THE EFFECT OF LONG MEMORY IN VOLATILITY ON STOCK MARKET FLUCTUATIONS

Bent Jesper Christensen and Morten Ørregaard Nielsen*

Abstract—Recent empirical evidence demonstrates the presence of an important long-memory component in realized asset return volatility. We specify and estimate multivariate models for the joint dynamics of stock returns and volatility that allow for long memory in volatility without imposing this property on returns. Asset pricing theory imposes testable cross-equation restrictions on the system that are not rejected in our preferred specifications, which include a strong financial leverage effect. We show that the impact of volatility shocks on stock prices is small and short lived, in spite of a positive risk-return tradeoff and long memory in volatility.

I. Introduction

The definition, computation, and analysis of realized volatility in financial return series has attracted considerable interest in recent literature, for example, French, Schwert, and Stambaugh (1987), Schwert (1989), Andersen and Bollerslev (1998), and Barndorff-Nielsen and Shephed (2002). Essentially, integrated instantaneous variance is estimated consistently by its sample analog based on high-frequency return observations. This approach allows gathering much more detailed information on the properties of financial market volatility than previously. A striking finding in the recent empirical literature is that realized volatility exhibits long memory. This finding is consistent across a number of studies, and financial theory may accommodate long memory in volatility as well; see, for example, Robinson (1991, 2001), Crato and de Lima (1994), Baillie, Bollerslev, and Mikkelsen (1996), Ding and Granger (1996), Breidt, Crato, and de Lima (1998), Comte and Renault (1998), and Andersen, Bollerslev, Diebold, and Labys (2003) and the references therein. However, so far no study has related the long-memory property of volatility to the level of asset prices themselves. The time series behavior of volatility may be expected to matter for asset prices, since volatility enters the risk premium. In this paper, we establish theoretically and empirically the consequences of long memory in volatility for asset prices.

Received for publication June 24, 2004. Revision accepted for publication May 9, 2006.

* University of Aarhus and CREATES; and Cornell University and CREATES, respectively.

We are grateful to seminar participants at the University of Aarhus and the Multivariate Modelling in Finance and Risk Management 2006 conference at Sandbjerg Manor, as well as three anonymous referees for useful and constructive comments. We are also grateful to Tim Bollerslev for supplying the realized volatility data used in our empirical work, to Rolf Karlsen for research assistance, and to the Danish Social Science Research Council (grant no. FSE 275-03-0220), the Center for Analytical Finance (CAF), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support.

1 Alternative models, such as nonlinear models as in Granger and Ding (1996) and Diebold and Inoue (2001), have been proposed to capture the persistence properties of volatility, but the long-memory model remains the most popular in both theoretical and empirical studies.

An important ingredient in the relation between the level of asset prices and the memory properties of volatility is the fundamental risk-return tradeoff, that is, the contemporaneous relation between the conditional mean and variance of returns, which has been the subject of extensive research. Early theoretical and empirical contributions on the functional form of this relation were due to Merton (1973, 1980). The risk-return tradeoff has been studied in the time-varying volatility case using GARCH-type models by Engle, Lilien, and Robins (1987), Bollerslev, Engle, and Wooldridge (1988), Chou (1988), Campbell and Hentschel (1992), Chou, Engle, and Kane (1992), Backus and Gregory (1993), and others. A two-factor model including a hedging component was considered by Scruggs (1998), who found that the risk-return component indicated a positive tradeoff. Recent work in asset pricing has examined cross-sectional risk premia induced by covariance between innovations in volatility and stock returns. This literature finds negative premia (for example, Ang, Hodrick, Xing, & Zhang, 2006). The idea is that since innovations in volatility are higher during recessions, stocks that covary with volatility are stocks that pay off in bad states, and these stocks should require a smaller risk premium. Thus, there is mixed evidence on the sign of the risk-return relation in the literature. A negative relation is also found in some GARCH-type models, but Harrison and Zhang (1999) use a seminonparametric density approach and show that GARCH models may be misspecified and lead to an erroneous indication of a negative risk-return relation, and they conclude that the true relation is positive. Recently, a positive risk-return tradeoff has been indicated by several studies using efficient high-frequency data volatility measurement, for example Ghysels, Santa-Clar, and Valkanov (2005), who use weighted rolling sample windows in the variance measurements. Similar conclusions are reached by Brandt and Kang (2004) using a latent VAR methodology. For a survey of these and related studies, see Lettau and Ludvigson (forthcoming).

Common to the risk-return tradeoff studies is that they do not incorporate the long-memory characteristic of volatility, which has been established in the realized volatility literature. In the latter, realized volatility is shown to be well described by a fractionally integrated or $I(d)$ process, with long-memory parameter $d$ in the vicinity of $0.3–0.4$ (see Andersen, Bollerslev, Diebold, & Eben, 2001, and Andersen et al., 2003), whereas the risk-return literature takes volatility to be $I(0)$, as in standard GARCH models.

Among the most important uses of models for expected returns is the discounting of streams of expected future cash flows, and thus the calculation of asset prices. While an
important ingredient, the risk-return tradeoff is not the only determinant of asset values. Poterba and Summers (1986) elaborate on the asset valuation aspect and derive the way in which both the risk-return tradeoff and the serial correlation in volatility jointly determine the level of stock prices. Their results show that a stronger positive risk-return relation and higher serial correlation in volatility both contribute toward numerically higher (namely, more negative) elasticity of stock prices with respect to volatility. They argue that effects of shocks to volatility decay rapidly and hence only affect required returns and thereby asset prices over short intervals. Thus, it is difficult to ascribe long-term movements in the stock market to volatility changes.

Obviously, this argument was made before the advent of long-memory models in financial economics and is called into question in view of the amassing recent literature on long memory in financial volatility series. The new evidence calls for a reassessment of the relation between the risk-return tradeoff, serial dependence in volatility, and the level of asset prices. The received evidence on the positive risk-return tradeoff and manifest long memory in volatility seem to suggest possible long-lasting effects of volatility shocks on required returns, and hence on the level of the stock market. A high elasticity (in absolute terms) of stock prices with respect to volatility would be expected.

In the present paper, we examine this possibility. We establish the relation between the risk-return tradeoff, serial dependence in volatility, and the elasticity of asset values with respect to volatility in the presence of long memory in volatility, and we estimate all three empirically. Our volatility measurements are efficient high-frequency realized return volatilities as well as model free implied volatilities from option prices. Our results confirm those from recent literature on both the positive risk-return tradeoff and the long-memory property of volatility. On the other hand, we show that the consequences for asset values are not as anticipated, based on the foregoing interpretation. Thus, the effect of volatility shocks on stock prices is modest and short lived, in our empirically supported model, in spite of the positive risk-return tradeoff and long memory in volatility.

