COMMON PERSISTENCE IN CONDITIONAL VARIANCES[1]

By Tim Bollerslev and Robert F. Engle

Since the introduction of the autoregressive conditional heteroskedastic (ARCH) model in Engle (1982), numerous applications of this modeling strategy have already appeared. A common finding in many of these studies with high frequency financial or monetary data concerns the presence of an approximate unit root in the autoregressive polynomial in the univariate time series representation for the conditional second order moments of the process, as in the so-called integrated generalized ARCH (IGARCH) class of models proposed in Engle and Bollerslev (1986). In the IGARCH models shocks to the conditional variance are persistent, in the sense that they remain important for forecasts of all horizons. This idea is readily extended to a multivariate framework. Even though many time series may exhibit persistence in variance, it is likely that several different variables share the same common long-run component. In that situation, the variables are naturally defined to be co-persistent in variance, and the co-persistent linear combination is interpretable as a long-run relationship. Conditions for co-persistence to occur in the multivariate linear GARCH model are presented. These conditions parallel the conditions for linear co-integration in the mean, as developed by Engle and Granger (1987). The presence of co-persistence has important implications for asset pricing relationships and in optimal portfolio allocation decisions. An empirical example relating to the time series properties of nominal U.S. dollar exchange rates for the deutschemark and the British pound provides a simple illustration of the ideas.

KEYWORDS: Persistence in variance, co-persistence in variance, generalized autoregressive conditional heteroskedasticity (GARCH), integrated GARCH (IGARCH), factor GARCH, asset pricing, exchange rate dynamics.

1. INTRODUCTION

THE ECONOMIC THEORY OF BEHAVIOR under uncertainty forms the basis for much of modern finance and monetary theory. An agent must make a decision based upon the distribution of a random variable some time in the future. In many rational expectations models it is simply assumed that only the mean of this conditional distribution affects the decision, but for more general utility functions and risk averse agents, some measure of the dispersion will also be of primary importance. However, until recently conventional econometric methods have not been responsive to the need to develop corresponding quantitative measures of risk and uncertainty. This is particularly true in the econometric analysis of time series, where the convenient but often implausible assumption

of linear covariance stationary models with finite unconditional second moments and time invariant conditional variances and covariances remains prominent.

To generalize this assumption, Engle (1982) introduced the so-called autoregressive conditional heteroskedastic (ARCH) class of models. In the ARCH model the conditional variance is allowed to change through time as a function of current and past information. Although this new class of time series models allows for a much wider class of nonlinear dynamic econometric models, the linear $\text{ARCH}(q)$ and the generalized ARCH model in Bollerslev (1986), $\text{GARCH}(p, q)$, have been found to be particularly useful parameterizations in the modeling of monetary and financial data.

A common finding in most of these studies concerns the presence of an approximate unit root in the estimated autoregressive polynomial for the conditional second moments, as in the integrated GARCH, or IGARCH, class of models proposed in Engle and Bollerslev (1986). In the IGARCH models shocks to the variance are persistent in the sense that current information remains important for forecasts of all horizons.

Although many economic or financial time series may exhibit persistence in their conditional variances, as previously noted by Engle (1987a, 1987b), a nontrivial linear combination of such variables may have no persistence in variance. In that situation the variables are naturally defined to be co-persistent in variance, and the co-persistent linear combination may be interpreted as a long-run relationship. This idea provides a generalization of linear co-integration in the mean, as in Granger (1981) and Engle and Granger (1987), to linear co-persistence in the variance. From an economic perspective the presence of co-persistence is potentially very important in the understanding of many modern asset pricing relationships. The price of an asset typically depends on the conditional covariance with some benchmark portfolio, so that the pricing of long-term contracts may differ radically from that of one-period contracts if shocks to the conditional variance or covariance with the benchmark portfolio persist. At the same time, the pricing of certain linear combinations of the assets, or portfolios, may not be very sensitive to current information if the assets involved show co-persistence in variance.

The plan of the rest of the paper is as follows. The notation is set out in the next section, which also briefly motivates the ideas within the context of a simple asset pricing paradigm. The statistical concept of persistence in variance is formally defined, and discussed in some detail as it relates to the linear $\text{GARCH}(p, q)$ class of models in Section 3. Building on these results, Section 4 introduces the concept of co-persistence in variance, together with a set of necessary and sufficient conditions for co-persistence to occur in the linear $\text{GARCH}(p, q)$ model. The empirical relevance of the ideas is illustrated in Section 5 by a simple example for the nominal U.S. dollar exchange rate versus the deutschemark and the British pound. Whereas the estimated bivariate system appears persistent in variance, the conditional variances of the two exchange rates are tied together in the long run. In particular, the bilateral deutschemark/British pound rate shows no persistence in variance, suggesting

that most of the volatility in the nominal U.S. dollar rates may be attributable to dollar-related news. Section 6 concludes.

2. NOTATION AND ECONOMIC MOTIVATION

Let $\{y_t\}$ denote the $N \times 1$ discrete time vector stochastic process of interest, with conditional mean and variance functions:

$$(1) \qquad E_{t-1}(y_t) = M_t$$
$$(t = 0, 1, \ldots),$$
$$(2) \qquad \mathrm{var}_{t-1}(y_t) = H_t.$$

The content of the information set used in forming the conditional moments will depend upon the particular application, though often attention will be restricted to the past of the process $\{y_t\}$ only. M_t is an $N \times 1$ stochastic vector, and the stochastic $N \times N$ symmetric covariance matrix H_t is almost surely positive definite for all t. Also, let the $N \times 1$ innovation vector or shocks to the conditional mean be denoted by

$$(3) \qquad \varepsilon_t \equiv y_t - M_t \qquad\qquad (t = 0, 1, \ldots).$$

For notational convenience, the dependence on the parameters of the process has been suppressed. Of course, if H_t is changing through time, other conditional moments of the process will likely be time varying also. From an economic or financial theory point of view, however, the interest typically centers around the first two moments of the process.

Before proceeding with the technical details defining persistence and co-persistence in variance, it is helpful to motivate the issues in a financial markets context. Suppose that the $N \times 1$ vector of asset excess returns, y_t, satisfies a strict factor structure,

$$\varepsilon_t = \sum_{k=1}^{K} \theta_k \eta_{kt} + \nu_t \qquad\qquad (t = 0, 1, \ldots,),$$

where

$$E_{t-1}(\eta_{kt}) = E_{t-1}(\eta_{kt}\eta_{jt}) = 0$$
$$(k, j = 1, \ldots, K; \, j \neq k),$$
$$E_{t-1}(\nu_t) = E_{t-1}(\nu_t \eta_{kt}) = 0.$$

The K factors, η_{kt}, are each normalized to have a conditional mean of zero. The K $N \times 1$ vectors of factor loadings are denoted θ_k, while ν_t is an $N \times 1$ stochastic vector of idiosyncratic shocks. The factors can be interpreted as the news sources common to the returns on all of the N assets. Given this return generating structure, the arbitrage pricing theory relates the asset excess returns to the factor loadings, so that the total expected excess returns are simply the composite of the expected returns for each factor; see Ross (1976). More specifically, if the risk premium for factor k at time t is denoted δ_{kt}, and the

idiosyncratic risk is not priced, exact factor pricing predicts that

$$M_t = \sum_{k=1}^{K} \delta_{kt}\theta_k.$$

If a factor is particularly undesirable for an investor, it is easy to form portfolios which have no component of that factor; e.g., $w'\theta_k = 0$. Obviously, there are many such portfolios.

