THE KOREAN ECONOMIC REVIEW
Volume 13, Number 1, Summer 1997.

THE TEMPORAL AGGREGATION EFFECT ON THE PREDICTABILITY OF EXCHANGE RATE VOLATILITY*

KEUN YEONG LEE**

The temporal aggregation effect on autocorrelation functions for the squares of exchange rate changes is theoretically and empirically analyzed. As the data frequency decreases, autocorrelation functions converge to zero. The paper also compares weighting schemes and out-of-sample performances of several competing models - GARCH(1, 1), homoskedastic, kernel, flat rolling regressions, and Foster and Nelson's weighted rolling regressions models. Temporal aggregation generally aggravates out-of-sample performances of GARCH(1, 1) models and weighted rolling regressions, compared with other models. Low-frequency GARCH(1, 1) models derived from high-frequency GARCH(1, 1) models are not worse than direct low-frequency GARCH(1, 1) models.

I. INTRODUCTION

It is widely known that ARCH effects tend to weaken with less frequently sampled data.[1] In several previous papers, the Ljung-Box test showed that autocorrelations for the squares of exchange rate changes are highly significant for daily data, but insignificant for monthly data(e.g., R. T. Baillie and T. Bollerslev, 1989). Kurtosis also decreases with low-frequency data(F. X. Diebold, 1986). F. C. Drost and T. E. Nijman(1993) also suggested that low-frequency GARCH parameter estimates can be derived from high-frequency GARCH parameter estimates. It is generally known that the GARCH(1, 1) model can forecast high volatility periods better than other models(e.g., K. Y. Lee, 1991; K. D. West and D.

* I would like to thank two anonymous referees for very helpful and detailed comments.

** National Intelligence Educational and Research Institute.

[1] It has well recognized that financial variables, including exchange rates, exhibit volatility clustering. This empirical fact has stimulated explictly modelling time variation in second moments. The Autoregressive Conditional Heteroskedasticity(ARCH) model of Engle(1982) is one of the most popular methods. Most asset pricing theories relate first moments to second moments. Therefore, forecast of volatility is also crucially important in asset pricing theories and dynamic hedging strategies. See Bollerslev, Chou, and Kroner(1992) for a review of the theory and empirical evidence in this field.

Cho, 1992). Therefore, it is interesting to study the temporal aggregation effect on the predictability of exchange rate volatility, using several important models for the conditional variance.

First, in the paper, the temporal aggregation effect on the autocorrelation function and excess kurtosis is theoretically and empirically examined. Serial autocorrelation in squared residuals for high-frequency data is indicative of strong conditional heteroskedasticity. But if squared exchange rate changes are not autocorrelated, as the frequency of data decreases, the homoskedastic model may be better than a conditional heteroskedastic model under temporal aggregation. Second, weighting schemes among GARCH(1, 1), kernel, flat rolling regressions, and Foster and Nelson's(1991) weighted rolling regressions models are compared. Some models have similar weighting schemes. For example, the GARCH model is a kind of one-sided weighted rolling regression. Third, the paper examines out-of-sample accuracy of these models using daily, weekly, and monthly data. The predictability of exchange rate volatility can be influenced by both temporal aggregation and weighting schemes in models. Drost and Nijman(1993) argued that a strong GARCH process aggregates to some weak GARCH process. It suggests that it is also interesting to examine out-of-sample accuracy of the GARCH model based on daily exchange rate volatility for every day of the past weeks or months. Finally, it is explicitly tested whether forecast errors differ among competing models.

The paper is organized as follows: In Section Ⅱ, theoretical background is discussed. Temporal aggregation of the autocorrelation function is considered. Section Ⅲ considers conditional variance models for out-of-sample comparisons and estimation methods. Weighting schemes in GARCH, kernel, and rolling regressions models are also compared. Section Ⅳ contains empirical results. Out-of-sample performances of competing models are examined. The null hypothesis of no difference in the accuracy of competing models is also explicitly tested. Section Ⅴ summarizes the major findings of this study.

Ⅱ. THEORETICAL BACKGROUND

Baillie and Bollerslev(1989) pointed out that autocorrelations for the squares of exchange rate changes are highly significant for daily data, but insignificant for monthly data and that kurtosis also decreases with less frequently sampled data. In this Section, a theoretical framework is provided to analyze the temporal aggregation effect on the autocorrelation function and excess kurtosis. First, the autocorrelation function for the aggregated squared series is derived by using autocorrelation functions for the original squared series and excess kurtosis for the original series. For high-frequency financial data, large squared changes tend to be followed by large squared changes. Serial autocorrelation in squared residuals for high-frequency data is indicative of strong conditional heteroskedasticity. Autocorrelation functions for the aggregated squared series can be employed to check whether strong conditional heteroskedasticity also exists under temporal aggregation. Second, excess kurtosis for the aggregated process was obtained. Excess kurtosis is used to investigate whether the unconditional distribution of squared exchange rate changes converges to normality under temporal aggregation. The following notation is employed:

$\gamma_n^*(\gamma_n)$: the nth-order autocovariance for the aggregated(original) squared series

$\gamma_0^*(\gamma_0)$: the variance for the aggregated(original) squared series

$\rho_n^*(\rho_n)$: the nth-order autocorrelation function for the aggregated(original) squared series

$EK_y(EK_\varepsilon)$: excess kurtosis for the aggregated(original) series

Proposition 1

Given

(a) $\varepsilon_t = \eta_t \sigma_t, \quad \eta_t \sim i.i.d.\ F(0, 1)$,

where $F(0, 1)$ specifies an arbitrary distribution with mean zero and unit variance;

(b) the k-period temporal aggregate defined as:

$$(1) \quad y_{k,t} = \sum_{i=1}^{k} \varepsilon_{k(t-1)+i}, \; t = 1, \cdots, \text{int}[x],$$

where int[x] denotes the integer part of x;
then the nth-order autocovariance (γ_n^*) for the aggregated squared series only depends on the autocovariances (γ_{kn+i}) for the original squared series:

$$(2) \quad \gamma_n^* = \sum_{i=1-k}^{k-1} (k - |i|)\gamma_{kn+i}, \; k > 1$$

Proof. See appendix.

Proposition 2

Given

(a) the conditions of Proposition 1;

(b) excess kurtosis defined as:

(3) $EK_\varepsilon = \{E(\varepsilon_t^4) - 3[E(\varepsilon_t^2)]^2\}/[E(\varepsilon_t^2)]^2$;

then the variance (γ_0^*) for the aggregated squared series depends on the variance (γ_0) and the autocovariances (γ_i) for the original squared series, and excess kurtosis (EK_ε) for the original series:

$$(4)\quad \gamma_0^* = \left[k + \frac{2k(k-1)}{EK_\varepsilon + 2}\right]\gamma_0 + 6\sum_{i=1}^{k-1}(k-i)\gamma_i,\ k > 1$$

Proof. See appendix.

Proposition 3

Under the conditions of Proposition 2, the nth-order autocorrelation function (ρ_n^*) for the aggregated squared series and excess kurtosis (EK_y) for the aggregated series depend on the autocorrelation functions (ρ_i) for the original squared series and excess kurtosis (EK_ε) for the original series respectively:

$$(5)\quad \rho_n^* = \frac{\sum_{i=1-k}^{k-1}(k - |i|)\rho_{kn+i}}{k + 6\sum_{i=1}^{k-1}(k-i)\rho_i + \frac{2k(k-1)}{(EK_\varepsilon + 2)}},\quad k > 1$$

and

$$(6)\quad EK_y = \frac{EK_\varepsilon}{k} + 6(EK_\varepsilon + 2)\frac{\sum_{i=1}^{k-1}(k-i)\rho_i}{k^2},\quad k > 1$$

Proof. See appendix.