For the joint analysis of asset returns and realized volatility, we introduce a general simultaneous bivariate system of the form

\[ A(L) X_t = B(L) u_t, \]

where \( X_t \) is a vector containing the data on return and volatility in period \( t \), \( A(L) \) and \( B(L) \) are lag polynomials, and \( u_t \) are the system errors. We derive and test parametric restrictions on the general vector autoregressive moving average (VARMA) system (1) implied by asset pricing theory. The restrictions are derived under alternative specifications for both the short- and long-memory properties of the volatility process and the functional form of the central risk-return tradeoff. An additional important and well-documented empirical feature in financial time series is the financial leverage effect; see, for example, Black (1976), Engle and Ng (1993), and Yu (2005). The standard argument from Black (1976) is that bad news decreases the stock price and hence increases the debt-to-equity ratio (that is, financial leverage), making the stock riskier and increasing the equity risk premium.

The Merton (1973, 1980) linear relation between the equity risk premium, and the dividend yield. For plausible values of the parameters the elasticity varies wildly with \( \alpha \), the equity risk premium, and volatility \( \sigma^2 \) of equity returns is employed, namely,

\[ \alpha_t = \gamma \sigma^2_t. \]

Thus, the sign and magnitude of the risk-return tradeoff is given by the parameter \( \gamma \). The specific short-memory model used is a stationary AR(1) specification for monthly stock market variances,

\[ \sigma^2_{t+1} = \rho_0 + \rho_1 \sigma^2_t + \nu_{t+1}. \]

Under these assumptions, the elasticity of the stock market level with respect to volatility is given by

\[ \frac{d \log P_t}{d \log \sigma^2_t} = \frac{-\bar{\alpha}}{1 + r + \bar{\alpha} - p_1 (1 + g)}, \]

where \( \bar{\alpha} \) is the mean value of \( \alpha \), and \( g \) is the constant growth rate of expected dividends, so that \( E_i(D_{t+1}) = (1 + g)D_t \). Expression (5) is evaluated at the mean values of the risk premium and the dividend yield. For plausible values of the parameters the elasticity varies wildly with \( p_1 \), the short-memory parameter of the volatility process. For example, a
doubling of volatility from its average level reduces the value of the market by only 0.6% if \( \rho_1 = 0 \), but by 38.7% if \( \rho_1 = 0.99 \). For the empirical estimates of \( \rho_1 \), the drop is of the order 1.4%–2.2%, and in alternative AR(12) and IMA(1,3) models up to 22.5%.

Expression (5) is a convenient representation of the fact that the risk-return tradeoff and serial dependence in volatility have separate impacts on asset prices. An increase in either \( \gamma \) (and hence \( \alpha \), through equation [3]) or \( \rho_1 \) increases the magnitude of the (negative) elasticity. The recent empirical findings of a significantly positive risk-return tradeoff (\( \gamma > 0 \) in the present context) and long memory in volatility might now be expected to translate into an asset price elasticity of great magnitude. However, the derivation of equation (5) was based on AR(1) volatility. Hence, motivated by the recent empirical literature, we generalize the framework to allow for both short and long memory in volatility. In particular, we allow for long memory of the fractional integration or \( I(d) \) type. As the IMA(1,3) model implies nonstationarity of the standard \( I(1) \) unit root type, while the AR(1) model (of class \( I(0) \)) approaches the unit root model as \( \rho_1 \) tends to 1, it might be expected that fractional integration of volatility would lead to elasticity estimates in the same range as the empirical estimates quoted above, between 1.4% and 22%. However, perhaps surprisingly, our theoretical and empirical results point toward a true elasticity in the presence of long memory in volatility probably not exceeding 1%.

It is important to note the structure of the model, in particular that \( \sigma_i^2 \) is the conditional variance process, as is common in long-memory stochastic volatility models (see, for example, Breidt et al., 1998; and Robinson, 2001), even though it enters the mean specification (2) explicitly. Thus, writing \( r_t \) for the return over period \( t \),

\[
r_t = \frac{P_{t+1} - D_t - P_t}{P_t},
\]

returns are conditionally heteroskedastic with a time-varying risk premium,

\[
r_t = r_f + \alpha_t + \sigma_t \epsilon_t,
\]

where \( \epsilon_t \) is an \( \mathcal{F}_{t-1} \)-martingale difference in the mean and the variance, that is, \( E(\epsilon_t^2|\mathcal{F}_t) = 0 \) and \( E(\epsilon_t^2|\mathcal{F}_t) = 1 \), and it is assumed that \( \sigma_t \) is measurable with respect to \( \mathcal{F}_t \), the information set at time \( t \). Throughout the paper, we consider risk premium specifications of the type

\[
\alpha_t = g(\sigma_t^2, \sigma_{t-1}^2, \ldots),
\]

where \( g(\cdot) \) is a well-behaved function, and since \( \sigma_t^2 \) is measurable with respect to \( \mathcal{F}_t \), so is the risk premium \( \alpha_t \). It follows that

\[
E(r_t^2|\mathcal{F}_t) = \alpha_t^2 + \sigma_t^2 E(\epsilon_t^2|\mathcal{F}_t) = \alpha_t^2 + \sigma_t^2,
\]

and the conditional variance of returns is

\[
\text{Var}(r_t|\mathcal{F}_t) = E(r_t^2|\mathcal{F}_t) - \left( E(r_t|\mathcal{F}_t) \right)^2 = \sigma_t^2.
\]

Hence, \( \sigma_t^2 \) is indeed the conditional variance of the returns. This is a conditional heteroskedasticity-in-mean-type specification similar to Engle et al. (1987) and Bollerslev et al. (1988).