Now suppose some or all of the factors have time varying conditional variances with

$$\text{Var}_{t-1}(\eta_{kt}) = h_{kt} \qquad\qquad\qquad (k = 1, \ldots, K).$$

Assuming that the idiosyncratic shocks are uncorrelated across the N assets with time invariant variances, the full conditional covariance matrix of the returns becomes

$$H_t = \sum_{k=1}^{K} \theta_k\theta_k' h_{kt} + D,$$

where D equals the $N \times N$ diagonal matrix of asset specific variances. Unless there are some redundancies between the factors, every asset which loads onto one of the factors with a time varying variance will have a temporally dependent conditional variance also. A subset of these factors may have variances which are persistent or integrated. As discussed below, these are non-mean reverting variance processes that typically have infinite unconditional variances. If the number of factors with non-mean reverting variances is less than N, then there will always be a portfolio with a nonpersistent variance, however. It is this ability to form portfolios from assets with persistent variances which are then nonpersistent, that we define as co-persistence. In our formal discussion of co-persistence below, the portfolio weights or co-persistent vectors are constrained to be time invariant, but the consideration of time-varying portfolio weights would certainly be possible.

Assuming that the market price of risk is determined by a standard capital asset pricing model with benchmark portfolio $w'_M y_t$, it follows that

$$M_t = \delta\,\text{Cov}_{t-1}(y_t, w'_M y_t) = \delta H_t w_M = \delta \sum_{k=1}^{K} \theta_k(\theta_k' w_M h_{kt}) + \delta D w_M.$$

If the benchmark portfolio is well diversified, the last term is effectively zero, and the risk premium associated with each of the factors is

$$\delta_{kt} = \delta\theta_k' w_M h_{kt} \qquad\qquad\qquad (k = 1, \ldots, K).$$

Even though an agent might choose a portfolio which avoids the persistent factors, that is not an obvious choice. Since the persistent variance factors generally are more variable, the market will reward investors for bearing that risk provided $\theta_k' w_M \neq 0$. The risk premia will be high just when the variance is high. The optimal portfolio allocation will depend upon the preference toward

risk and may differ radically for short and long-term investment horizons in the presence of persistent factor variances.

Given our main focus of the paper, we shall not pursue this discussion any further. It is obvious, however, that to satisfactorily address many of the issues related to long-term portfolio allocation decisions in this and more general situations with time varying conditional return variances and covariances, a formal framework for thinking about and analyzing persistence and co-persistence in variance is called for. We now turn to the development of such a statistical framework.

3. PERSISTENCE IN VARIANCE

In order to consider the persistence of the process $\{y_t\}$, define the $N \times 1$ vector stochastic process

$$(4) \qquad M_t^*(s) \equiv E_s(M_t) - E_0(M_t), \qquad t > s > 0.$$

The persistence of a shock to the conditional mean is then naturally thought of in terms of the limiting distribution of $M_t^*(s)$ for t going to infinity. It is important to recognize that this notion of persistence is a joint property of the underlying stochastic process and the associated information structure. Following standard practice in time series econometrics, we shall assume that the information set used in forming the time s conditional moments is based on an increasing sequence of sigma algebras generated by $\{y_s, y_{s-1}, \ldots\}$, possibly augmented by other random variables known at time s. Shocks to the mean of the process may then be characterized as having only transitory effects if $\limsup_{t \to \infty} |M_t^*(s)| = 0$ almost surely (a.s.) for all $s > 0$. Conversely, if $\limsup_{t \to \infty} |\{M_t^*(s)\}_i| \neq 0$ a.s. for some $s > 0$ and some $i = 1, 2, \ldots, N$, certain shocks will have a permanent effect on the optimal forecast of the process.[2] Of course, the persistence of the process may be further differentiated depending upon whether $\limsup_{t \to \infty} |M_t^*(s)|$ is bounded in probability, or $O_p(1)$. This persistence measure corresponds exactly to the idea of infinite dependence on initial conditions, as discussed within the context of linear models by Beveridge and Nelson (1981).

To illustrate, consider the mltivariate $ARMA(k, l)$ process,

$$M_t = \mu + \Phi(L) y_t + \Theta(L) \varepsilon_t \qquad (t = 0, 1 \ldots),$$

with $\Phi(L)$ and $\Theta(L)$ kth and lth order matrix lag polynomials. Ruling out cancellations, it is well known that the persistence of a shock to the mean depends upon the roots of the characteristic equation $\det[I - \Phi(\lambda^{-1})] = 0$, where $\det[\cdot]$ denotes the determinant operator. If all the roots of this determinant equation lie inside the unit circle, the ARMA process for $\{y_t\}$ may be represented as a time-invariant, possibly infinite-order moving average of the

---

[2] It is important to recognize that $\limsup |\{M_t^*(s)\}_i| = 0$ must hold for all $s > 0$ and $i = 1, 2, \ldots, N$. Consider the univariate process, $y_t = y_{t-2} + \varepsilon_t$ if t is even and $y_t = \varepsilon_t$ if t is odd, where ε_t is i.i.d. It follows directly that $M_t^*(1) = 0$ for all t, while $M_t^*(2) = 0$ for t odd and $M_t^*(2) = y_2 - y_0$ for t even.

martingale difference sequence $\{\varepsilon_t\}$. Thus,

$$\limsup_{t \to \infty} |M_t^*(s)| = \limsup_{t \to \infty} |\Theta_{t-s}\varepsilon_s + \Theta_{t-s+1}\varepsilon_{s-1} + \cdots + \Theta_{t-1}\varepsilon_1|$$
$$= 0 \quad \text{a.s.}$$

for all $s > 0$, where Θ_i denotes the ith order moving average matrix. This definition of transient includes both covariance stationary and strictly stationary ARMA processes with finite first, but possibly infinite second order moments. On the other hand, if $\lambda \geq 1$ solves the characteristic equation, $\limsup_{t \to \infty} |\{M_t^*(s)\}_i| \neq 0$ a.s. for some $s > 0$ and some $i = 1, 2, \ldots, N$. Unit root, or $I(1)$, processes in which the norm of the largest root(s) equals one have recently received a lot of attention in the empirical modeling of macroeconomic and financial time series; for a discussion of the theoretical implications and the empirical evidence see, for instance, Nelson and Plosser (1982), Watson (1986), Baillie and Bollerslev (1989a), and the many references therein.

Whereas there is little ambiguity about what constitutes persistence in linear models, as pointed out by Nelson (1990, 1991), for nonlinear models different measures of convergence may give rise to different notions of persistence. This is particularly true when describing higher order conditional moments.

Following the discussion above, a natural characterization of the persistence in variance of a process is determined by the influence of the initial conditions on the optimal forecasts for the future conditional variances as the forecast horizon increases. This notion of persistence is also the motivation behind the forecast profiles calculated in Gallant, Rossi, and Tauchen (1992). Thus, in analogy to (4) define the $N(N \times 1)/2 \times 1$ vector stochastic process

$$(5) \qquad H_t^*(s) \equiv E_s\big(\text{vech}\,(H_t)\big) - E_0\big(\text{vech}\,(H_t)\big), \qquad t > s > 0,$$

where $\text{vech}(\cdot)$ denotes the vector half operator that stacks the lower triangular elements of an $N \times N$ matrix as an $N(N+1)/2 \times 1$ vector. As for the conditional mean, if $\limsup_{t \to \infty} |H_t^*(s)| = 0$ a.s. for all $s > 0$, the influence of a shock to the conditional variance may be thought of as having only transitory effects. This suggests the following formal definition:

DEFINITION: The stochastic process $\{y_t\}$ is defined to be *persistent in variance* if $\limsup_{t \to \infty} |\{H_t^*(s)\}_i| \neq 0$ a.s. for some $s > 0$ and some $i = 1, 2, \ldots,$ $N(N+1)/2$.