In equations (5) and (6), ρ_n^* and $EK_y \rightarrow 0$, as $k \rightarrow \infty$. That is, conditional heteroskedasticity disappears and the convergence towards normality is to be expected, as k increases.

A simple Monte Carlo experiment was carried out in order to check the validaty of the proposition 3.[2)] We generated 1000 samples of size 1000 by the sim-

[2] The ARCH(p) process($\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \eta_t$) has the following Yule-Walker equations:

$$\rho_n = \sum_{i=1}^{p} \alpha_i \rho_{n-i},\ n = 1, 2, \cdots$$

ple first order ARCH process as:

(7) $\varepsilon_t = N_t(0, 1)(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{1/2}$

For each of the 1000 samples, ρ_1^* and EK_y are calculated from equations (5) and (6) respectively for given α_0 and α_1. Table 1 shows $\overline{\rho_1^*}$ and $\overline{EK_y}$ as a function of k. $\overline{\rho_1^*}$ and $\overline{EK_y}$ imply the average of the first-order autocorrelation coefficient for the k-period aggregated squared series and excess kurtosis for the k-period aggregated series for 1000 samples respectively. As k increases, $\overline{\rho_1^*}$ and $\overline{EK_y}$ decrease. Autocorrelation coefficients for the aggregated squared series can be employed to check whether strong conditional heteroskedasticity also exists under temporal aggregation.

Ⅲ. MODELS FOR CONDITIONAL VARIANCE

The stylized facts about exchange rates are that they are linearly unpredictable and conditionally heteroskedastic. Some empirical papers examined the out-of-sample forecasting performance to evaluate whether nonlinearities in exchange rate models are important. Most failed to find important nonlinearities. The results prompt researchers to take interest in the conditional variance, under the assumption that the conditional mean follows a random walk. Consider the following time series process:

(8) $(\ln S_t - \ln S_{t-1}) \times 100 = b + \varepsilon_t$

(9) $(\varepsilon_t \mid \Omega_{t-1}) \sim F(0, \sigma_t^2)$

(10) $\sigma_t^2 = m(\Omega_{t-1}) = E(\varepsilon_t^2 \mid \Omega_{t-1})$

Then, using equations (5) and (6) with Yule Walker equations, we can derive ρ_n^* and EK_y for given α_i and EK_y. For example, in the simple ARCH(1) case, the autocorrelation function for the squared series is $\rho_n = \alpha_1^n$, $n \geq 0$. Plugging it into equations (5) and (6), we obtain ρ_n^* and EK_y for given α_1 and EK_y. As shown in Table, ρ_1^* and EK_y converge to zero, as k increases.

	k	ρ_n^*	EK_y
$\alpha_1 = 0.3$, $EK_\varepsilon = 3$	1	0.300	3.000
	5	0.027	2.439
	20	0.003	0.747
$\alpha_1 = 0.5$, $EK_\varepsilon = 5$	1	0.500	5.000
	5	0.065	6.145
	20	0.008	2.140

[Table 1] Monte Carlo simulations

	k	$\overline{\rho_1^*}$	$\overline{EK_y}$
$\alpha_0 = 0.5$ $\alpha_1 = 0.5$	1 5 20	0.416 0.039 0.001	5.222 3.969 1.301
$\alpha_0 = 0.3$ $\alpha_1 = 0.7$	1 5 20	0.482 0.062 0.004	15.139 11.953 4.231
$\alpha_0 = 0.1$ $\alpha_1 = 0.9$	1 5 20	0.513 0.082 0.006	39.112 31.265 11.498

Note: $\overline{\rho_1^*}$ and $\overline{EK_y}$ imply the average of the first-order autocorrelation coefficient for the k-period aggregated squared series and excess kurtosis for the k-period aggregated series for 1000 samples respectively.

where S is a spot exchange rate and the conditioning set $\Omega_{t-1} = [\varepsilon_{t-1}, \cdots, \varepsilon_{t-\infty}]$. F is an arbitrary distribution. The regression estimator m can be derived using a variety of parametric and nonparametric methods. Several model specifications stem from the empirical fact that large squared exchange rate changes tend to be followed by large squared changes. The paper examines four important models for the conditional variance - GARCH(1, 1), kernel, flat rolling regressions, and weighted rolling regressions models.

3.1. GARCH

Several authors have argued that the GARCH(1, 1) model is useful for describing exchange rate volatility(e.g., R. F. Engle and T. Bollerslev, 1986; D. A. Hsieh, 1989; Baillie and Bollerslev, 1989). In the paper, the following GARCH(1, 1) model is used to investigate exchange rate volatility:

$$(11) \quad \varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim N(0, 1)$$

$$(12) \quad \sigma_t^2 = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

where ε_t is an error term with zero mean and conditional variance $E(\varepsilon_t^2 \mid \Omega_t) = \sigma_t^2$. The conditional normal distribution is usually considered in the GARCH model. In the GARCH model, the conditional variance is a linear function of lagged conditional variances and past errors, leaving the unconditional variance constant. The GARCH model can be interpreted as an analogy to the ARMA model for the conditional mean. By substituting $\alpha_0 + \alpha \varepsilon_{t-2}^2 + \beta \sigma_{t-2}^2$ into σ_{t-1}^2 in equation (12) repeatedly, one obtains:

$$(13)\quad \sigma_t^2 = \tau + \sum_{i=1}^{t-1} \lambda_{it} \varepsilon_i^2$$

where $\lambda_{it} = \alpha \beta^{t-i-1}$ and $\tau = \alpha_0 (1 - \beta^{t-1})/(1-\beta) + \beta^{t-1} \sigma_1^2$. Maximum likelihood estimates for the normal distribution in the GARCH(1, 1) model were obtained, using the BFGS(Broyden, Fletcher, Goldfarb, and Shanno) algorithm.

Following Engle and Bollerslev(1986) in the GARCH(1, 1) model, the one-step-ahead conditional variance forecast can be written as:

$$(14)\quad \hat{\sigma}_{T+1/T}^2 = \hat{\alpha}_0 + \hat{\alpha} \hat{\varepsilon}_T^2 + \hat{\beta} \hat{\sigma}_T^2$$

Lee(1991) and West and Cho(1992) constructed weekly time series models based on the past Wednesday data and compared their out-of-sample accuracy. But in the same situation, it is also possble to build weekly time series models based on daily exchange rate volatility. To analyze the temporal aggregation effect on a GARCH process, Drost and Nijman(1993) defined three types of the GARCH processes; strong GARCH, semi-strong GARCH, and weak GARCH. Rescaled innovations are independent in strong GARCH and are uncorrelated in semi-strong GARCH. In a weak GARCH process, only projections of the conditional variance are considered. They argued that if the original squared series follows the symmetric weak GARCH(1, 1) model, then the aggregated squared series also follows the symmetric weak GARCH(1, 1) process.[3] That is, a strong or semi- strong GARCH process aggregates to some weak GARCH process that is not semi-strong GARCH. But their simulation results showed that the quasi-maximum likelihood estimates (QMLE) are not different from the true parameters, even if the low-frequency model follows a weak GARCH process. High-frequency ARMA processes generally aggregate to low-frequency ARMA processes(e.g., F. C. Palm and T. E. Nijman, 1984). In Drost and Nijman(1993) low-frequency GARCH(1, 1) estimates are derived from high-frequency GARCH(1, 1) estmates as:

---

[3] In Proposition 1, a strong GARCH assumption was invoked. But this assumption is strong. The proof in Propositions 1~3 just needs:

$$E(\varepsilon_{kt-p}\varepsilon_{kt-q},\ \varepsilon_{k(t-n)-r}\varepsilon_{k(t-n)-s}) = 0$$
$$\text{if } p \neq q \text{ or } r \neq s (p, q, r, s = 0, \cdots, k-1)$$

Drost and Nijman(1993) replaced this technical condition with the symmetry condition.