In section II below, we derive the relevant elasticity of the market level with respect to volatility under long memory in \( \sigma_t^2 \). We show that the elasticity is smaller in magnitude than earlier estimates, and much more stable under variations in the long-memory parameter than in the short-memory case. We also consider a new specification of the equity risk premium consistent with the stylized fact that equity returns do not exhibit long memory, as the simple specification (3) would require under long memory in \( \sigma_t^2 \), and the relevant elasticity remains small with this specification. In section III, we introduce our new bivariate modeling strategy, where volatility and equity returns are modeled jointly in the VARMA system (1). We derive testable cross-equation restrictions on the system from asset pricing theory. Section IV presents our empirical analysis, using monthly stock market returns along with high-frequency realized volatilities and model free VIX implied volatilities. We find evidence in favor of our VARMA specification for volatilities and returns, with a positive risk-return relation and long memory in volatility as well as a strong leverage effect, consistent with recent literature. However, in spite of the latter characteristics, our theoretical and empirical results show that the impact of volatility shocks on stock prices is small, with an elasticity probably not exceeding 1%, and short lived, dying out after about six months. Section V concludes.

### II. Consequences of Long Memory in Volatility

To model the long memory in volatility, we consider the class of autoregressive fractionally integrated moving average (ARFIMA) processes introduced by Granger and Joyeux (1980) and Hosking (1981). For excellent surveys on long-memory processes and fractional models, see Robinson (1994, 2003) and Baillie (1996), and for a textbook treatment see, for example, Beran (1994). A process is labeled an ARFIMA\((p, d, q)\) process if its \( d \)th difference is a stationary and invertible ARMA\((p, q)\) process. Here, \( d \) may be any real number and if \( -1/2 < d < 1/2 \) the process is stationary and invertible, which we assume throughout. For a precise statement, \( x_t \) is an ARFIMA\((p, d, q)\) process if

\[
\phi(L) \Delta^d x_t = \theta(L) \mu_t,
\]

where \( \phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i \) and \( \theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i \) are polynomials of order \( p \) and \( q \) in the lag operator \( L \).
EFFECT OF LONG MEMORY IN VOLATILITY

maximum likelihood estimator is empirically relevant region for the volatility processes that (2001; and Andersen et al., 2003). In particular, it is the
tions in finance (Lobato & Velasco, 2000; Andersen et al.,
0

concerned mainly with the stationary long-memory case
the short-memory case. Throughout this paper, we shall be
faster exponential rate of decline of the autocorrelations in
long memory, since the autocorrelations die out at a hyper-

The parameter \( d \) determines the (long) memory of the
process. If \( d > -1/2 \) the process is invertible and posses-
ses a linear (Wold) representation, and if \( d < 1/2 \) it is
covariance stationary. If \( d = 0 \) the spectral density is
bounded at the origin and the process has only weak
dependence (short memory). Furthermore, if \( d < 0 \) the
process is said to be antipersistent, and has mostly negative
autocorrelations, but if \( d > 0 \) the process is said to have
long memory, since the autocorrelations die out at a hyper-
bolic rate (and are not summable), in contrast to the much
faster exponential rate of decline of the autocorrelations in
the short-memory case. Throughout this paper, we shall be
concerned mainly with the stationary long-memory case
\( 0 \leq d < 1/2 \). This interval is relevant for many applica-
tions in finance (Lobato & Velasco, 2000; Andersen et al.,
2001; and Andersen et al., 2003). In particular, it is the
empirically relevant region for the volatility processes that
we study.

To estimate the parameters of model (8) we use the
maximum likelihood procedure of Sowell (1992). The max-
imum likelihood estimator is \( \sqrt{T} \)-consistent and asymptot-
ically normal. For details on the asymptotic normal distri-
bution and the associated regularity conditions, we refer the
reader to Sowell (1992). The estimator is asymptotically
efficient in the classical sense when the model is correctly
specified and Gaussian.

The well-documented long-memory property of volatility
suggests replacing the Poterba and Summers (1986) short-
memory volatility process (4) by

\[
\phi(L) \Delta^d (\sigma_t^2 - \bar{\sigma}^2) = \theta(L) \mu_t,
\]

where \( \bar{\sigma}^2 \) denotes the unconditional mean of \( \sigma_t^2 \), but retain-
ing their other assumptions. In particular, we maintain
relations (2) and (3). This allows isolating the effect of long
memory in volatility on the level of stock prices. We have the
following.

**Proposition 1.** If stock return volatility is governed by
the general ARFIMA\((p, d, q)\) process (10), then the elasticity
of the stock market level with respect to volatility, at the
mean values of the risk premium and the dividend yield, is
given by

\[
\frac{d \log P_t}{d \log \sigma_t^2} = \sum_{j=0}^{\infty} \psi_j \frac{-\tilde{\alpha}(1 + g)}{(1 + r_f + \tilde{\alpha})^{j+1}},
\]  

(11)

where \( \psi_j \) is the \( j \)-th impulse-response of \( \sigma_t^2 \).

This proposition shows that the simple form (5) of the
elasticity is special and a consequence of imposing the
short-memory AR(1) process (4) on volatility. The general
form of the elasticity from proposition 1 is an infinite
weighted sum of terms similar to the elasticity (5), with
weights given by the impulse-responses \( \psi_j \), which are sim-
ply the coefficients in the infinite-order moving average
representation of volatility,

\[
\sigma_t^2 - \bar{\sigma}^2 = \sum_{j=0}^{\infty} \psi_j \mu_{t-j}.
\]  

(12)

Square summability of the impulse-responses follows from
stationarity of the AR polynomial \( \phi(\cdot) \) and since \( d < 1/2 \).

Of course, higher-order short-memory ARMA\((p, q)\) models
for volatility, instead of the simple AR(1), do not in
general reproduce the summable geometric series and hence
closed-form solution for the elasticity (5). Nonetheless, we
include long memory from the outset, due to the recent
empirical evidence favoring this property.

Within the long-memory class there are simplifications in
relevant special cases. In particular, when \( p = q = 0 \), the
impulse-responses are given by \( \psi_j = \Gamma(j + d)/(\Gamma(d) \Gamma(j + 1)) \), and we have the following.