In order to illustrate this definition and the new concepts discussed below, explicit reference will be made to the linear GARCH(p, q) class of models. The same ideas extend directly to other parametric nonlinear time series models, including the exponential GARCH model in Nelson (1991) and the bi-integrated processes in Hansen (1989).

In particular, consider the popular vector GARCH$(1, 1)$ model introduced by Bollerslev, Engle, and Wooldridge (1988) with H_0 positive definite and

$$(6) \qquad \text{vech}\,(H_t) = W + A_1 \,\text{vech}\,(\varepsilon_{t-1}\varepsilon_{t-1}') + B_1 \,\text{vech}\,(H_{t-1}) \quad (t = 1, 2, \ldots).$$

Conditions on A_1 and B_1 for H_t to be positive definite a.s. for all $t \geq 1$ are given in Baba, Engle, Kraft, and Kroner (1990). Let

$$(7) \qquad A_1 + B_1 = Q\Gamma Q^{-1}$$

denote the Jordan decomposition in terms of the $(N(N+1)/2) \times (N(N+1)/2)$ matrices Q and Γ. If the eigenvalues for $A_1 + B_1$ are distinct, Γ equals the diagonal matrix of the $N(N+1)/2$ eigenvalues, and Q is given by the corresponding matrix of right eigenvectors. If some of the eigenvalues coincide, Γ is upper triangular with the eigenvalues along the diagonal, while Q is a nonsingular matrix.

Upon recursive substitution and repeated use of the law of iterated expectations, it follows that the optimal forecast for the conditional covariance matrix may be written as

$$(8) \qquad E_s\big(\mathrm{vech}\,(H_t)\big) = W + (A_1 + B_1)E_s\big(\mathrm{vech}\,(H_{t-1})\big)$$

$$= W + Q\Gamma Q^{-1}E_s\big(\mathrm{vech}\,(H_{t-1})\big)$$

$$= \cdots$$

$$= Q\sum_{i=0}^{t-s-2}\Gamma^i Q^{-1}W + Q\Gamma^{t-s-1}Q^{-1}\,\mathrm{vech}\,(H_{s+1}),$$

$$t - 1 > s > 0,$$

where Γ^0 is equal to the identity matrix by definition. Thus,

$$(9) \qquad H_t^*(s) = Q\Gamma^{t-s-1}Q^{-1}$$

$$\times \left[\mathrm{vech}\,(H_{s+1}) - (A_1 + B_1)^s\,\mathrm{vech}\,(H_1) - \sum_{i=0}^{s-1}(A_1 + B_1)^i W \right].$$

Let λ denote the eigenvalue for $A_1 + B_1$ with the maximum norm. If the eigenvalues are distinct and $|\lambda| < 1$, $\{\Gamma^t\}$ converges to zero element by element. For $|\lambda| = 1$ the norm of the corresponding element in Γ^t equals unity for all t, whereas for $|\lambda| > 1$ at least one of the diagonal elements in Γ^t increases without a bound for t going to infinity. For a formal discussion of the results involving the more general Jordan canonical form in which some of the eigenvalues for $A_1 + B_1$ coincide, see Andersen (1971).

For any arbitrary finite first and second moment initialization of $\mathrm{vech}\,(H_0)$ and ε_0 respectively, it follows now by analogy to the proof for the univariate case in Bollerslev (1986), that the vector GARCH $(1,1)$ process $\{\varepsilon_t\}$ is asymptotically covariance stationary if and only if $|\lambda| < 1$. Also, from (8) the optimal forecast will approach the unconditional covariance matrix of the process

$$\lim_{t \to \infty} E_s\big(\mathrm{vech}\,(H_t)\big) = Q\sum_{i=0}^{\infty}\Gamma^i Q^{-1}W = (I - A_1 - B_1)^{-1}W, \qquad s > 0,$$

and $\limsup_{t \to \infty}|H_t^*(s)| = 0$ a.s. for all $s > 0$.

On the other hand if $|\lambda| = 1$, shocks to the conditional covariance matrix remain important for forecasts of all horizons. In particular, lim sup of $H_t^*(s)$

equals the nonzero random vector,

$$\limsup_{t \to \infty} |H_t^*(s)|$$

$$= \left| Q e_j e_j Q^{-1} \left[\operatorname{vech}(H_{s+1}) - (A_1 + B_1)^s \operatorname{vech}(H_1) \right. \right.$$

$$\left. \left. - \sum_{i=0}^{s-1} (A_1 + B_1)^i W \right] \right|,$$

where e_j refers to the $N(N+1)/2 \times 1$ vector of zeros except for unity in the jth element corresponding to the position of the largest eigenvalue in Γ. Finally, if $|\lambda| > 1$ it follows that the effect of a shock to the covariance matrix will explode over time as $\limsup_{t \to \infty} |H_t^*(s)|$ is unbounded in probability.

These results for the vector GARCH$(1,1)$ process extend straightforwardly to the higher order linear vector GARCH(p,q) class of models,

$$(10) \qquad \operatorname{vech}(H_t) = W + \sum_{i=1}^{q} A_i \operatorname{vech}(\varepsilon_{t-i} \varepsilon_{t-i}') + \sum_{i=1}^{p} B_i \operatorname{vech}(H_{t-i})$$

$$= W + A(L) \operatorname{vech}(\varepsilon_t \varepsilon_t') + B(L) \operatorname{vech}(H_t) \quad (t = 1, 2, \dots),$$

where W is an $N(N \times 1)/2 \times 1$ vector, and the $(N(N+1)/2) \times (N(N+1)/2)$ matrices A_i and B_i are restricted to ensure that H_t is a.s. positive definite for all $t > 0$; see Baba, Engle, Kraft, and Kroner (1990). In particular we have the following theorem.

THEOREM 1: *The vector GARCH(p,q) process $\{\varepsilon_t\}$ defined in (10) is covariance stationary if and only if, all the roots of the characteristic polynomial,*

$$(11) \qquad \det\left[I - A(\lambda^{-1}) - B(\lambda^{-1}) \right] = 0,$$

lie inside the unit circle, in which case $\limsup_{t \to \infty} |H_t^*(s)| = 0$ a.s. for all $s > 0$.

PROOF: See Appendix.

The conditions in terms of the matrix polynomial in (11) for the vector GARCH(p,q) process to be covariance stationary provide a generalization of the results for the univariate case analyzed in Bollerslev (1986). Interestingly, however, in many of the empirical applications with the univariate GARCH(p,q) model to financial or monetary data, the estimated values for $A_1 + \cdots + A_q + B_1 + \cdots + B_p$ are often found to be very close to one; for a recent survey of the extensive literature on ARCH models and their empirical applications in finance see Bollerslev, Chou, and Kroner (1992). This pronounced empirical regularity provides a motivation for the so-called integrated GARCH(p,q), or IGARCH(p,q), class of models introduced by Engle and Bollerslev (1986). In the univariate IGARCH(p,q) model $A_1 + \cdots + A_q + B_1 + \cdots + B_p = 1$, and in conformity with the univariate ARIMA$(k, 1, l)$ class of

models for the conditional mean, shocks to the conditional variance will have a permanent effect as $\limsup_{t \to \infty} |H_t^*(s)| \neq 0$ a.s. for some $s > 0$.[3] From Theorem 1 these results for the univariate IGARCH class of models extends in an obvious way to a multivariate framework with $|\lambda| = 1$.

As it stands, the unrestricted linear vector GARCH(p, q) model in (10) involves a total of $N(N + 1)[1 + N(N + 1)(p + q)/2]/2$ unique parameters, and in practice some simplifying assumptions will have to be imposed. For instance, in the so-called diagonal GARCH(p, q) model in Bollerslev, Engle, and Wooldridge (1988) the A_i and B_i matrices are taken to be diagonal.