(15) $\alpha_0^* = k\alpha_0[1-(\alpha+\beta)^k]/[1-(\alpha+\beta)]$

(16) $\alpha^* = (\alpha+\beta)^k - \beta^*$

(17) $$\frac{\beta^*}{1+\beta^*} = \frac{a(\alpha,\beta,EK_\varepsilon,k)(\alpha+\beta)^k - b(\alpha,\beta,k)}{a(\alpha,\beta,EK_\varepsilon,k)[1+(\alpha+\beta)^{2k}] - 2b(\alpha,\beta,k)}$$

$|\beta^*| < 1$ is the solution of the quadratic equation (17).

where

$$a(\alpha,\beta,EK_\varepsilon,k) = k(1-\beta)^2 + 2k(k-1)(1-\alpha-\beta)^2(1-2\alpha\beta-\beta^2)/ (EK_\varepsilon+2)[1-(\alpha+\beta)^2] + 4[k-1-k(\alpha+\beta) + (\alpha+\beta)^k][\alpha-\alpha\beta(\alpha+\beta)]/[1-(\alpha+\beta)^2]$$

and

$$b(\alpha,\beta,k) = [\alpha-\alpha\beta(\alpha+\beta)][1-(\alpha+\beta)^{2k}]/[1-(\alpha+\beta)^2]$$

Using equations (15), (16), and (17), the one-step-ahead conditional variance forecast in the low-frequency GARCH(1, 1) model based on the high-frequency GARCH(1, 1) model can be obtained as:

(18) $\hat{\sigma}^{*2}_{T+1/T} = \hat{\alpha}_0^* + \hat{\alpha}^*\hat{\varepsilon}^{*2}_T + \hat{\beta}^*\hat{\sigma}^{*2}_T$

By substituting $\hat{\alpha}_0^* + \hat{\alpha}^*\hat{\varepsilon}^{*2}_{T-1} + \hat{\beta}^*\hat{\sigma}^{*2}_{T-1}$ into equation(18) repeatedly, we can rewrite(18) as:

(19) $\hat{\sigma}^{*2}_{T+1/T} = \hat{\tau}^* + \sum_{i=1}^{T}\hat{\lambda}^*_{iT+1}\hat{\varepsilon}^{*2}_i$,

where $\hat{\lambda}^*_{iT+1} = \hat{\alpha}^*\hat{\beta}^{*T-i}$ and $\hat{\tau}^* = \alpha_0^*(1-\hat{\beta}^{*T})/(1-\hat{\beta}^*) + \hat{\beta}^{*T}\hat{\sigma}^{*2}_1$.

3.2. Kernel methods

The nonparametric approach could also be employed to approximate the conditional variance. Nonparametric estimation methods allow one to estimate an unknown response function, without reference to a specific form. One of the simplest nonparametric methods is kernel estimation. In this study we use the following nonparametric kernel estimation model, which has previously been studied by A. R. Pagan and A. Ullah(1988) and A. R. Pagan and G. W. Schwert(1990).[4]

$$(20)\quad \hat{\sigma}_t^2 = \sum_{i=1}^{T} W_{it}\,\hat{\varepsilon}_i^2, \quad \sum_{i=1}^{T} W_{it} = 1,$$

where T is the sample size. By letting z_t be the $r \times 1$ vector containing the elements in $\Omega_{t-1}(z_t = [\hat{\varepsilon}_{t-1}, \cdots, \hat{\varepsilon}_{t-r}]')$, the weighting sequence W_{it} is defined by

$$(21)\quad W_{it} = K_h(z_t - z_i) / \sum_{k=1}^{T} K_h(z_k - z_t),$$

where $K_h(u)$ is the kernel with the bandwidth h. The following convolution kernel is chosen in this study which was described by H. J. Bierens(1990).

$$(22)\quad K_h(u_1, \cdots, u_r) = \prod_{j=1}^{r} K_h(u_j),$$

$$\begin{aligned} K_h(u_j) &= 1/2\,|u_j|^3 - u_j^2 + 2/3 && \text{if } 0 \le |u_j| < 1, \\ &= 1/6\,|u_j|^3 + u_j^2 - 2|u_j| + 4/3 && \text{if } 1 \le |u_j| < 2, \\ &= 0 && \textit{elsewhere}, \end{aligned}$$

where $u_j = (\hat{\varepsilon}_{i-j} - \hat{\varepsilon}_{t-j})/h$. The bandwidth is set to $c\,\hat{\sigma}_{tj} T^{-\omega}$ $(0 < c < \infty,\ 0 < \omega < 1/6)$, where $\hat{\sigma}_{tj}$ is the sample standard deviation of $\hat{\varepsilon}_{t-j}$. The procedure has been applied, using a grid search with 30 grid points for c from 0.1 to 3 for the out-of-sample analysis. ω is given to 0.05.

In nonparametric kernel models, the one-step-ahead conditional variance forecast can be expressed as:

$$(23)\quad \hat{\sigma}_{T+1/T}^2 = \sum_{i=1}^{T} W_{iT+1}\,\hat{\varepsilon}_i^2,$$

where $W_{iT+1} = K_h(z_{T+1} - z_i) / \sum_{k=1}^{T} K_h(z_k - z_{T+1})$ and $z_{T+1} = \hat{\varepsilon}_T$.

3.3. Rolling regressions

There are several strategies for estimating time-varying variances and covariances - block constant covariance estimation methods(e. g., R. C. Merton, 1980; J. M. Poterba and L. H. Summers, 1986; K. R. French, G. W. Schwert, and R. F. Stambaugh, 1989), one-sided rolling regression methods(e. g., E. F. Fama and J. D. Macbeth, 1973), and two-sided rolling regression methods(e. g., R. R. Officer,

[4] When the bandwidth h goes to infinity, the predicted value of kernel models is the same as that of the homoskedastic model, because it is equal to the mean of $\hat{\varepsilon}_t^2$.

1973; Merton, 1980). In one-sided flat rolling regressions, the conditional variance estimate is:

$$(24)\quad \hat{\sigma}_t^2 = \sum_{i=t-n}^{t-1} \delta_{it}\hat{\varepsilon}_i^2,$$

where

$$(25)\quad \delta_{it} = 1/n \quad \text{if } t-n \le i \le t-1,$$
$$= 0 \quad \text{elsewhere,}$$

and where n is the window length.

D. P. Foster and D. B. Nelson(1991) suggested that flat weighting schemes such as one-sided or two-sided rolling regressions or block-constant estimators are inefficient. They altered the shape of weights from a block shape to an exponential decline:

$$(26)\quad \delta_{it}^* = 3^{\frac{1}{2}}/n\ \mathrm{EXP}[-3^{\frac{1}{2}}(t-i)/n]$$

They showed that the asymptotic variance of the normalized measurement error process derived using δ_{it}^* is lower than its asymptotic variance derived by using δ_{it}.[5] They also derived asymptotically optimal window lengths as well as optimal weights for two-sided weighted rolling regressions.