**Corollary 2.** In the situation from proposition 1, the
special case of an ARFIMA\((0, d, 0)\) or fractionally
integrated noise process for volatility yields the elasticity

\[
\frac{d \log P_t}{d \log \sigma_t^2} = \sum_{j=0}^{\infty} \frac{-\tilde{\alpha}(1 + g)}{(1 + r_f + \tilde{\alpha})^{j+1} \Gamma(d) \Gamma(j + 1)}.
\]

(13)

It is worth investigating the quantitative economic con-
sequences of these changes in asset price elasticities. Po-
terba and Summers (1986) evaluate the short-memory elas-
ticity (5) for the representative (monthly) parameter values
\( r_f = 0.035, \tilde{\alpha} = 0.006, \) and \( g = 0.00087 \). They find that
the elasticity varies from \(-0.006\) to \(-0.387\) as \( \rho_1 \) varies
from 0 to 0.99. In comparison, for the ARFIMA\((0, d, 0)\)
volatility specification from corollary 2 and using the same
values for \( r_f, \tilde{\alpha}, \) and \( g \), we get the long-memory elasticities
shown in table 1. Clearly, these are less variable than the
short-memory elasticities. Even as the memory parameter \( d \)
approaches 1/2, the boundary for stationarity, the elasticity
is only about \(-0.03\), an order of magnitude less than the
short-memory elasticities near the boundary for stationarity
of the AR(1) given by \( \rho_1 = 1 \). Given the mounting empirical
evidence favoring long memory in volatility, this suggests
that the high elasticities reported earlier should be inter-
preted with considerable caution. More generally, the results
TABLE 1.—Elasticity Estimates

<table>
<thead>
<tr>
<th>Memory Parameter (d)</th>
<th>$d \log P_t$</th>
<th>$d \log \sigma_t^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-0.006</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>-0.007982</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>-0.011053</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>-0.015307</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>-0.021198</td>
<td></td>
</tr>
<tr>
<td>0.49</td>
<td>-0.028415</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table presents values of the elasticity in corollary 2 using the parameter values $\epsilon_1 = 0.035$, $\bar{a} = 0.006$, and $\gamma = 0.00087$ from Poterba and Summers (1986).

indicate that the way in which volatility is entered in the model is crucial and should be considered carefully.

Inspection of relation (3) shows that in fact the memory properties of the volatility process carry over to the stock return process through the risk premium link. This is cause of concern since it implies that if volatility is stationary fractionally integrated, as empirical literature demonstrates, then so are stock returns, and indeed, the two are stationary fractionally cointegrated (see Christensen & Nielsen, 2006). Such strong serial dependence in stock returns seems empirically highly implausible, thus calling specification (3) into question. Of course, as finance theory suggests, a link between risk and return should be allowed. The problem with the tight link (3) is that the entire movement in the deviation from the mean $\sigma_t^2 - \bar{\sigma}^2$ in volatility spills over fully into risk premium deviations, $\alpha_t - \bar{\alpha} = \gamma(\sigma_t^2 - \bar{\sigma}^2)$, including long-memory components, which is not empirically warranted. Hence, an alternative form of the general link (7) is needed, and we consider the specification

$$\alpha_t - \bar{\alpha} = \gamma \Delta^d(\sigma_t^2 - \bar{\sigma}^2),$$

(13)

where the fractional filter $\Delta^d = (1 - L)^d$ removes spillover of the long-memory components into the risk premium, while still permitting a risk-return tradeoff. This specification is in the spirit of Ang et al. (2006), that is, only the short-memory component of volatility impacts expected returns. If increases in this component primarily occur in bad states and returns covary with volatility, then potentially the required premium is negative in bad states, and this is accommodated through the possibility of a negative value of $\gamma$. Lacking a model for the conditional mean of volatility, Ang et al. (2006) regress changes in VIX on stock returns to construct a factor mimicking innovations in market volatility, and similarly insert the innovations in a linear risk-return specification. Our long-memory model (10) in effect provides the model for the conditional mean of volatility and hence allows directly identifying the short-memory component of volatility, for use in equation (13).

With the new specification (13), the elasticity of the stock market level with respect to volatility is again given by equation (11), but now with the terms $\psi_i$ given by the impulse-responses of $\Delta^d(\sigma_t^2 - \bar{\sigma}^2)$, in other words, an ARMA($p$, $q$) process. For example, if volatility exhibits long memory of the ARFIMA(1, $d$, 0) type, then the resulting elasticity is again of the original form (5). Thus, what is important in assessing the relevance of the latter form of the stock price-volatility elasticity is not only the memory properties of volatility, but the combination of these with the risk premium link.

The specification of the risk premium link (13) still imposes ARMA properties on $\alpha_t$ if volatility is governed by the ARFIMA process (10) with $p$ or $q$ positive. The same problem arises in the original setup, with specifications (3) and (4), since $\alpha_t$ is restricted to follow an AR(1) process in that case, and with the same memory parameter $p_1$ as volatility. Even ARMA memory may be too much to force upon returns, and a martingale difference scheme for returns would seem a competitive alternative, based on casual empiricism. Hence, in the following we also consider the specification

$$\alpha_t - \bar{\alpha} = \gamma \mu_t,$$

(14)

that is, it is exactly the innovation to volatility from equation (10) that impacts the current risk premium and generates the risk-return relation, cf. Ang et al. (2006). By the fundamental structure of the stock return equation (2), the risk premium in specification (14) guarantees that the stock return (dividend included) is a martingale difference sequence. Whether this specification is preferred now becomes an empirical question. Under equation (14), the elasticity of the stock market level with respect to volatility is simply the constant $-\bar{\alpha}$, that is, minus the average risk premium. Using the same parameter value as in table 1, this is $-0.006$, again much smaller in magnitude than the elasticities from the short-memory model for positive $p_1$.

Of course, testable cross-restrictions are implied by the risk premium links (3), (13), and (14), since volatility and risk premia share the same innovation sequence. This is explored below.

III. Bivariate Modeling Strategy

In this section, we consider explicitly the bivariate time series model for returns and volatility. Rewriting equation (2) explicitly in terms of the expectation error $\varepsilon_t = \sigma_t \varepsilon_t$, we have

$$r_t = r_f + \alpha_t + \varepsilon_t.$$

(15)

We write $\omega^2_{\varepsilon}$ for var $(\varepsilon_t)$ and $\omega_{\alpha \mu}$ for cov $(\varepsilon_t, \mu_t)$, with $\mu_t$ the innovation in the volatility equation (4). The model based on equations (3) and (4) then in fact implies the bivariate structure

$$\begin{bmatrix} r_t - \bar{r} \\ \sigma_t^2 - \bar{\sigma}^2 \end{bmatrix} = \begin{bmatrix} 0 & \gamma p_1 \\ 0 & p_1 \end{bmatrix} \begin{bmatrix} r_{t-1} - \bar{r} \\ \sigma_{t-1}^2 - \bar{\sigma}^2 \end{bmatrix} + u_t,$$