In the K-factor GARCH(p, q) parameterization proposed by Engle (1987a) and Engle, Ng, and Rothschild (1990), the conditional covariance matrix H_t is expressed as a linear combination of past values of $\varepsilon_{t-i}\varepsilon'_{t-i}$ and H_{t-i}, each post- and pre-multiplied by the same rank one matrices,

$$(12) \qquad H_t = V + \sum_{i=1}^{q} \sum_{k=1}^{K} a_{ik}^2 g_k f_k' \varepsilon_{t-i} \varepsilon'_{t-i} f_k g_k' + \sum_{i=1}^{p} \sum_{k=1}^{K} b_{ik}^2 g_k f_k' H_{t-i} f_k g_k'$$

$$(t = 1, 2, \ldots).$$

Here a_{ik} and b_{ik} denote scalar constants, and g_k and f_k are $N \times 1$ vectors with the property that $f_k' g_k = 1$, and $f_k' g_j = 0$ for $k \neq j$.

Define the $\mathrm{vec}\,2(\cdot)$ operator for the $N \times 1$ vector γ by $\mathrm{vec}\,2(\gamma) \equiv \mathrm{vech}\,(2\gamma\gamma' - \mathrm{diag}(\gamma)\,\mathrm{diag}(\gamma))$.[4] Then for any symmetric $N \times N$ matrix Ω, $\gamma'\Omega\gamma = \mathrm{vec}\,2(\gamma)'\,\mathrm{vech}\,(\Omega)$. Using this $\mathrm{vec}\,2(\cdot)$ operator, the conditional covariance matrix for the K-factor GARCH(p, q) model in (12) may be written in $\mathrm{vech}(\cdot)$ format as

$$(13) \qquad \mathrm{vech}\,(H_t) = W - \sum_{k=1}^{K} \mathrm{vech}\,(g_k g_k') \omega_k + \sum_{k=1}^{K} \mathrm{vech}\,(g_k g_k') h_{kt}$$

$$(t = 1, 2, \ldots),$$

where $W = \mathrm{vech}\,(V)$ and

$$(14) \qquad h_{kt} \equiv \mathrm{vec}\,2(f_k)' \,\mathrm{vech}\,(H_t)$$

$$= \omega_k + \sum_{i=1}^{q} a_{ik}^2 (f_k' \varepsilon_{t-i})^2 + \sum_{i=1}^{p} b_{ik}^2 h_{kt-i} \qquad (t = 1, 2, \ldots),$$

denotes the conditional variance of $\{f_k' \varepsilon_t\}$, and $\omega_k \equiv \mathrm{vec}\,2(f_k)'W$. The optimal forecast for the full covariance matrix can be computed from the forecasts of

---

[3] This analogy to the linear ARIMA class of models is not complete. Even though the unconditional variance for the IGARCH(p, q) model does not exist, under the additional assumptions of conditionally normal errors the process is strictly stationary and ergodic; see Nelson (1990) and Bougerol and Picard (1992).

[4] To illustrate,

$$\mathrm{vec}\,2((\gamma_1, \gamma_2)') = (\gamma_1^2, 2\gamma_1\gamma_2, \gamma_2^2)', \quad \text{and}$$

$$\mathrm{vec}\,2((\gamma_1, \gamma_2, \gamma_3)') = (\gamma_1^2, 2\gamma_1\gamma_2, 2\gamma_1\gamma_3, \gamma_2^2, 2\gamma_2\gamma_3, \gamma_3^2)'.$$

the conditional variances for the K univariate processes $\{f_k' \varepsilon_t\}$.[5] From Theorem 1, the multivariate K-factor GARCH(p, q) process is covariance stationary if and only if all of the univariate processes, $\{f_k' \varepsilon_t\}$ are covariance stationary, i.e., $a_{1k}^2 + \cdots + a_{qk}^2 + b_{1k}^2 + \cdots + b_{pk}^2 < 1$, in which case $\limsup_{t \to \infty} |h_{kt}^*(s)| \neq 0$ a.s. for all $s > 0$ and $k = 1, \ldots, K$. A similar result holds true for the general multivariate factor ARCH model, in which the h_{kt}'s need not correspond to the variance of a particular linear combination of the process.

4. CO-PERSISTENCE IN VARIANCE

Granger (1981) first proposed the concept of co-integration, which has been further developed by Engle and Granger (1987) and Granger (1986) among others. The basic idea is that two or more time series may show persistence in the mean as discussed above, while at the same time certain linear combinations of the variables may exhibit no persistence, or be covariance stationary. In this case the variables are said to be co-integrated. This implies that long-term forecast of the individual series may depend nontrivially on shocks at the forecast origin, but for the co-integrating linear combination of the variables, the influence of a shock at the forecast origin will vanish asymptotically as the forecast horizon increases. In particular, if $\{y_t\}$ is an $N \times 1$ vector $I(1)$ process with conditional mean M_t, the effect of a shock to the mean is persistent as $\limsup_{t \to \infty} |M_t^*(s)| \neq 0$ a.s. for some $s > 0$, whereas for the nonzero co-integrating vector θ,

$$\limsup_{t \to \infty} |E_s(\theta' M_t) - E_0(\theta' M_t)| = \limsup_{t \to \infty} |\theta' M_t^*(s)| = 0$$

a.s. for all $s > 0$. If $\{y_t\}$ has finite second order moments, this particular linear combination defines a covariance stationary long-run equilibrium relationship among the variables. Although the forecasts of the individual elements in $\{y_t\}$ diverge, the multivariate long-run forecasts are tied together.

Similarly, while many financial or monetary time series appear to be persistent in variance, certain linear combinations of the variables may have no persistence, so that the variables exhibit a common long-run component. The following definition formalizes this idea.

DEFINITION: The multivariate stochastic process $\{y_t\}$ is defined to be *co-persistent in variance* if there exist a vector $\gamma \in \mathscr{R}^N$ such that $\{\text{vec} 2(\gamma)\}_i \neq 0$ and $\limsup_{t \to \infty} |\{H_t^*(s)\}_i| \neq 0$ a.s. for some $s > 0$ and some $i = 1, 2, \ldots, N(N + 1)/2$, while

$$\limsup_{t \to \infty} |E_s(\gamma' H_t \gamma) - E_0(\gamma' H_t \gamma)| = \limsup_{t \to \infty} |\text{vec} 2(\gamma)' H_t^*(s)| = 0$$

a.s. for all $s > 0$.

---

[5] In the asset pricing applications in Engle, Ng, and Rothschild (1990) and Ng, Engle, and Rothschild (1992), the portfolios defined by $f_k/(f_k' f_k)$ are referred to as factor representing portfolios. In the K-factor GARCH(p, q) model the conditional covariance between any two factor representing portfolios is constant. Also, any portfolio w such that $w' g_k = 0$ for $k = 1, \ldots, K$ will have a constant variance.

This definition is analogous to the idea of co-integration in the mean. While shocks persist in the optimal forecasts for some of the individual variances and covariances of the process, the variances and covariances are tied together in the long-run, and for the particular univariate stochastic process defined by the linear combination $\{\gamma'y_t\}$, shocks to the conditional variance have only temporary effects. Below we shall refer to γ as the co-persistent vector. Of course, γ is only unique up to scale, and some normalization scheme will have to be imposed. Also, the number of such normalized linearly independent co-persistent vectors might exceed one.