In one-sided flat and weighted rolling regressions, the one-step-ahead conditional variance forecast is:

$$(27)\quad \hat{\sigma}_{T+1/T}^2 = \sum_{i=T-n+1}^{T} \delta_{iT+1}\hat{\varepsilon}_i^2 \ (\text{or} \sum_{i=T-n+1}^{T} \delta_{iT+1}^*\hat{\varepsilon}_i^2),$$

where

$$\sigma_{iT+1} = 1/n \quad (\text{or } 3^{1/2}/n\ \mathrm{EXP}[-3^{1/2}(T-i+1)/n]) \quad \text{if } T-n+1 \le i \le T,$$
$$= 0 \quad \text{elsewhere}$$

[5] The normalized measurement error process is defined as:

$${}_HQ_t \equiv H^{-1/4}({}_H\hat{\Omega}_t - {}_HQ_t)$$

where H is a finite time interval, ${}_H\Omega_t$ is the conditional variance-covariance matrix of a random process, and ${}_H\hat{\Omega}_t$ is the rolling regression estimator of ${}_H\Omega_t$. Then, the normalized measurement error process is asymptotically distributed $N(0, {}_HC_T)$.

4. Comparisons

Weighting schemes in equation (26) are different from those in equation (20). Weights in one-sided weighted rolling regressions depend on time. When we predict the conditional variance at time $T+1$, the shape of weights in equation (26) is exponentially declining, as the distance from $T+1$ increases. But in nonparametric kernel methods, the shape depends on z_t, the $r \times 1$ vector containing the elements in $\Omega_{t-1}(z_t = \hat{\varepsilon}_{t-1}$ in this paper). If z_i $(= \hat{\varepsilon}_{i-1}, i = 1, \cdots, T)$ is nearer to z_{T+1} $(= \hat{\varepsilon}_T)$, we give $\hat{\varepsilon}_i^2$ more weight.[6]

The GARCH model amounts to a one-sided weighted rolling regression(weights are non-negative). By considering ARCH(p) models for arbitrarily large but finite order, we can use GARCH models to approximate Foster and Nelson's (1991) weighted rolling regressions. For $p \geq 1$, equation (13) can be rewritten as:

$$(28) \quad \sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i \hat{\varepsilon}_{t-i}^2$$

where $a_0 = \alpha_0(1-\beta^p)/(1-\beta) + \beta^p \sigma_{t-p}^2$ and $a_i = \alpha\beta^{i-1}$. If a_0 in equation (28) is ignored, the weighting scheme of the GARCH(1, 1) model is similar to that of the weighted rolling regression. The shape of weights in equation (28)$(a_i = \alpha\beta^{i-1})$ is declining, as that in equation (26), when time is far away from t. Table 2 compares weighting schemes of GARCH(1, 1), kernel, flat rolling regressions, and Foster and Nelson's weighted rolling regressions models.

Ⅳ. EMPIRICAL RESULTS

1. A Data and summary statistics

Daily spot exchange rates used in this study are from the EHRA macro data tape from the Federal Reserve Board. The data are quoted by banks at noon New York time. The sample period is from January 27, 1975 through January 9, 1991 and consists of 4001 observations for three spot exchange rates relative to the U. S. dollar - the Japanese yen, the German mark, the British pound. We delete all missing values(e.g., holidays) in daily exchange rate series. All are measured in US $ per unit of foreign currency. Our interest centers on percentage exchange rate changes of which we have 4000 daily observations. In order to investigate the temporal aggregation effect on the predictability of exchange rate

---

[6] The shape of weights of flat rolling regressions is also similar to that of N-nearest neighbor methods. Foster and Nelson's(1991) weighted rolling regressions are similar to locally weighted regressions, because the conditional variance estimate is a weighted, not a simple, average of $\hat{\varepsilon}_t^2$. In flat and weighted rolling regressions, the shape depends on time, but not on the elements in Ω_{t-1}.

[Table 2] Weighting schemes of models

model	$\hat{\sigma}^2_{T+1/T}$	weighting scheme	comparison
GAR	$a_0+\sum_{i=1}^{p} a_i \varepsilon^2_{T-i+1}$	$a_i=\alpha\beta^{i-1}$	time$\rightarrow -\infty \Rightarrow a_i \downarrow$
KER	$\sum_{i=1}^{T} W_{iT+1}\hat{\varepsilon}^2_i$	$W_{iT+1}=K_h(z_{T+1}-z_i)/\sum_{k=1}^{T} K_h(z_k-z_{T+1})$	$z_i \rightarrow z_{T+1} \Rightarrow W_{iT+1}\uparrow$
FLAT	$\sum_{i=T-n+1}^{T} \delta_{iT+1}\hat{\varepsilon}^2_i$	$\delta_{iT+1}=1/n$	δ_{iT+1} : constant
F-N	$\sum_{i=T-n+1}^{T} \delta^*_{iT+1}\hat{\varepsilon}^2_i$	$\delta^*_{iT+1}=3^{\frac{1}{2}}/n\mathrm{EXP}[-3^{\frac{1}{2}}(T-i+1)/n]$	time$\rightarrow -\infty \Rightarrow \delta^*_{iT+1}\downarrow$

Note: 1. $z_{T+1}=[\hat{\varepsilon}_T,\cdots,\hat{\varepsilon}_{T-r+l}]'$

2. GAR, KER, FLAT, and F-N imply GARCH(1, 1), kernel, flat rolling regressions, and Foster and Nelson's weighted rolling regressions models respectively.

volatility, the paper also studies weekly and monthly(every fourth week) series derived from daily data. Every fourth week is used, instead of an end of month series as monthly data. We have 800 weekly observations and 200 monthly observations. That is, consider a time series of daily log differences $[D_t]_{t=1}^{4000}$ and form the k period temporal aggregate:[7]

$$(29)\quad Y_{k,t}=\sum_{i=1}^{k} D_{k(t-1)+i},\ t=1,\cdots,\mathrm{int}[x],$$

where int[x] denotes the integer part of x. The weekly time series corresponds to the $k=5$ business day aggregate $[W_{k,j}]_{j=1}^{800}$ and the monthly exchange rate series corresponds to the $k=20$ business day aggregate $[M_{k,s}]_{s=1}^{200}$.

Baillie and Bollerslev(1989) showed that using the Ljung-Box test, autocorrelations for the squared exchange rate changes are highly significant for daily se-

[7] Let $s_t=\ln S_t$ and $D_t=s_t-s_{t-1}$ where S is a daily spot exchange rate.

$t=0$ s_0	1 s_1	2 s_2	3 s_3	4 s_4	5 s_5	6 s_6	7 s_7	8 s_8	9 s_9	10 s_{10}
Daily log difference	s_1-s_0 D_1	s_2-s_1 D_2	s_3-s_2 D_3	s_4-s_3 D_4	s_5-s_4 D_5	s_6-s_5 D_6	s_7-s_6 D_7	s_8-s_7 D_8	s_9-s_8 D_9	$s_{10}-s_9$ D_{10}
Weekly log difference			s_5-s_0 $W_{5,1}$					$s_{10}-s_5$ $W_{5,2}$		
Monthly log difference										$s_{20}-s_0$ $M_{20,1}$

[Table 3] Summary statistics on ε_t and ε_t^2 (sample size = 4000/*k*)