(16)

where $\bar{r} = r_f + \gamma \bar{\sigma}^2$ and $\bar{\sigma}^2 = \rho_0(1 - \rho_1)$ are the unconditional means of the return and volatility processes,
respectively, and the vector \( u_t = (u_{1t}, u_{2t})' \) of system errors is given by

\[
\begin{bmatrix}
1 & \gamma \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
\varepsilon_t \\
\mu_t
\end{bmatrix},
\]

\[
\text{var}(u_t) = \Omega_u = \begin{bmatrix}
\omega_0^2 + 2\gamma \omega_{\mu} + \gamma^2 \omega_\mu^2 & \gamma \omega_{e\mu} + \omega_{e\mu}
\\
\gamma \omega_{e\mu} + \omega_{e\mu} & \omega_\mu^2
\end{bmatrix}.
\]

(17)

Note that equation (16) is a first-order vector autoregressive (VAR(1)) model with two testable zero restrictions. Furthermore, from equation (17) the errors \( u_{1t} \) and \( u_{2t} \) are correlated (through the risk premium parameter \( \gamma \)) even when \( \varepsilon_t \) and \( \mu_t \) are not. Indeed, imposing \( \omega_{e\mu} = 0 \) generates a testable cross-equation restriction linking mean and variance parameters (the ratio of the nonzero elements in the autoregressive coefficient matrix from equation [16] equals the ratio of \( \text{cov}(u_{1t}, u_{2t}) \) to var \( (u_{2t}) \) from equation [17]). Whether or not \( \omega_{e\mu} = 0 \), efficient estimation of this model would require simultaneous estimation of both equations, since the system errors in \( u_t \) are contemporaneously correlated in any case.

It is important to note that even though the unconditional variance of returns \( \text{var}(r_t) = (1 - \gamma^2 p^2)^{-1}(\omega_0^2 + 2\gamma \omega_{\mu} + \gamma^2 \omega_\mu^2) \) is constant in the model, it is the time-varying \( \sigma_t^2 = \text{var}(r_t | \mathcal{F}_t) \) which is modeled in the second equation of the system (16). In our empirical work below, we use either the VIX model free implied volatility for \( \sigma_t^2 \), or realized volatility estimated by a sum of squared intra-period returns (see section IV). Our estimation procedure uses the constant unconditional variance \( \Omega_u \) from equation (17) for the VAR system errors, which entails a potential efficiency loss relative to explicitly allowing for time-varying conditional variance of \( u_{1t} \) in the estimation, but retains consistency and asymptotic normality of the parameter estimates and is much simpler to implement and interpret, for example, in terms of the above parameter restrictions.

The VAR(1) model (16) is not empirically realistic, since it imposes short memory on volatility, contradicting findings of long memory in volatility from the literature. Long memory in volatility may be entered through the ARFIMA(\( d, q \)) process (10). This yields the alternative triangular fractionally cointegrated system

\[
\begin{align*}
r_t - \bar{r} &= \gamma(\sigma_t^2 - \bar{\sigma}^2) + \varepsilon_t \\
\Delta^4(\sigma_t^2 - \bar{\sigma}^2) &= \phi^{-1}(L)\theta(L)\mu_t
\end{align*}
\]

(18)

(19)

where \( \bar{r} \) and \( \bar{\sigma}^2 \) continue to denote the unconditional means, although these may now be different from before. The problem with this model is that it imposes long memory in the return equation (18), which is unwarranted empirically.

As a consequence of the empirical problems arising in the models (16) and (18)–(19) (short memory in volatility respectively long memory in returns), we turn to alternative specifications, where equation (13) or (14) is substituted for equation (3) to produce a better empirically founded return equation. An additional important and well-documented empirical feature in financial time series is the financial leverage effect; see Black (1976), Engle and Ng (1993), and Yu (2005). The standard argument from Black (1976) is that bad news decreases the stock price and hence increases the debt-to-equity ratio (that is, financial leverage), making the stock riskier and increasing future expected volatility. This generates a negative relationship between volatility and the lagged return. Thus, we consider either equation (10) for the volatility process or the specification

\[
\phi(L)\Delta^4(\sigma_t^2 - \bar{\sigma}^2) = \lambda r_{t-1} + \theta(L)\mu_t,
\]

(20)

where \( \lambda r_{t-1} \) accommodates the financial leverage effect.

The implications for the resulting bivariate systems are explored in the following propositions.

Proposition 3. Let \( X_t = (r_t - \bar{r}, \Delta^4(\sigma_t^2 - \bar{\sigma}^2))' \). Then the bivariate system for returns and volatilities is given by the vector ARMA (VARMA) model

\[
A(L)X_t = B(L)u_t,
\]

(21)

where \( u_t \) is given by equation (17), and the risk premium is given by equation (13).

A) No leverage effect: Assume stock return volatility is governed by the ARFIMA(\( d, q \)) process (10). Then the lag polynomials \( A(L) = I_2 + \sum_{i=1}^{p} A_i L^i \) and \( B(L) = I_2 + \sum_{i=1}^{q} B_i L^i \) satisfy the restrictions

\[
A_i = \begin{bmatrix}
0 & \gamma \phi_i \\
0 & \phi_i
\end{bmatrix}, \quad i = 1, \ldots, p,
\]

\[
B_i = \begin{bmatrix}
0 & \gamma \theta_i \\
0 & \theta_i
\end{bmatrix}, \quad i = 1, \ldots, q,
\]

where \( \phi_i \) and \( \theta_i \) are defined as in equation (20). In particular, there are \( 2(p + q) \) zero restrictions and \( \max \{ p + q - 1, 0 \} \) cross-equation restrictions. If, in addition, we assume \( \omega_{e\mu} = 0 \), then there is one additional cross-equation restriction when \( p + q \leq 1 \). Finally, \( \gamma \) is not identified if \( \omega_{e\mu} \neq 0 \) and \( p = q = 0 \).

B) Leverage effect included: Assume stock return volatility is governed by the levered ARFIMA(\( d, q \)) process (20). Then the lag polynomials \( A(L) = I_2 + \sum_{i=1}^{p} A_i L^i \) and \( B(L) = I_2 + \sum_{i=1}^{q} B_i L^i \) satisfy the restrictions

\[
A_i = \begin{bmatrix}
\gamma \lambda \lambda_1(i = 1) & \gamma \phi_1(i \neq 0) \\
\lambda_1(i = 1) & \phi_1(i \neq 0)
\end{bmatrix},
\]

\[
i = 1, \ldots, \max \{ 1, p \},
\]

\[
B_i = \begin{bmatrix}
0 & \gamma \theta_i \\
0 & \theta_i
\end{bmatrix}, \quad i = 1, \ldots, q,
\]

with \( 1(\cdot) \) the usual indicator function. There are \( 2(1(p = 0) + \max \{ 0, p - 1 \} + q) \) zero restrictions and \( p + q \).
cross-equation restrictions. If we assume \( \omega_{\varepsilon u} = 0 \), then there
is one additional cross-equation restriction.