In analogy to co-integration in the mean, the above definition of co-persistence in variance is restricted to linear combinations of the variables having no persistence. For $\{y_t\}$ a vector of asset returns this corresponds exactly to the portfolio with weights $\gamma/(\gamma'\gamma)$ having no persistence. While it is possible that $\limsup_{t\to\infty}|\theta'H_t^*(s)| \neq 0$ for some vector $\theta \in \mathcal{R}^{N(N+1)/2}$ not of the form $\mathrm{vec}\,2(\gamma)$, the interpretation of such relationships are difficult. The concept of co-persistence in variances suggested here could also be extended to cover common persistence across both means and variances, by considering situations where $\limsup_{t\to\infty}|(M_t^*(s)', H_t^*(s)')'| \neq 0$, but $\limsup_{t\to\infty}|\theta'(M_t^*(s)', H_t^*(s)')'| = 0$ for some $\theta \in \mathcal{R}^{N+N(N+1)/2}$.[6] Of course, this would subsume the above definition of co-persistence in variance for $\theta' = (0', \mathrm{vec}\,2(\gamma)')$, as well as common co-persistence in the mean and variance for $\theta' = (\gamma', \mathrm{vec}\,2(\gamma)')$. More generally, a variable persistent in the mean might enter the conditional variance of some other variable and $\limsup_{t\to\infty}|\theta'(M_t^*(s)', H_t^*(s)')'| = 0$ without a particular linear combination of the variables eliminating this persistence.[7] We leave further work exploring these more general notions of co-persistence for future work.

Returning to the definition of co-persistence in variance set out above, the univariate IGARCH class of models introduced in Engle and Bollerslev (1986), and their multivariate extensions discussed in Section 3, can now be formally analyzed in a multivariate context allowing for the possibility of common long-run dynamics in second order moments.

THEOREM 2: *Let* $|\lambda_1| \geq \cdots \geq |\lambda_r| \geq 1 > |\lambda_{r+1}| \geq \cdots \geq |\lambda_n|$ *denote the ordered roots from the characteristic polynomial for the vector* $GARCH(p,q)$ *process* $\{\varepsilon_t\}$ *in* (10), *and* $v_1, v_2, \ldots, v_n$ *the corresponding* $N(N+1)/2 \times 1$ *right eigenvectors,*

$$(15) \qquad A(\lambda_i^{-1})v_i + B(\lambda_i^{-1})v_i = v_i.$$

The process is then co-persistent in variance, if and only if

$$(16) \qquad \mathrm{vec}\,2(\gamma)'v_i = 0 \qquad\qquad\qquad (i = 1, 2, \ldots r),$$

for some nonzero vector $\gamma \in \mathcal{R}^N$.

[6] We are grateful to a referee for pointing out this possible extension.
[7] An interesting empirical illustration of such co-persistence across means and variances is provided by Glosten, Jagannathan, and Runkle (1990) where the short term interest rate is found to enter significantly in the conditional variance equation for stock returns. Similarly, Lamoureux and Lastrapes (1990a) find that contemporaneous trading volume helps explain stock return volatility.

PROOF: See Appendix.

From the proof of Theorem 2 the univariate stochastic process $\{\gamma'\varepsilon_t\}$ is covariance stationary, although in general the conditional variance for $\{\gamma'\varepsilon_t\}$ will depend upon past realizations of all the elements in the $N \times 1$ vector stochastic process $\{\varepsilon_t\}$, as well as past elements in the $N \times N$ conditional covariance matrices, $\{H_t\}$. However, in certain situations $\{\gamma'\varepsilon_t\}$ reduces to a simple univariate GARCH(p,q) process.

LEMMA 1: *Linear combinations, $\{\gamma'\varepsilon_t\}$, of the vector GARCH(p,q) process in (10) will follow a univariate GARCH(p,q) process if and only if for some scalar constants $\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p$,*

$$(17) \qquad \text{vec}\,2(\gamma)'A_i = \alpha_i\,\text{vec}\,2(\gamma)' \qquad\qquad (i = 1, 2, \ldots, q),$$

$$\text{vec}\,2(\gamma)'B_i = \beta_i\,\text{vec}2(\gamma)' \qquad\qquad (i = 1, 2, \ldots, p).$$

PROOF: See Appendix.

It follows immediately, that if the vector GARCH(p,q) process is co-persistent in variance, and $\{\gamma'\varepsilon_t\}$ has a univariate GARCH(p,q) representation, where γ is a co-persistent vector, the sum of the scalar parameters $\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p$ must be less than one. In particular, for the K-factor GARCH(p,q) process, we have the following lemma.

LEMMA 2: *The K-factor GARCH(p,q) process $\{\varepsilon_t\}$ in (12) with*

$$\sum_{i=1}^{q} a_{ik}^2 + \sum_{i=1}^{p} b_{ik}^2 \geq 1 \qquad\qquad (k = 1, \ldots, r)$$

and

$$\sum_{i=1}^{q} a_{ik}^2 + \sum_{i=1}^{p} b_{ik}^2 < 1 \qquad\qquad (k = r+1, \ldots, K)$$

is co-persistent in variance, and $\{f_k'\varepsilon_t\}$, $k = 1, 2, \ldots, K$, have univariate GARCH(p,q) representations with parameters $a_{1k}^2, \ldots, a_{qk}^2, b_{1k}^2, \ldots, b_{pk}^2$.

PROOF: Since $\text{vec}\,2(f_k)'\text{vech}(g_j g_j') = 0$ for $k \neq j$, the lemma follows directly from post-multiplication in (13) with $\text{vec}\,2(f_k)'$ for $k = r+1, \ldots, K$. Q.E.D.

Note, for the 1-factor GARCH(p,q) model, any vector γ such that

$$\text{vec}\,2(\gamma)'\text{vech}(g_1 g_1') = (\gamma'g_1)^2 = 0$$

will result in a conditionally homoskedastic process, $\{\gamma'\varepsilon_t\}$.

5. EMPIRICAL EXAMPLE

One of the areas where the ARCH methodology has found the widest use has been in the modeling of high frequency foreign exchange rate data. As an illustration of the ideas introduced above, we shall here consider a simple example based on daily data for the deutschemark (DM) and the British pound (BP) exchange rate vis a vis the U.S. dollar. Both the DM and the BP are quoted as the number of U.S. dollars per foreign currency. The data cover the period January, 1980 until February, 1985, for a total of 1245 observations excluding vacations.[8] For a more detailed description of the data see Baillie and Bollerslev (1989a, 1989b) and Bollerslev (1987), where the same set of data has been analyzed from a different perspective.

Since the appropriate procedures for conducting inference in multivariate persistent variance processes have not yet been fully worked out, the empirical results discussed below should be interpreted with some caution. As noted in Engle and Bollerslev (1986), it is possible that some of the well known difficulties for conducting inference in $I(1)$ processes carry over to models with persistence in variance also. However, the quasi maximum likelihood estimates (QMLE), obtained under the assumption of conditional normality, essentially do a GLS type correction which might therefore ensure standard asymptotic properties of the estimators and associated test statistics; see Hong (1988), Lee and Hansen (1991), and Lumsdaine (1990) for a formal treatment of the univariate IGARCH$(1, 1)$ model.

Turning to the data, for neither of the two time series is the null hypothesis of a unit root in the autoregressive polynomial in the univariate time series representation for the mean rejected at conventional levels;[9] see Kim and Schmidt (1992) for a discussion of testing for unit roots in the mean in the presence of IGARCH type effects. After first differencing, none of the resulting return series show any remaining serial correlation. Testing for co-integration in the mean between the two rates as in Engle and Granger (1987) also fails to reject the null hypothesis of no co-integration. These preliminary tests therefore suggest the simultaneous modeling of the mean adjusted logarithmic first differences of the two spot rates as a simple bivariate martingale sequence. However, from the univariate analysis in Baillie and Bollerslev (1989b) and Bollerslev (1987), both the DM and the BP exhibit substantial time variation in their conditional variances. For instance, Bollerslev (1987) on estimating a univariate GARCH$(1, 1)$ model for each of the two rates finds $\hat{\alpha}_1 + \hat{\beta}_1$ equal to .966 and .971 respectively, and for neither the DM nor the BP is the null hypothesis of an IGARCH$(1, 1)$ process, i.e., $\alpha_1 + \beta_1 = 1$, rejected by a conventional t test.