	Japan			Germany			U.K.		
k	1	5	20	1	5	20	1	5	20
SD	0.621	1.484	3.305	0.641	1.457	3.304	0.645	1.482	3.242
SK	0.450	0.569	0.504	−0.053	0.253	0.353	−0.059	0.042	0.273
EK	4.290	2.483	0.752	4.531	1.003	0.687	4.531	1.063	0.232
$Q^2(10)$	306.4	31.67	8.037	278.8	69.44	12.41	416.7	76.12	31.11
[p-value]	[0.00]	[0.00]	[0.63]	[0.00]	[0.00]	[0.26]	[0.00]	[0.00]	[0.00]
R/S	2.22**	1.58	1.39	2.13**	1.73*	1.50	2.01**	1.40	1.15
ρ_1	0.171	0.117	0.047	0.084	0.089	0.163	0.110	0.067	−0.031
ρ_2	0.084	0.022	−0.020	0.105	0.108	0.037	0.105	0.025	−0.023
ρ_3	0.095	0.086	0.076	0.076	0.205	0.007	0.089	0.155	0.358
ρ_4	0.048	0.068	0.054	0.100	0.075	−0.112	0.139	0.156	0.025
ρ_5	0.116	0.066	0.142	0.110	0.027	0.060	0.124	0.047	0.003

Note: 1. The weekly(monthly) series corresponds to the $k = 5(20)$ business day aggregate.
2. SD: Standard deviation
3. SK(EK): Skewness(excess kurtosis)
4. $Q^2(10)$: Ljung-Box statistics for 10th-order correlation in $\hat{\varepsilon}_t^2$.
5. Let $\varphi(r) = \sum_{t=1}^{[rT]} (\hat{\varepsilon}_t^2 - \hat{\mu}_2)$, where $0 \leq r \leq 1$, T is the sample size, $[\cdot]$ is the "integer part of", and $\hat{\mu}_2$ is the sample variance of $\hat{\varepsilon}_t$. Then the mo- dified R/S statistic Q_n is: $Q_n = [Max\ \varphi(r) - Min\ \varphi(r)]/(T\hat{s})^{1/2}$, where $\hat{s}$ is an estimate of the asymptotic variance. * and ** indicate significance at the 0.10 and 0.05 levels respectively (see Table 1a in Haubrich and Lo, 1989).
6. ρ_i: the *i*th-order sample autocorrelation coefficient for $\hat{\varepsilon}_t^2$.

ries, but insignificant for monthly series. They also pointed out that excess kurtosis tends to decrease, as the frequency of data decreases. Summary statistics on ε_t and ε_t^2 are reported to examine their arguments. As shown in Table 3, excess kurtosis decreases, but skewness increases with less frequently sampled data. The null hypothesis of conditional homoskedasticity is tested, using the Ljung-Box test statistic $Q^2(p)$ for the *p*th-order serial correlation in $\hat{\varepsilon}_t^2$. The null hypothesis of uncorrelated squared changes is decisively rejected for daily data, but not for monthly data. It implies that the homoskedastic model may be good for monthly data, but not for daily data. Low-frequency autocorrelation coefficients are generally smaller than high-frequency autocorrelation coefficients. The modified rescaled range statistic[8] (R/S in Table 3) is also considered, in order to test the constancy of the unconditional variance. The test statistic is mentioned in Table 3. The null hypothesis of constancy of the unconditional variance is rejected for the full daily sample. Stationary models such as nonparametric methods may be inappropriate for the full daily data. But we cannot assert that test statistics for the whole sample size are closely related to the out-of-sample performance, because half subsamples are used in a rolling manner in out-of-sample comparisons.

2. Out-of-sample RMSE

We examine the temporal aggregation effect on the one-step-ahead out-of-sample predictive performance of exchange rate volatility, using GARCH(1, 1), kernel, homoskedastic, one-sided flat, and weighted rolling regressions models.

As shown in Section Ⅲ, low-frequency GARCH(1, 1) estimates can be derived from high-frequency GARCH(1, 1) estimates. Therefore, we also examine the out-of-sample forecasting performance in weekly or monthly series using daily GARCH(1, 1) parameter estimates as well as direct weekly or monthly GARCH (1, 1) parameter estimates. Then, which estimates are better in out-of-sample accuracy? In ARMA processes, H. Lütkepohl(1986) showed that if more information is employed, better forecasts can be obtained, by using asymptotic theory and small sample simulation results. But W. P. Cleveland and G. C. Tiao(1979) suggested that for some time series a different model may be needed for different seasons. T. G. Anderson and T. Bollerslev(1994) argued that the intradaily GARCH models don't conform closely to the theoretical aggregation results.

Table 4 shows GARCH(1, 1) parameter estimates. In Table 4, weekly (or monthly) GARCH(1, 1) estimates implied by daily GARCH(1, 1) estimates are derived by plugging daily GARCH(1, 1) estimates into equations (15), (16), and (17). In the case of Japan, $\hat{\alpha} + \hat{\beta}$ is greater than 1 in the full daily and weekly data. But parameter estimates for the whole sample size don't seem to be related to the out-of-sample performance, because half subsamples are used in a rolling manner in out-of-sample comparisons.

Data are divided into two equal subsamples. The first is used for a within-sample fit, while the second is used for out-of-sample comparisons.[9] The root meam square error (RMSE) is employed to measure out-of-sample accuracy.[10]

[8] See J. G. Haubrich and A. W. Lo(1989), and M. Loretan and P. C. B. Phillips(1992) for the modified rescaled range statistic.

[9] Observations 2001 through 4000 for daily data, observations 401 through 800 for weekly data, and observations 101 through 200 for monthly data were reserved, in order to compare the one-step-ahead out-of-sample performance in each exchange rate volatility series.

[10] The RMSE is the statistical standard criterion which is commonly used. It is also useful and interesting to compare conditional variance models with economic loss functions(e.g., R. F. Engle, T. Hong, A. Kane, 1990; K. D. West, H. J. Edison, and D. Cho, 1993; J. A. Lopez, 1994).

[Table 4] GARCH (1, 1) parameter estimates
(12) $\sigma_t^2 = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

		α	β
Daily	Japan	0.059 (0.008)	0.943 (0.007)
	Germany	0.107 (0.010)	0.888 (0.009)
T=4000	U.K.	0.087 (0.008)	0.903 (0.009)
Direct	Japan	0.097 (0.015)	0.909 (0.014)
weekly	Germany	0.151 (0.039)	0.835 (0.042)
T=800	U.K.	0.101 (0.028)	0.843 (0.037)
Weekly	Japan	0.125	0.884
implied	Germany	0.189	0.789
by daily	U.K.	0.141	0.807
Direct	Japan	0.000 (0.002)	0.994 (0.019)
monthly	Germany	0.071 (0.035)	0.864 (0.076)
T=200	U.K.	0.091 (0.051)	0.755 (0.112)
Monthly	Japan	0.256	0.781
implied	Germany	0.301	0.616
by daily	U.K.	0.183	0.625

Note: 1. Standard errors are in parentheses.
2. Weekly (or monthly) implied by daily are estimated by plugging daily GARCH(1, 1) estimates into equations (15), (16), and (17).

Foster and Nelson(1991) derived asymptotically optimal window lengths for two-sided rolling regressions. But two-sided rolling regressions cannot be considered in the out-of-sample analysis, because leading terms are unknown. Hence, only one-sided rolling regression is considered in the paper. Kernel methods also have bandwidth selection problems.

For rolling regressions, the smallest of RMSE's that was obtained from 20 window lengths($n = 5 \times j$, $j = 1, \cdots, 20$) was reported in Tables 6, 7, and 8. With kernel methods, the smallest of RMSE's which was obtained from 30 bandwidths was also reported.[11] In order to model the conditional variance of the series $\hat{\varepsilon}_t$, conditioning variables z_t must be chosen. Experiments are conducted with $z_t = \hat{\varepsilon}_{t-1}$ for out-of-sample comparisons among competing models.[12] In the

[11] The Gaussian kernel was also used. But its out-of-sample performances were worse than those of the convolution kernel employed here. There exists some controversy about choice of kernel(e.g., D. W. K. Andrews, 1991; W. K. Newey and K. D. West, 1994).