The simultaneous system approach for asset returns and volatility behind this proposition is novel and the main contribution of this paper. Far more information can be obtained from the system than from analysis of each equation in isolation. Efficient estimation and inference requires a system approach such as full information maximum likelihood. Note that by construction \( Y_t \) has zero mean. The number of empirically testable restrictions from asset pricing theory is considerable. In part A with no leverage, the total number of parameters in the conditional mean equations of the corresponding unrestricted VARMA\((p, q)\) model is \( 4(p + q) \). When leverage is included in part B, the total number of parameters in the unrestricted VARMA\((\max \{1, p\}, q)\) model is \( 4(\max \{1, p\} + q) \). To illustrate the restrictions in this case, consider the special case where stock return volatility is governed by an ARFIMA\((0, d, 1)\) and the financial leverage effect is included. The bivariate process for returns and volatility is given by the VARMA\((1, 1)\) model

\[
r_t - \bar{r} = \gamma \lambda (r_{t-1} - \bar{r}) + \varepsilon_t + \gamma \mu_t + \gamma \theta \mu_{t-1},
\]

\[
\Delta^q(\sigma^2_t - \bar{\sigma}^2) = \lambda (r_{t-1} - \bar{r}) + \mu_t + \theta \mu_{t-1}.
\]

In the associated unrestricted VARMA\((1,1)\) model there are two \( 2 \times 2 \) matrices in the mean equations, for a total of eight parameters. Only the first-order moving-average coefficient \( \theta \) (corresponding to \( \theta_i \) in equation \((20)\)) and the parameters \( \gamma \) and \( \lambda \) appear in the restricted system above, that is, there are five constraints. Thus, this count matches the general number of restrictions from proposition 3B for the case \((p, q) = (0, 1)\).

Although the autoregressive polynomial in the ARFIMA model for volatility is of order \( p = 0 \) in this example, the autoregressive order of the resulting simultaneous VARMA system is one. This corresponds to the more general result in proposition 3B, that an autoregression of order \( p \) for volatility translates into an autoregressive component in the full system of order \( \max \{1, p\} \). All these restrictions are of course empirically testable. Finally, in the specific example considered here, if the additional assumption \( \omega_{\varepsilon u} = 0 \) is invoked, that is, the expectation revisions in returns are uncorrelated with the innovations to volatility, then the risk premium parameter, \( \gamma \), may be recovered from the estimated covariance matrix for the (stochastically correlated) system errors \( u_t \) defined in equation \((17)\) as \( \gamma = \text{cov}(u_{1t}, u_{2t})/\text{var}(u_{2t}) \), and this generates one additional restriction, for a total of six constraints.

Both the \( A_t \) and \( B_t \) coefficient matrices \((i \geq 1)\) exhibit nonzero terms in the first row (the return equation), even in the restricted form of the model in proposition 3. This degree of serial dependence in asset returns is potentially in conflict with empirical evidence, as well as efficient-markets theory. Hence, for the empirical analysis, it is worthwhile to compare with a version based on the alternative risk premium specification \((14)\).

**Proposition 4.** Let the risk premium be given by equation \((14)\). Then the bivariate system for returns and volatilities is given by the VARMA model \((21)\).

A) No leverage effect: Assume stock return volatility is governed by the ARFIMA\((p, d, q)\) process \((10)\). Then the lag polynomials \( A(L) = I_2 - \sum_{i=1}^{\max \{1, p\}} A_iL^i \) and \( B(L) = I_2 + \sum_{i=1}^{\max \{1, q\}} B_iL^i \) satisfy the restrictions

\[
A_i = \begin{bmatrix} 0 & 0 \\ 0 & \phi_i \end{bmatrix}, \quad i = 1, \ldots, p,
\]

\[
B_i = \begin{bmatrix} 0 & 0 \\ 0 & \theta_i \end{bmatrix}, \quad i = 1, \ldots, q,
\]

with \( \phi_i, \theta_i \) defined as in equation \((10)\). In particular, there are \( 3(p + q) \) zero restrictions.

B) Leverage effect included: Assume stock return volatility is governed by the levered ARFIMA\((p, d, q)\) process \((20)\). Then the lag polynomials \( A(L) = I_2 - \sum_{i=1}^{\max \{1, p\}} A_iL^i \) and \( B(L) = I_2 + \sum_{i=1}^{\max \{1, q\}} B_iL^i \) satisfy the restrictions

\[
A_i = \begin{bmatrix} 0 & 0 \\ \lambda \theta 1(i = 1) & \phi_i(p \neq 0) \end{bmatrix}, \quad i = 1, \ldots, \max \{1, p\},
\]

\[
B_i = \begin{bmatrix} 0 & 0 \\ 0 & \theta_i \end{bmatrix}, \quad i = 1, \ldots, q,
\]

with \( 1(\cdot) \) the usual indicator function and \( \phi_i, \theta_i \) defined as in equation \((20)\). Thus, there are \( 3(\max \{1, p\} + q) + 1 \) \((p = 0) - 1 \) zero restrictions.

Finally, in both parts A and B, if \( \omega_{\varepsilon u} \neq 0 \) the parameter \( \gamma \) is not identified, and if \( \omega_{\varepsilon u} = 0 \) then \( \gamma = \text{cov}(u_{1t}, u_{2t})/\text{var}(u_{2t}) \).

With this specification, there are \( 4(p + q) \) VARMA parameters in the unrestricted model in part A and 4 \((\max \{1, p\} + q) \) in part B. The restricted coefficient matrices \( A_i \) and \( B_i \) have zeros in the first row (the return equation) for \( i \geq 1 \). Hence, the process for asset returns is a martingale difference sequence in the restricted model. This potentially accords better with data than the model from proposition 3, a possibility explored in the empirical work below.