---

[8] There is a three year overlap with the sample period used here and the earlier sample for the BP studied by Gallant, Hsieh, and Tauchen (1991). However, as the analysis in Baillie and Bollerslev (1989b) and Bollerslev (1987) indicate, the BP was merely "recalcitrant" during the 1970's.

[9] A more detailed description of the preliminary test results and the model diagnostics discussed below is available in the earlier working paper version of this article.

Of course, this apparent persistence in the volatility process may very well be common to the two series. To further investigate this idea, the conditional covariance matrix for the continuously compounded percentage daily rate of returns, $y_t \equiv (100 \cdot \Delta \log(s_t^{DM}), 100 \cdot \Delta \log(s_t^{BP}))$, is here parameterized as a bivariate GARCH$(1, 1)$ model. In order to ensure that for some nonzero vector $\gamma \in \mathscr{R}^2$, the conditional variance for the scalar stochastic process $\{\gamma' y_t\}$ takes the form of a simple univariate GARCH$(1, 1)$ model, the conditions given in Lemma 1 were imposed in the estimation. Compared to the unrestricted GARCH$(1, 1)$ model in (6), the six nonlinear restrictions, $\mathrm{vec}\, 2(\gamma)' A_1 = \alpha_1 \mathrm{vec}\, 2(\gamma)$ and $\mathrm{vec}\, 2(\gamma)' B_1 = \beta_1 \mathrm{vec}\, 2(\gamma)$ for some $\gamma \in \mathscr{R}^2$ and $\alpha_q, \beta_1 \in \mathscr{R}$, result in three over-identifying restrictions, leaving 15 as opposed to 18 free parameters in A_1 and B_1 to be estimated. Normalizing the first element in γ to unity, the QMLE of the model parameters, with asymptotic standard errors in parentheses, are[10]

$$y_t = \begin{bmatrix} -.048 \\ (.018) \\ -.048 \\ (.017) \end{bmatrix} + \varepsilon_t,$$

$$(18) \quad \mathrm{vech}\,(H_{t+1}) = \begin{bmatrix} .009 \\ (.004) \\ -.009 \\ (.004) \\ .015 \\ (.005) \end{bmatrix} + \begin{bmatrix} .068 & .000 & .024 \\ (-) & (-) & (.009) \\ -.048 & .161 & -.059 \\ (.013) & (-) & (.016) \\ .048 & -.146 & .130 \\ (.012) & (.035) & (-) \end{bmatrix} \cdot \mathrm{vech}\,(\varepsilon_t \varepsilon_t')$$

$$+ \begin{bmatrix} .894 & .000 & .000 \\ (-) & (-) & (-) \\ .123 & .709 & .092 \\ (.018) & (-) & (.027) \\ .000 & .176 & .762 \\ (-) & (.042) & (-) \end{bmatrix} \cdot \mathrm{vech}\,(H_t),$$

$$\hat{\gamma} = \begin{bmatrix} 1.000 & -1.176 \\ (-) & (.058) \end{bmatrix}', \quad \hat{\alpha}_1 = .247, \quad \hat{\beta}_1 = .605.$$
$$\qquad\qquad\qquad\qquad\qquad (.026) \qquad\quad (.034)$$

After some preliminary analysis, the nonbinding restrictions $\{B_1\}_{13} = \{B_1\}_{31} = 0$ were pre-imposed in the estimation of the model in order to guarantee a well defined conditional variance process. Also, in the preliminary unrestricted estimation the parameters for the covariance terms entering the conditional variance equation for the DM, i.e., $\{A_1\}_{12}$ and $\{B_1\}_{12}$, were both found to be small and insignificant, and were subsequently fixed at zero to ease convergence of the numerical optimization algorithm. Indeed, a $T \cdot R^2$ type LM test for the

---

[10] Note, it is not possible to write the model in (18) in the factor GARCH form in equation (12).

hypothesis that $\{A_1\}_{12} = \{B_1\}_{12} = 0$ takes the value 3.832, which is insignificant at conventional levels in the asymptotic chi-squared distribution with two degrees of freedom.[11] Similarly, the LM test for the three overidentifying restrictions from Lemma 1 imposed in the estimation of (18) equals only 2.488.

To further check the validity of the model specification, a series of additional diagnostic tests were calculated. For instance, the LM test statistic for an unrestricted $AR(1)$, or equivalently $MA(1)$, error structure is 4.926, far below the five percent critical value in the chi-squared distribution with four degrees of freedom. The LM test for additional serial correlation in the second order moments in the form of a diagonal $ARCH(2)$ term, i.e., $\{A_2\}_{11} = \{A_2\}_{22} = \{A_2\}_{33} = 0$, equals 3.458, again insignificant at conventional levels. Also, the Pagan and Sabau (1987) consistency tests based on the regressions of $\{\hat{\varepsilon}_t\}_i^2$ on a constant and $\{\hat{H}_t\}_{ii}$ result in t statistics for the slope coefficients equal to one of 1.436 and 1.447, respectively. Thus, subject to the caveats noted earlier, the bivariate model in (18) seems to provide a reasonably good description of the joint conditional first and second order moments of the two exchange rates.

Turning to the actual estimation results, the ordered eigenvalues for the characteristic polynomial in (11), i.e., $\det[I - \hat{A}_1\lambda^{-1} - \hat{B}_1\lambda^{-1}] = 0$, equal $\hat{\lambda}_1 = .982$, $\hat{\lambda}_2 = .889$, and $\hat{\lambda}_3 = .852$ respectively, with corresponding eigenvectors $\hat{v}_1 = (.629, .575, 523)$, $\hat{v}_2 = (-.283, .394, .874)$, and $\hat{v}_3 = (-.150, -.697, .701)$. In light of the standard errors for the other parameter estimates, and the downward bias in the univariate estimates for $\alpha_1 + \beta_1$ in the $IGARCH(1,1)$ model documented in Lumsdaine (1991), this implied estimate for $|\lambda_1|$ is therefore suggestive of a high degree of persistence in the variance for the bivariate model; see Theorem 1.[12] However, at the same time $\text{vec}\, 2(\hat{\gamma})'\hat{v}_1 = \text{vec}\, 2(\hat{\gamma})'\hat{v}_2 = 0$, and by Theorem 2, the estimation results point to $\hat{\gamma}$ as a co-persistent vector for the system.

As noted by a referee, imposing the constraints in Lemma 1 in the estimation implies that $\text{vec}\, 2(\hat{\gamma})'(\hat{A}_1 + \hat{B}_1) = (\hat{\alpha}_1 + \hat{\beta}_1)\text{vec}\, 2(\hat{\gamma})'$, and therefore makes $\text{vec}\, 2(\hat{\gamma})$ a left eigenvector for $\hat{A}_1 + \hat{B}_1$. Since the matrix of left eigenvectors is equal to the inverse of the matrix of right eigenvectors, it follows by construction that $\text{vec}\, 2(\hat{\gamma})$ must be orthogonal to all but one of the estimated right eigenvectors for $\hat{A}_1 + \hat{B}_1$. However, nothing in the estimation imposes that $\text{vec}\, 2(\hat{\gamma})$ be orthogonal to the right eigenvectors corresponding to the two largest eigenvalues; i.e. $\hat{v}_1$ and $\hat{v}_2$.

Thus, whereas the effect of a shock in terms of the optimal forecasts for each of the conditional variances and the covariance will persist, or die out very

---

[11] With conditionally leptokurtic errors, the actual size of the OPG $T \cdot R^2$ LM test exceeds the nominal size, yielding even stronger support for the null hypothesis in the absence of a rejection at a particular level; see Bollerslev and Wooldridge (1992) for related Monte Carlo evidence.