[12] We can also expand the information set z_t to multiple lags. But these experiments do not generally outperform the one-lag case because of low dimensionality restriction(Pagan and Schwert, 1990).

[Table 5] Bandwidth and window length specifications

	Daily			Weekly			Monthly		
	Japan	Germany	U K	Japan	Germany	U K	Japan	Germany	U K
c (KER)	0.3	0.8	0.5	0.8	2.7	1.8	0.7	0.6	0.6
n (FLAT)	55	20	35	35	30	25	100	95	85
n (F-N)	40	30	35	40	55	40	100	30	70

Note: 1. KER, FLAT, and F-N imply kernel, flat rolling regressions, and Foster- Nelson's weighted rolling regressions respectively.

real world, optimal bandwidths are not known and should be chosen ex ante by cross validation. Table 5 reports bandwidth and window length specifications employed in out-of-sample comparisons.

Flat rolling regressions in the literature generally use fewer than $n=24$ lags(e. g., French, Schwert, and Stambaugh, 1987). R. G. Donaldson, M. Kamstra, and H. Y. Kim(1993) suggested that restricting $n \leq 40$ don't hamper the performance of flat rolling regressions. In this paper, when n is specified between 20 lags and 55 lags in daily and weekly series, flat and weighted rolling regressions have the smallest RMSE. But this is not the case with monthly series.

First, with daily exchange rate volatility series, GARCH(1, 1) models and Foster and Nelson's weighted rolling regressions are better than other models. Large squared changes tend to be followed by large squared changes for high-frequency exchange rate data. The homoskedastic model has the worst performance in all cases.

Second, with weekly exchange rate volatility series, GARCH(1, 1) models are generally worse than other models in the cases of Japan and Germany. Flat rolling regressions beat other models in the cases of Germany and the U. K. In Japan case, the weekly GARCH(1, 1) model derived from the daily GARCH(1, 1) model have the best performance.

Third, with monthly exchange rate volatility series, kernel models beat other models in the cases of Japan and Germany. GARCH(1, 1) models are relatively not good. Direct monthly GARCH(1, 1) estimates have the worst out-of-sample performance in the cases of Japan and U. K., while monthly GARCH(1, 1) estimates based on weekly GARCH(1, 1) estimates have the best performance in the case of U. K.

GARCH(1, 1) model have high persistence; which may account for why it can explain high volatility periods better than homoskedastic, nonparametric kernel, and flat rolling regressions models. Generally speaking, temporal aggregation aggravates out-of-sample performances of GARCH(1, 1) models and Foster and Nelson's rolling regressions. Large squared changes don't tend to be followed by large squared changes in low-frequency exchange rate data. Tables 7 and 8 indi-

[Table 6] Out-of-sample RMSE (daily)

	Japan	Germany	U.K.
1. GARCH	$.879^{1}$	$.951^{2}$	1.089^{2}
2. HO	$.892^{5}$	$.961^{5}$	1.125^{5}
3. KER	$.888^{3}$	$.960^{4}$	1.115^{4}
4. FLAT	$.888^{4}$	$.953^{3}$	1.092^{3}
5. F-N	$.884^{2}$	$.948^{1}$	1.089^{1}
H_s [p-value]	3.43 [0.064] (12)	4.49 [0.034] (6)	3.74 [0.053] (13)
H_m [p-value]	4.65 [0.325] (7)	12.78 [0.012] (7)	14.08 [0.007] (7)

Note: 1. HO, KER, FLAT, and F-N imply homoskedastic, kernel, flat rolling regressions, and Foster and Nelson's weighted rolling regressions models respectively.

2. The number in upper-right corner of RMSE indicates the rank of the model.
3. H_s: the hypothesis of equal MSE's for the best model and the worst model.
4. H_m: the hypothesis of equal MSE's across all five models. H_s and H_m rows give χ^2 statistics.
5. The number in parenthesis was computed by Newey and West(1994)'s automatic lag selection method.

[Table 7] Out-of-sample RMSE (weekly)

	Japan	Germany	U.K.
1. GARCH	5.151^{6}	4.015^{6}	4.202^{2}
2. HO	5.063^{3}	3.921^{3}	4.270^{6}
3. KER	5.053^{2}	3.920^{2}	4.255^{5}
4. FLAT	5.103^{5}	3.896^{1}	4.170^{1}
5. F-N	5.095^{4}	3.926^{4}	4.209^{3}
6. GARCH implied by daily	5.051^{1}	3.969^{5}	4.220^{4}
H_s [p-value]	5.94 [0.015] (1)	0.96 [0.033] (2)	0.88 [0.347] (7)
H_m [p-value]	13.59 [0.018] (0)	7.01 [0.220] (4)	1.30 [0.934] (3)

Note: 1. HO, KER, FLAT, and F-N imply homoskedastic, kernel, flat rolling regressions, and Foster and Nelson's weighted rolling regressions models respectively.

2. GARCH implied by daily is derived by plugging daily GARCH(1, 1) estimates into equations (15), (16), and (17).
3. The number in upper-right corner of RMSE indicates the rank of the model.
4. H_s: the hypothesis of equal MSE's for the best model and the worst model.
5. H_m: the hypothesis of equal MSE's across all six models. H_s and H_m rows give χ^2 statistics.
6. The number in parenthesis was computed by Newey and West(1994)'s automatic lag selection method.

[Table 8] Out-of-sample RMSE (monthly)

	Japan	Germany	U.K.
1. GARCH	20.95[7]	18.44[3]	19.00[7]
2. HO	20.37[2]	18.49[4]	18.83[5]
3. KER	20.11[1]	18.20[1]	18.71[4]
4. FLAT	20.37[3]	18.43[2]	18.70[3]
5. F-N	20.59[4]	18.52[5]	18.85[6]
6. GARCH implied by daily	20.94[6]	19.14[7]	18.70[2]
7. GARCH implied by weekly	20.66[5]	18.74[6]	18.69[1]
H_s [p-value]	1.22 [0.270] (1)	3.33 [0.068] (3)	2.72 [0.099] (3)
H_m [p-value]	n. a.	8.48 [0.205] (1)	12.12 [0.059] (1)

Note: 1. HO, KER, FLAT, and F-N imply homoskedastic, kernel, flat rolling regressions, and Foster and Nelson's weighted rolling regressions models respectively.

2. GARCH implied by daily (or weekly) is derived by plugging daily (or weekly) GARCH(1, 1) estimates into equations (15), (16), and (17).
3. The number in upper-right corner of RMSE indicates the rank of the model.
4. H_s: the hypothesis of equal MSE's for the best model and the worst model.
5. H_m: the hypothesis of equal MSE's across all seven models. H_s and H_m rows give χ^2 statistics.
6. The number in parenthesis was computed by Newey and West(1994)'s automatic lag selection method.

cate that homoskedastic, kernel, and flat rolling regressions models are relatively better than GARCH(1, 1) models in low-frequency data. But kernel methods and rolling regressions have window length specification problems. Low-frequency GARCH(1, 1) models based on high-frequency GARCH(1, 1) models beat other models for Japan case in weekly series and for U. K. case in monthly series respectively.

3. Statistical tests

In this Section, using generalized method of moments(L. P. Hansen, 1982), the null hypothesis of equal forecast accuracy of the following pairs of competing models was formally tested:

H_s: the hypothesis of equal MSE's from the best model and the worst model.
H_m: the hypothesis of equal MSE's across all competing models.