Before turning to the empirical analysis, we briefly return to the simple example from above, where volatility is governed by an ARFIMA\((0, d, 1)\) model and financial leverage is included. In this case, with \( q = 1 \), the restricted simultaneous VARMA\((1, 1)\) system from part B of proposition 4 is given by

\[
r_t - \bar{r} = \varepsilon_t + \gamma \mu_t,
\]

\[
\Delta^q(\sigma^2_t - \bar{\sigma}^2) = \lambda (r_{t-1} - \bar{r}) + \mu_t + \theta \mu_{t-1}.
\]
EFFECT OF LONG MEMORY IN VOLATILITY

TABLE 2.—Summary Statistics

<table>
<thead>
<tr>
<th>Data Series:</th>
<th>$r_{t}^{\text{nom}}$</th>
<th>$r_{t}^{\text{rft}}$</th>
<th>$g_{t}^{\text{nom}}$</th>
<th>$\pi_{t}$</th>
<th>Variance × 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Realized variance 1988:1–2002:12</td>
<td>0.9006%</td>
<td>0.4229%</td>
<td>0.3108%</td>
<td>0.2183%</td>
<td>0.2260</td>
</tr>
<tr>
<td>Mean</td>
<td>Std. dev.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: VIX implied variance 1990:1–2005:2</td>
<td>0.8507%</td>
<td>0.3493%</td>
<td>0.3363%</td>
<td>0.2057%</td>
<td>0.2521</td>
</tr>
<tr>
<td>Mean</td>
<td>Std. dev.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel C: Monthly volatility 1988:1–2002:12</td>
<td>0.2521%</td>
<td>0.1559%</td>
<td>2.4535%</td>
<td>0.6466%</td>
<td>0.1665</td>
</tr>
<tr>
<td>Mean</td>
<td>Std. dev.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Summary statistics for nominal one-month returns ($r_{t}^{\text{nom}}$), risk-free interest rates ($r_{t}^{\text{rft}}$), growth rates of dividends ($g_{t}^{\text{nom}}$), and inflation rates ($\pi_{t}$) are reported for the data period for which we have high-frequency realized variance observations (panel A) and the data period for which we have VIX implied variance observations (panel B). The last column of each panel presents the mean and standard deviation ($\times 100$) of the respective variance measures.

Even though returns form a martingale difference sequence, the system errors remain contemporaneously correlated, and simultaneous estimation is called for. This is so even if $\omega_{\text{eq}} = 0$, that is, if expectation revisions in returns are uncorrelated with innovations to volatility. In this case $\gamma$ may be recovered as stated in proposition 4. Note that there are six testable zero restrictions on the VARMA parameters in this example.

The various models above are explored empirically in the following section.

IV. Empirical Analysis

We use five-minute returns on the S&P 500 stock market index using linear interpolation following Müller et al. (1990), Dacorogna et al. (1993), and Barucci and Reno (2002), among others, to form monthly realized volatilities (annualized sample sum-of-squares), resulting in roughly 2,000 intramonthly return observations for each realized volatility (97 per day and approximately 20 trading days per month). The returns data cover the period January 1, 1988, to December 31, 2002. For more details on the construction of the realized volatility data, see Andersen, Bollerslev, and Diebold (2007) or Andersen, Bollerslev, Diebold, and Vega (2004). The use of five-minute returns can be justified on the basis of bias considerations when dealing with market microstructure noise induced effects which are expected to be small at the five-minute frequency; see, for example, the simulations by Nielsen and Frederiksen (forthcoming). Indeed, much recent work has been devoted to integrated variance estimation in the presence of noise, such as Bandi and Russell (2005a, 2006); Hansen and Lunde (2006); Oomen (2005, 2006); and Zhang, Mykland, and Ait-Sahalia (2005)–Bandi and Russell (forthcoming) and Barndorff-Nielsen and Shephard (2005, 2007) review this literature. There is some evidence that the optimal (in a mean squared error sense) sampling frequency is higher than five minutes when estimating daily realized volatilities. However, we consider monthly observations on realized volatility, hence employing roughly 2,000 return observations for each realized volatility. Consequently, we expect our monthly realized volatility measures to be very precise and not contaminated by market microstructure induced bias.

In addition, we use the monthly nominal dividend series for the index to form monthly growth rates $g_{t}^{\text{nom}}$. For the risk-free rate ($r_{t}^{\text{rft}}$), we use monthly data on the three-month Treasury bill rate (secondary market, middle rate), and we use the growth rate in the consumer price index (nondurables) for the inflation series ($\pi_{t}$). These data and the one-month nominal returns ($r_{t}^{\text{nom}}$) are obtained from Datastream. As Poterba and Summers (1986), we also compare with results where the monthly volatility series is based on implied volatilities from option prices. We use the model free (new) VIX implied volatility series from the Chicago Board of Options Exchange (CBOE). The new VIX is based on Britten-Jones and Neuberger (2000) and measures expected volatility over a thirty-day period by averaging the weighted prices of out-of-the-money puts and calls. Unlike the old VIX (now VXE), the new VIX is based on the S&P 500 index and is available from January 1990.

Our sample covers the period January 1990 to February 2005, and all variables are stated on a monthly basis.

Summary statistics are in table 2. Panel A shows statistics for the data set covering the period 1988:1–2002:12 (180 observations), for which we have high-frequency realized volatility data for $\sigma_{t}^{2}$. Panel B shows statistics for the data set covering the period 1990:1–2005:2 (182 observations), for which we use VIX implied volatilities for $\sigma_{t}^{2}$. As the standard deviations of returns (first column) are about equal in the two panels, the higher average volatility (last column) in panel B is due to the use of implied instead of realized volatility, perhaps reflecting a negative volatility risk premium in option prices. Furthermore, the standard deviation of volatility is considerably lower in panel B (0.17 versus 0.25), consistent with the notion that implied volatility measures expected future volatility which should be less volatile than actual realized volatility. In both data sets, average returns are considerably higher than interest rates, which in turn exceed inflation. As expected, the standard deviations of returns far exceed those of the risk-free rates.