[12] Of course, when comparing persistence across different models using varying data frequencies it is misleading to simply focus on the numerical value of the largest estimated eigenvalue, $|\hat{\lambda}_1|$, unless $|\hat{\lambda}_1| = 1$. In particular, to gain some intuition about the degree of persistence implied by the largest root, it is useful to think about the half life of a shock associated with that root; i.e., the number k such that $\hat{\lambda}_1^k = .5$. For $\hat{\lambda}_1 = .982$ the half life is about 38 trading days. The half life associated with $\hat{\lambda}_3 = .852$ is only about four trading days.

slowly according to the largest characteristic root $|\hat{\lambda}_1| = .982$, the long-run forecasts are tied together.[13] In fact, the particular linear combination of the two exchange rates defined by the co-persistent vector $\hat{\gamma}$, i.e., $\{\hat{\gamma}'y_t\}$, follows a simple univariate GARCH$(1,1)$ model with $\hat{\alpha}_1 = .247$ and $\hat{\beta}_1 = .605$.

Interestingly, the QMLE for γ is close to $(1, -1)$. By the absence of triangular arbitrage this particular linear combination of the two rates corresponds exactly to the logarithmic first difference of the deutschemark/British pound exchange rate.[14] Indeed, the t statistic for $\alpha_1 + \beta_1 = 1$ in the univariate GARCH$(1,1)$ model for the deutschemark/British pound rate equals 6.033, overwhelmingly rejecting the null hypothesis of a persistent variance process for the bilateral European rate. This suggest that most of the apparent persistence in the volatility in the U.S. foreign exchange market may be attributed to dollar-related news.

6. CONCLUSION

In this paper a definition of co-persistence in variance was proposed. The relevance of this new idea was illustrated by an empirical example for two nominal U.S. dollar exchange rates. Similar findings are likely to obtain with different foreign currencies and in the study of other related speculative prices. This has important implications for many interesting questions in theoretical and empirical finance and monetary economics.

For instance, the finding of a common long-run component in the volatility processes may prove helpful in the construction of long-term forecast intervals and in the calculation of optimal hedge portfolios. According to modern asset pricing theories and in the pricing of derivative assets, the price today is a function of the conditional variance of the future asset returns, or the covariance with some benchmark portfolio(s). If shocks to the conditional variance or covariance have only short lived temporary effects, the risk premium associated with long term contracts will be affected very little. On the other hand, if the conditional variance is persistent, the pricing of long term contracts will be a nontrivial function of today's information set. At the same time, if the assets in a portfolio are co-persistent in variance with the co-persistent vector proportional to the vector of asset shares, the risk premium for long term contracts in that portfolio will tend to be time invariant. An empirical illustration is provided by the factor GARCH model for short term Treasury bills in Engle, Ng, and Rothschild (1990), where the estimates are suggestive of co-persistence in

---

[13] As stated by Nelson (1991), "the usual cautions about interpreting a unit root as evidence of truly infinite persistence of shocks apply." Any parametric time series model imposes restrictions on the relation between the low and high frequencies, so that in practice much of the information about the low frequency behavior may implicitly be derived from the behavior at high frequencies. Of course, changes in the underlying structure may also result in apparent persistence; see Lamoureux and Lastrapes (1990b) and Schwert (1987).

[14] Fixing $\gamma \equiv (1, -1)$, the QMLE for α_1 and β_1 take the values .256 (.028) and .583 (.040) respectively, with corresponding eigenvalues .981, .879, and .839. The quasi likelihood ratio test statistic for $\gamma = (1, -1)$ equals 4.603.

variance across the different maturities. Similarly in the factor GARCH model for stock returns estimated by Ng, Engle, and Rothschild (1992), the size sorted portfolios appear to be co-persistent in variance. We leave further theoretical and empirical work exploring these ideas for future research.

Dept. of Finance, J. L. Kellogg Grad. School, Northwestern University, Evanston, Ill. 60208, U.S.A.

and

Dept. of Economics, D-008, University of California, San Diego, LaJolla, CA 92093, U.S.A.

APPENDIX

1. PROOF OF THEOREM 1: From the definition of the process,

$$E_s\left(\text{vech}\left(H_t\right)\right) = W + \sum_{i=1}^{q} A_i E_s\left(\text{vech}\left(H_{t-i}\right)\right) + \sum_{i=1}^{p} B_i E_s\left(\text{vech}\left(H_{t-i}\right)\right),$$

where $t - s > m \equiv \max\{p,q\}$. Stacking $\text{vech}(H_{t+s}),\ldots,\text{vech}(H_{t+s-m+1})$ as an $n \times 1$ vector, $\text{vech}_n(H_t)$, where $n \equiv mN(N+1)/2$, this may be written more conveniently in companion first order form as

(A1) $E_s\left(\text{vech}_n\left(H_t\right)\right) = F_n' W + C_n E_s\left(\text{vech}_n\left(H_{t-1}\right)\right),$

where the $N(N+1)/2 \times n$ selection matrix F_n has zeros everywhere except for unity in the (i,i)th elements $i = 1,2,\ldots,N(N+1)/2$, and the $n \times n$ companion matrix C_n is defined by

(A2) $C_n = \begin{bmatrix} A_1+B_1 & A_2+B_2 & \cdots & A_{m-1}+B_{m-1} & A_m+B_m \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix},$

with $A_i \equiv 0$ for $i > q$, and $B_i \equiv 0$ for $i > p$. It follows therefore from the analogy to the GARCH(1,1) model in (8) that the vector GARCH(p,q) process is covariance stationary if and only if all the roots of

(A3) $\det[\lambda I - C_n] = 0$

lie inside the unit circle. Let $\lambda_i \neq 0$ denote a characteristic root with corresponding $n \times 1$ characteristic vector x_i, i.e.,

$$C_n x_i = \lambda_i x_i.$$

From the structure of C_n, $x_i' = (v_i', \lambda_i^{-1} v_i', \ldots, \lambda_i^{-m+1} v_i')$, where the $N(N+1)/2 \times 1$ vector v_i satisfies

(A4) $(A_1+B_1)v_i + (A_2+B_2)\lambda_i^{-1}v_i + \cdots + (A_m+B_m)\lambda_i^{-m+1} = \lambda_i v_i.$

Dividing through by λ_i it follows that $\det[I - A(\lambda_i^{-1}) - B(\lambda_i^{-1})] = 0$.   Q.E.D.