The null hypothesis that $m_i = c$(m_i: MSE of model i, c: constant) can be tested by the following procedure:

$$(30) \quad u_t = m_t - c,$$

where $u_t = (u_{1t}, \cdots, u_{qt})$,
$m_t = (m_{1t}, \cdots, m_{qt})$,

q implies the number of competing models. The orthogonality conditions state that $E(u_t \otimes z_t) = 0$, where z_t is a constant ($E(u_t) = 0$). Then there are 1 parameter and $1 \times q$ orthogonality conditions leaving $q - 1$ overidentifying restrictions. Since the true parameter is unknown, a GMM estimator can be estimated by defining the function $g^*(c) \equiv E(u_t)$. This function has a zero at $c = c^*$, since it is equal to zero under the null hypothesis. The method of moments estimator of the function $g^*(c)$ for a sample is:

$$(31) \quad g_T(c) = T^{-1} \sum_{t=1}^{T} u_t,$$

where T is the sample size. The parameter c can be estimated by minimizing the following criterion function:

$$(32) \quad J_T(c) = g_T(c)' W_T g_T(c),$$

where W_T is a symmetric weighting matrix of orthogonality conditions. The consistent estimate of W_T is formed by:

$$(33) \quad W_T = [R_T(0) + \sum_{j=1}^{k-1} (R_T(j) + R_T(j)')]^{-1},$$

where

$$(34) \quad R_T(j) = T^{-1} \sum_{t=1+j}^{T} u_t u_{t-j}'.$$

The two-step procedure is employed in the paper. F. X. Diebold and R. S. Mariano(1991) argued that k-step-ahead forecast errors tend to be approximately characterized by moving average processes of order $(k-1)$. But, in the paper, j was determined using the automatic lag selection method in Newey and West (1994).

First, with daily series, the H_s test of equal MSE's from the best model and the worst model is rejected at the 0.10 level in all cases. It implies that the GAR-

CH(1, 1) model is significantly better than the homoskedastic model. It is also tested whether MSE's are the same across all models. The H_m test of equal MSE's for five competing models is rejected at the 0.05 level in the cases of Germany and U. K.

Second, in weekly series comparisons, the weekly GARCH(1, 1) model derived from the daily GARCH(1, 1) model is also considered. The H_s test of equal MSE's for the best model and the worst model is rejected at the 0.05 level in the cases of Japan and Germany. The H_m test of equal MSE's across six competing models is rejected at the 0.05 level in the case of Japan.

Third, in monthly series comparisons, the monthly GARCH(1, 1) model based on the daily (or weekly) GARCH(1, 1) model is also considered. The H_s test of equal MSE's for best model and the worst model is rejected at the 0.10 level in the cases of Germany and U. K. The H_m test of equal MSE's for seven competing models is rejected at the 0.10 level in the case of U. K. The null hypothesis of equal MSE's among competing models is rejected more frequently with the daily series than with the monthly series.

V. CONCLUSIONS

The temporal aggregation effect on the predictability of exchange rate volatility has been examined, using daily, weekly, and monthly data. First, temporal aggregation of the autocorrelation function and excess kurtosis is considered. The theoretical and empirical results show that autocorrelations for the squares of exchange rate changes and excess kurtosis decrease with low-frequency data.

The paper further confirms the theoretical and empirical results, as mentioned above, by investigating whether out-of-sample accuracy of a GARCH(1, 1) model is still better than that of the homoskedastic model under temporal aggregation. This study also investigated out-of-sample accuracy of kernel methods, flat rolling regressions and Foster and Nelson's weighted regressions. Weighting schemes of some models are related to each other. In summary, GARCH(1, 1) models are good using daily series, while homoskedastic, kernel, and flat rolling regressions models perform better using monthly series. Kernel models and rolling regressions have window length selection problems. Temporal aggregation generally aggravates out-of-sample performances of GARCH(1, 1) models and Foster and Nelson's weighted rolling regressions. Large squared changes don't tend to be followed by large squared changes in low-frequency exchange rate data. Low-frequency GARCH(1, 1) models based on high-frequency GARCH(1, 1) models are not bad, compared with direct low-frequency GARCH(1, 1) models.

The null hypothesis of no difference in the accuracy of competing models is also explicitly tested. The null hypothesis of equal forecast accuracy is rejected more frequently with the daily series than with the monthly series.

APPENDICES

1. Proof of Proposition 1

(A1) $y^2_{k,t} = (\sum_{i=1}^{k} \varepsilon_{k(t-1)+i})^2$

(A2) $y^2_{k,t-n} = (\sum_{i=1}^{k} \varepsilon_{k(t-n-1)+i})^2$

$\gamma^*_n = cov(y^2_{k,t}, y^2_{k,t-n})$

From assumption,

$cov(\varepsilon_{kt-p}\, \varepsilon_{kt-q}, \varepsilon_{k(t-n)-r}\, \varepsilon_{k(t-n)-s}) = 0$
if $p \neq q$ or $r \neq s$ $(p, q, r, s = 0, \cdots, k-1)$

Then

(A3)
$$\begin{aligned}\gamma^*_n &= \sum_{j=0}^{k-1}\sum_{i=0}^{k-1} cov(\varepsilon^2_{kt-i}, \varepsilon^2_{k(t-n)-j}) = \sum_{j=0}^{k-1}\sum_{i=0}^{k-1}\gamma_{kn+j-i} \\ &= \gamma_{kn-(k-1)} + 2\gamma_{kn-(k-2)} + \cdots + (k-1)\gamma_{kn-1} \\ &\quad + k\gamma_{kn} + (k-1)\gamma_{kn+1} + \cdots\cdots + \gamma_{kn+(k-1)} \\ &= \sum_{i=1-k}^{k-1}(k-|i|)\gamma_{kn+i}.\end{aligned}$$

2. Proof of Proposition 2

(A4) $\gamma^*_0 = var(y^2_{k,t}) = E(y^4_{k,t}) - [E(y^2_{k,t})]^2$

(A5) $E(y^4_{k,t}) = \sum_{i=0}^{k-1} E(\varepsilon^4_{kt-i}) + 6E[\sum_{j=0}^{k-2}\sum_{i=j+1}^{k-1}(\varepsilon^2_{kt-i}\, \varepsilon^2_{kt-j})]$

(A6) $[E(y^2_{k,t})]^2 = \sum_{i=0}^{k-1}[E(\varepsilon^2_{kt-i})]^2 + 2\sum_{j=0}^{k-2}\sum_{i=j+1}^{k-1}[E(\varepsilon^2_{kt-i})E(\varepsilon^2_{kt-j})]$

Plug (A5) and (A6) into (A4),

(A7)
$$\begin{aligned}\gamma^*_0 &= \sum_{i=0}^{k-1} var(\varepsilon^2_{kt-i}) + 6\sum_{j=0}^{k-2}\sum_{i=j+1}^{k-1} cov(\varepsilon^2_{kt-i}, \varepsilon^2_{kt-j}) \\ &\quad + 4\sum_{j=0}^{k-2}\sum_{i=j+1}^{k-1}[E(\varepsilon^2_{kt-i})E(\varepsilon^2_{kt-j})] \\ &= k\gamma_0 + 6\sum_{i=1}^{k-1}(k-i)\gamma_i + 4\frac{k(k-1)}{2}[E(\varepsilon^2_{kt})]^2\end{aligned}$$

$$= \{ k + 4\frac{k(k-1)}{2\gamma_0} [E(\varepsilon_{kt}^2)]^2\} \gamma_0 + 6 \sum_{i=1}^{k-1} (k-i)\gamma_i .$$

