exchange Rates: A Conditional Variance Tale," *Journal of Business and Economic Statistics*, 7, 297-305.

BAILLIE, R.T. and T. BOLLERSLEV (1990), "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates," *Review of Economic Studies*, 58, 565-585.

BERNDT, E.K., B.H. HALL, R.E. HALL, and J. HAUSMAN (1974), "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 4, 653-665.

BOLLERSLEV, TIM (1986), :Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

BOLLERSLEV, Tim (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, "*The Review of Economics and Statistics,* 498-505.

BOLLERSLEV, T., R.F. ENGLE and J.M. WOOLDRIDGE (1988), "A Capital Asset Pricing Model with Time Varying Covariances," *Journal of Political Economy*, 96, 116-131.

BOLLERSLEV, T., R.Y. CHOU and K.F. KRONER (1992), "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, forthcoming.

CUMBY, ROBERT E. and JOHN HUIZINGA (1992), "Testing the Autocorrelation Structure of Disturbances in Ordinary Least Squares and Instrumental Variables Regressions," *Econometrica*, 60, 185-195.

DIEBOLD, F.X. (1987), "Testing for Serial Correlation in the Presence of ARCH," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 323-328.

DOMOWITZ, I. and C.S. HAKKIO (1988), "Conditional Variance and the Risk Premium in the Foreign Exchange Market," *Journal of International Economics*, 19, 47-66.

ENGLE, ROBERT F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

ENGLE, R.F., D.M. LILIEN and R.P. ROBINS (1987). "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391-407.

ENGLE, R.F., VICTOR Ng, M. ROTHSCHILD (1990), "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45. 213-238.

GIOVANNINI, ALBERTO and PHILIPPE JORION (1989), "The Time Variation of Risk and Return in the Foreign Exchange and Stock Markets," *The Journal of Finance*, 44, 307-325.

GRANGER, C.W.J. (1983), "Forecasting White Noise," *Applied Time Series Analysis of*

Economic Data, A. Zellner, ed., Washington D.C., US Department of Commerce, Bureau of the Census.

GRANGER, C.W.J. (1991), "Forecasting Stock Market Prices: Lessons for Forecasters," UCSD Working Paper, 91-23.

HARVEY, ANDREW (1990), *The Econometric Analysis of Time Series*, MIT Press.

HSIEH, DAVID A. (1989), "Modeling Heteroskedasticity in Daily Foreign Exchange Rates," *Journal of Business and Economic Statistics*, 7, 307-317.

JOHANSEN, S. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

KARIYA, T., T. TSUKUDA and J. MARU (1990), "Testing the Random Walk Hypothesis for Japanese Stock Prices in S. Taylor Models," Graduate School of Business, University of Chicago, Working Paper 90-94.

LEE, TAE-HWY, HALBERT WHITE AND CLIVE W.J. GRANGER (1992), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Methods," *Journal of Econometrics*, forthcoming.

MCLEOD, A.I. and W.K. LI (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations," *Journal of Time Series Analysis*, 4, 169-176.

NELSON (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Ecocometrica*, 59, 347-370.

PHILLIPS, P.C.B. (1987), "Time Series Regressions with Unit Roots," *Econometrica*, 55, 277-302.

SAWA, T. (1978), "Information Criteria for Determining Among Alternative Regression Models," *Econometrica*, 46, 1273-1291.

TAYLOR, S.J. (1988), *Modelling Financial Time Series*, Wiley, Chichester.

WHITE, HALBERT (1984), *Asymptotic Theory for Econometricians*, Academic Press.

WHITE, HALBERT (1989), "An Additional Hidden Unit Test for Neglected Nonlinearity in Multilayer Feedforward Networks," *Proceedings of the International Joint Conference on Neural Network*, Washington D.C., San Diego: SOS Printing, II:451-455.

WOOLDRIDGE, J.M. (1990), "A Unified Approach to Robust, Regression Based Specification Tests," *Econometric Theory*, 6, 17-43.