2. PROOF OF THEOREM 2: From (8) and the proof of Theorem 1 it follows that

$$(A5) \qquad E_s(\text{vech}_n(H_t)) = \sum_{i=0}^{t-s-m-1} Q\Gamma^i Q^{-1} F_n' W + Q\Gamma^{t-s-m} Q^{-1} E_s(\text{vech}_n(H_{m+s})),$$

where $t - s > m$, Γ refer to the $n \times n$ matrix of eigenvalues from (A3), and Q the corresponding matrix of right eigenvectors. Note, the first $N(N+1)/2$ elements in the jth column of Q is proportional to v_j, say $\delta_j v_j$. If $\text{vec}\,2(\gamma)'v_j = 0$ for $j = 1, 2, \ldots, r$, it follows that

$$(A6) \qquad E_s(\gamma' H_t \gamma) = E_s(\text{vec}\,2(\gamma)' \text{vech}(H_t))$$

$$= E_s(\text{vec}\,2(\gamma)' F_n \text{vech}_n(H_t))$$

$$= \sum_{i=0}^{t-s-m-1} (0, \ldots, 0, \delta_{r+1} \text{vec}\,2(\gamma)'v_{r+1}\lambda_{r+1}^i, \ldots) Q^{-1} F_n' W$$

$$+ (0, \ldots, 0, \delta_{r+1} \text{vec}\,2(\gamma)'v_{r+1}\lambda_{r+1}^{t-s-m}, \ldots) Q^{-1} E_s(\text{vech}_n(H_{m+s})).$$

Thus,

$$\limsup_{t \to \infty} |E_s(\gamma' H_t \gamma) - E_0(\gamma' H_t \gamma)| = \limsup_{t \to \infty} |\text{vec}\,2(\gamma)' H_t^*(s)| = 0$$

a.s. for all $s > 0$. Conversely, suppose the model is co-persistent in variance. By direct substitution,

$$(A7) \qquad \text{vec}\,2(\gamma)' H_t^*(1) = \text{vec}\,2(\gamma)'Q\Gamma^t$$

$$\cdot \left[\Gamma^{-m-1} Q^{-1} (E_1(\text{vech}_n(H_{m+1})) - F_n'W) \right.$$

$$\left. - \Gamma^{-m} Q^{-1} E_0(\text{vech}_n(H_m)) \right]$$

$$= (\delta_1 \text{vec}\,2(\gamma)'v_1\lambda_1^t, \ldots, \delta_i \text{vec}\,2(\gamma)'v_i\lambda_i^t, \ldots)$$

$$\cdot \left[\Gamma^{-m-1} Q^{-1} (E_1(\text{vech}_n(H_{m+1})) - F_n'W) \right.$$

$$\left. - \Gamma^{-m} Q^{-1} E_0(\text{vech}_n(H_m)) \right],$$

and since $\limsup_{t \to \infty} |\text{vec}\,2(\gamma)' H_t^*(1)| = 0$ a.s. it follows that

$$\text{vec}\,2(\gamma)'v_i = 0$$

for all i corresponding to $|\lambda_i| \geq 1$.   Q.E.D.

3. PROOF OF LEMMA 1: By direct substitution,

$$E_{t-1}((\gamma'\varepsilon_t)^2) = \text{vec}\,2(\gamma)' \text{vech}(H_t)$$

$$= \text{vec}\,2(\gamma)'W + \sum_{i=1}^{q} \text{vec}\,2(\gamma)' A_i \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}')$$

$$+ \sum_{i=1}^{p} \text{vec}\,2(\gamma)' B_i \text{vech}(H_{t-i})$$

$$= \omega + \sum_{i=1}^{q} \alpha_i(\gamma'\varepsilon_{t-i})^2 + \sum_{i=1}^{p} \beta_i(\gamma'H_{t-i}\gamma),$$

where the last equality follows from (17). Conversely, suppose $\{\gamma'\varepsilon_t\}$ reduces to a univariate GARCH(p,q) process with conditional variance $\{\gamma'H_t\gamma\}$ and parameters $\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p$. Then for all realizations of $\{\varepsilon_t\}$ and $\{H_t\}$,

$$\text{vec}\,2(\gamma)' A_i \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}') = \alpha_i(\gamma'\varepsilon_{t-i})^2 \qquad\qquad (i = 1, 2, \ldots, q),$$

and

$$\text{vec}\,2(\gamma)' B_i \text{vech}(H_{t-i}) = \beta_i(\gamma'H_{t-i}\gamma) \qquad\qquad (i = 1, 2, \ldots, p).$$

However,

$$\alpha_i(\gamma'\varepsilon_{t-i})^2 = \alpha_i \operatorname{vec} 2(\gamma)' \operatorname{vech}(\varepsilon_{t-i}\varepsilon'_{t-i}) \qquad (i = 1, 2, \ldots, q),$$

and

$$\beta_i(\gamma'H_{t-1}\gamma) = \beta_i \operatorname{vec} 2(\gamma)' \operatorname{vech}(H_{t-i}) \qquad (i = 1, 2, \ldots, p),$$

and (17) follows immediately. Q.E.D.

REFERENCES

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series*. New York: John Wiley and Sons.

BABA, Y., R. F. ENGLE, D. KRAFT, AND K. KRONER (1990): "Multivariate Simultaneous Generalized ARCH," UCSD, Department of Economics, unpublished manuscript.

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

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

BEVERIDGE, S., AND C. R. NELSON (1981): "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151–174.

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

———— (1987): "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542–547.

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 Modelling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

BOLLERSLEV, T., AND J. M. WOOLDRIDGE (1992): "Quasi-Maximum Likelihood Estimation of Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143–172.

BOUGEROL, P., AND N. PICARD (1992): "Stationarity of GARCH Processes and of Some Non-Negative Time Series," *Journal of Econometrics*, 52, 115–127.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987–1008.

———— (1987a): "Multivariate GARCH with Factor Structure—Cointegration in Variance," UCSD, Department of Economics, unpublished manuscript.

———— (1987b): "On the Theory of Cointegrated Economic Time Series," UCSD, Department of Economics, unpublished manuscript.

ENGLE, R. F., AND T. BOLLERSLEV (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50; 81–87.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

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

GALLANT, A. R., P. E. ROSSI, AND G. TAUCHEN (1992): "Stock Prices and Volume," *Review of Financial Studies*, 5, 199–242.

GALLANT, A. R., D. HSIEH, AND G. TAUCHEN (1991): "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate 1974–83," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen. Cambridge: Cambridge University Press.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. RUNKLE (1990): "Relationship Between the Expected Value and the Volatility of the Nominal Excess Returns on Stocks," University of Minnesota, Department of Finance, unpublished manuscript.

GRANGER, C. W. J. (1981): "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 23, 121–130.

———— (1986): "Developments in the Study of Cointegrated Economic Variables," *Oxford Bulletin of Economics and Statistics*, 48, 213–228.

HANSEN, B. E. (1989): "Bi-Integration and Heteroskedastic Cointegration," University of Rochester, Department of Economics, unpublished manuscript.

HONG, C. H. (1988): "The Integrated Generalized Autoregressive Conditional Heteroskedastic Model: The Process, Estimation and Monte Carlo Experiment," UCSD, Department of Economics, unpublished manuscript.

KIM, K., AND P. SCHMIDT (1992): "Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics*, forthcoming.

LAMOUREUX, C. G., AND W. D. LASTRAPES (1990a): "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects," *Journal of Finance*, 45, 221–229.

———— (1990b): "Persistence in Variance, Structural Change and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225–234.

LEE, S., AND B. E. HANSEN (1991): "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and the IGARCH Models," University of Rochester, Department of Economics, unpublished manuscript.

LUMSDAINE, R. L. (1990): "Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1, 1) and IGARCH(1, 1) Models," Princeton University, Department of Economics, unpublished manuscript.

———— (1991): "Finite Sample Properties of the Maximum Likelihood Estimator in GARCH(1, 1) Models: A Monte Carlo Investigation," Princeton University, Department of Economics, unpublished manuscript.

NELSON, D. B. (1990): "Stationarity and Persistence in the GARCH(1, 1) Model," *Econometric Theory*, 6, 318–334.

———— (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

NG, V., R. F. ENGLE, AND M. ROTHSCHILD (1992): "A Multi-Dynamic Factor Model for Stock Returns," *Journal of Econometrics*, 52, 245–266.

PAGAN, A. R., AND H. C. L. SABAU (1987): "Consistency Tests for Heteroskedasticity and Risk Models," University of Rochester, Department of Economics, unpublished manuscript.

ROSS, S. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

SCHWERT, G. W. (1987): "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data," *Journal of Monetary Economics*, 20, 73–103.

WATSON, M. W. (1986): "Univariate Detrending Methods with Stochastic Trends," *Journal of Monetary Economics*, 18, 49–75.