$[E(\varepsilon_{kt}^2)]^2 / \gamma_0$ in (A7) can be expressed as excess kurtosis:

$$\text{(A8)} \quad \gamma_0 / [E(\varepsilon_{kt}^2)]^2 = \{E(\varepsilon_{kt}^4) - 3[E(\varepsilon_{kt}^2)]^2\} / [E(\varepsilon_{kt}^2)]^2 + 2 = EK_\varepsilon + 2.$$

Plug (A8) into (A7),

$$\text{(A9)} \quad \gamma_0^* = \left[k + \frac{2k(k-1)}{EK_\varepsilon + 2} \right] \gamma_0 + 6 \sum_{i=1}^{k-1} (k-i)\gamma_i .$$

3. Proof of Proposition 3

From (A3) and (A9),

$$\text{(A10)} \quad \rho_n^* = \frac{\gamma_n^*}{\gamma_0^*} = \frac{\sum_{i=1-k}^{k-1} (k - |i|)\gamma_{kn+i}}{\left[k + \frac{2k(k-1)}{EK_\varepsilon + 2} \right] \gamma_0 + 6 \sum_{i=1}^{k-1} (k-i)\gamma_i}$$

$$= \frac{\sum_{i=1-k}^{k-1} (k - |i|)\rho_{kn+i}}{k + \frac{2k(k-1)}{EK_\varepsilon + 2} + 6 \sum_{i=1}^{k-1} (k-i)\rho_i} .$$

From (A5) and (A6),

$$\text{(A11)} \quad EK_y = \frac{E(y_{k,t}^4)}{[E(y_{k,t}^2)]^2} - 3$$

$$EK_y = \{ (kEK_\varepsilon + 15)[E(\varepsilon_{k,t}^2)]^2 + 6 \sum_{j=0}^{k-2} \sum_{i=j+1}^{k-1} cov(\varepsilon_{kt-i}^2, \varepsilon_{kt-j}^2) + 6 \sum_{j=0}^{k-2} \sum_{i=j+1}^{k-1} E(\varepsilon_{kt-i}^2) E(\varepsilon_{kt-j}^2)\} / k^2 [E(\varepsilon_{kt}^2)]^2 - 3$$

$$= \frac{EK_\varepsilon}{k} + \frac{6 \sum_{i=1}^{k-1} (k-i)\gamma_i}{k^2 [E(\varepsilon_{kt}^2)]^2}$$

$$= \frac{EK_\varepsilon}{k} + 6(EK_\varepsilon + 2) \frac{\sum_{i=1}^{k-1} (k-i)\rho_i}{k^2} .$$

REFERENCES

Anderson, T. G. and T. Bollerslev, "Intraday Seasonality and Volatility Persistence in Foreign Exchange and Equity Markets", Manuscript, Northwestern University, 1994.

Andrews, D. W. K., "Heteroskedasticity and Autocorrelation Consisitent Covariance Matrix Estimation", *Econometrica*, May 1991, 59: 817-858.

Baillie, R. T. and T. Bollerslev, "The Message in Daily Exchange Rates: A Conditional Variance Tale", *Journal of Business and Economic Statistics*, July 1989, 7: 297-305.

Bollerslev, T., R. Y. Chou, and K. F. Kroner, "ARCH Modeling Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, 1992, 52: 5-59.

Bierens, H. J., "Model-Free Asymptotically Best Forecasting of Stationary Economic Time Series", *Econometric Theory*, September 1990, 6: 348-383.

Clevelend, W. P. and G. C. Tiao, "Modeling Seasonal Time Series", *Economie Applique*, 1979, 32: 107-129.

Diebold, F. X., "Temporal Aggregation of ARCH Processes and the Distribution of Asset Returns", *Working Paper*, Federal Reserve Board, Washington, DC., 1986.

Diebold, F. X. and R. S. Mariano, "Comparing Predictive Accuracy I: An Asymptotic Test", Manuscript, *Department of Economics*, University of Pennsylvania, 1991.

Donaldson, R. G., M. Kamstra, and H. Y. Kim, "Evaluating Alternative Models for the Conditional Volatility of Stock Returns: Evidence form International Data", 1993, Manuscript, University of British Columbia.

Drost, F. C. and T. E. Nijman, "Temporal Aggregation of GARCH Processes", *Econometrica*, July 1993, 61: 909-927.

Engle, R. F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U. K. inflation", *Econometrica*, 1982, 50: 987-1008.

Engle, R. F. and T. Bollerslev, "Modeling the Persistence of Conditional Variance", *Econometric Reviews*, 1986, 5: 1-50.

Engle, R. F., T. Hong, and A. Kane, "Valuation of Variance Forecasts with Simulated Options Markets", Manuscript, Department of Economics, UC San Diego, 1990.

Fama, E. F. and J. D. Macbeth, "Risk, Return, and Equilibrium: Empirical Tests", *Journal of Political Economy*, May-June 1973, 81: 607-636.

Foster, D. P. and D. B. Nelson, "Rolling Regressions", Manuscript, University of Chicago, 1991.

French, K. R., G. W. Schwert, and R. F. Stambaugh, "Expected Stock Returns and Volatility", *Journal of Financial Economics*, September 1987, 19: 3-29.

Hansen, L. P., "Large Sample Properties of Generalized Method of Moments Estimations", *Econometrica*, July 1982, 50: 1029-1054.

Haubrich, J. G. and A. W. Lo, "The Sources and Nature of Long-term Memory in the Business Cycle", *NBER Working Paper* No. 2951, 1989.

Hsieh, D. A., "Modelling Heteroskedasticity in Daily Foreign Exchange Rates", *Journal of Business and Economic Statistics*, July 1989, 7: 307-317.

Lee, K. Y., "Are the GARCH Models Best in the Out-of-Sample Performance?", *Economic Letters*, November 1991, 37: 305-308.

Lopez, J. A., "Comparing the Predictive Accuracy of Volatility Models", Manuscript, University of Pennsylvania, 1994.

Loretan, M. and P. C. B. Phillips, "Testing the Covariance Stationarity of Heavy-Tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets", SSRI Working Paper No. 9208, University of Wisconsin-Madison, 1992.

Lütkepohl, H., "Forecasting Vector ARMA Processes with Systematically Missing Observations", *Journal of Business and Economic Statistics*, July 1986, 4: 375-390.

Merton, R. C., "On Estimating the Expected Return on the Market", *Journal of Financial Economics*, December 1980, 8: 323-361.

Newey, W. K. and K. D. West, "Automatic Lag Selection in Covariance Matrix Estimation", 1994, forthcoming, *Review of Economic Studies*.

Officer, R. R., "The Variability of the Market Factor of the New York Stock Exchange", *Journal of Business*, July 1973, 46: 434-453.

Pagan, A. R. and G. W. Schwert, "Alternative Models for Conditional Stock Volatility", *Journal of Econometrics*, July-August 1990, 45: 267-290.

Pagan, A. R. and A. Ullah, "The Econometric Analysis of Models with Risk Terms", *Journal of Applied Econometrics*, April 1988, 3: 87-105.

Palm, F. C. and T. E. Nijman, "Missing Observations in the Dynamic Regression Model", *Econometrica*, November 1984, 52: 1415-1435.

Poterba, J. M. and L. H. Summers, "The Persistence of Volatility and Stock Market Fluctuations", *American Economic Review*, December 1986, 76: 1142-1151.

West, K. D. and D. Cho, "The Predictive Ability of Several Models of Exchange Rate Volatility", Manuscript, University of Wisconsin-Madison, 1992.

West, K. D., H. J. Edison, and D. Cho, "A Utility Based Comparison of Some Models of Exchange Rate Volatility", *Journal of International Economics*, August 1993, 35: 23-46.