Before proceeding to the empirical analysis of the bivariate systems of section III, we examine the univariate time
Table 3.—Univariate Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>$d$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\phi_3$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>log-lik</th>
<th>AIC</th>
<th>SIC</th>
<th>Q(12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, d, 0)</td>
<td>0.3978</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>868.70</td>
<td>-1733.40</td>
<td>-1730.21</td>
<td>18.17</td>
</tr>
<tr>
<td>(0, d, 1)</td>
<td>0.3757</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>0.0446</td>
<td>——</td>
<td>——</td>
<td>868.76</td>
<td>-1731.53</td>
<td>-1725.14</td>
<td>18.33</td>
</tr>
<tr>
<td>(0, d, 2)</td>
<td>0.4163</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>0.0314</td>
<td>-0.1087</td>
<td>——</td>
<td>869.34</td>
<td>-1730.69</td>
<td>-1721.11</td>
<td>18.37*</td>
</tr>
<tr>
<td>(0, d, 3)</td>
<td>0.3035</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>0.1376</td>
<td>0.0092</td>
<td>0.2046</td>
<td>872.43</td>
<td>-1734.87</td>
<td>-1722.10</td>
<td>13.76</td>
</tr>
<tr>
<td>(1, d, 0)</td>
<td>0.3803</td>
<td>0.0347</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>868.75</td>
<td>-1731.50</td>
<td>-1725.11</td>
<td>18.28</td>
</tr>
<tr>
<td>(1, d, 1)</td>
<td>0.4116</td>
<td>0.7447</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>-0.7594</td>
<td>——</td>
<td>868.71</td>
<td>-1729.42</td>
<td>-1719.84</td>
<td>18.14*</td>
</tr>
<tr>
<td>(1, d, 2)</td>
<td>-0.4863</td>
<td>0.9934</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>-0.0598</td>
<td>——</td>
<td>870.11</td>
<td>-1730.22</td>
<td>-1717.45</td>
<td>18.50*</td>
</tr>
<tr>
<td>(2, d, 0)</td>
<td>0.4323</td>
<td>-0.0209</td>
<td>-0.0856</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>869.14</td>
<td>-1730.28</td>
<td>-1720.70</td>
<td>18.25*</td>
</tr>
<tr>
<td>(2, d, 1)</td>
<td>0.3842</td>
<td>0.4688</td>
<td>-0.0178</td>
<td>——</td>
<td>——</td>
<td>-0.4381</td>
<td>——</td>
<td>868.75</td>
<td>-1727.50</td>
<td>-1714.73</td>
<td>18.30*</td>
</tr>
<tr>
<td>(3, d, 0)</td>
<td>0.3454</td>
<td>0.0759</td>
<td>-0.0482</td>
<td>0.2034</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>872.03</td>
<td>-1734.07</td>
<td>-1721.30</td>
<td>15.02</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Model</th>
<th>$d$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\phi_3$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>log-lik</th>
<th>AIC</th>
<th>SIC</th>
<th>Q(12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0, d, 0)</td>
<td>0.4877</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>982.92</td>
<td>-1961.83</td>
<td>-1958.64</td>
<td>35.64**</td>
</tr>
<tr>
<td>(0, d, 1)</td>
<td>0.4263</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>991.61</td>
<td>-1977.22</td>
<td>-1970.83</td>
<td>16.53</td>
</tr>
<tr>
<td>(0, d, 2)</td>
<td>0.3198</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>993.81</td>
<td>-1979.62</td>
<td>-1970.04</td>
<td>15.73</td>
</tr>
<tr>
<td>(0, d, 3)</td>
<td>0.3469</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>993.87</td>
<td>-1977.75</td>
<td>-1964.98</td>
<td>15.56*</td>
</tr>
<tr>
<td>(1, d, 0)</td>
<td>0.3146</td>
<td>0.4509</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>992.48</td>
<td>-1978.96</td>
<td>-1972.57</td>
<td>18.06</td>
</tr>
<tr>
<td>(1, d, 1)</td>
<td>0.3251</td>
<td>0.3474</td>
<td>——</td>
<td>——</td>
<td>0.1172</td>
<td>——</td>
<td>——</td>
<td>992.83</td>
<td>-1977.65</td>
<td>-1968.07</td>
<td>16.30</td>
</tr>
<tr>
<td>(1, d, 2)</td>
<td>0.3312</td>
<td>-0.0790</td>
<td>——</td>
<td>——</td>
<td>0.5441</td>
<td>——</td>
<td>——</td>
<td>993.84</td>
<td>-1977.68</td>
<td>-1964.91</td>
<td>15.68*</td>
</tr>
<tr>
<td>(2, d, 0)</td>
<td>0.3421</td>
<td>0.4559</td>
<td>-0.0760</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>992.97</td>
<td>-1977.95</td>
<td>-1968.37</td>
<td>15.60</td>
</tr>
<tr>
<td>(2, d, 1)</td>
<td>0.3857</td>
<td>0.7713</td>
<td>-0.2303</td>
<td>——</td>
<td>——</td>
<td>-0.3579</td>
<td>——</td>
<td>993.23</td>
<td>-1976.46</td>
<td>-1963.69</td>
<td>14.79</td>
</tr>
<tr>
<td>(3, d, 0)</td>
<td>0.4171</td>
<td>0.3712</td>
<td>-0.0434</td>
<td>-0.1119</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>993.86</td>
<td>-1977.71</td>
<td>-1964.94</td>
<td>14.02</td>
</tr>
</tbody>
</table>

Note: Log-lik is the value of the maximized log-likelihood function, AIC and BIC are the Akaike and Schwarz information criteria, and Q(12) is the Ljung-Box statistic for twelve lags which is asymptotically $\chi^2$ distributed with $12 - 1 - p - q$ degrees of freedom. For the Ljung-Box test, one and two asterisks denote rejection at 5% and 1% significance level, respectively. Asymptotic standard errors are in parentheses.

The series properties of the two individual volatility series (realized and implied), using the ARFIMA($p$, $d$, $q$) model (10). Table 3 shows the results of the univariate analysis. The results for the realized variance data set, from 1988 to 2002, are in panel A, and results for the VIX implied variance data set, from 1990 to 2005, are in panel B. The reported estimates of the long-memory parameter $d$ and the autoregressive and moving-average parameters $\phi$ and $\theta$ are found by application of the Sowell (1992) maximum likelihood method to deviations from the sample mean, using the ARFIMA package v1.01 for PcGive 10.1; see Doornik (2001) and Doornik and Ooms (2001). Asymptotic standard errors are in parentheses. Also reported are log-lik, the value of the maximized log likelihood function; AIC and BIC, the Akaike and Schwarz information criteria; and Q(12), the Ljung-Box statistic for twelve lags which is asymptotically $\chi^2$ distributed with $12 - 1 - p - q$ degrees of freedom. For the Ljung-Box test, one and two asterisks denote rejection at 5% and 1% significance level, respectively. Across both panels, the long-memory parameter $d$ is very strongly significant and estimated to values between 0.3 and 0.5, except a single case in panel A where a negative $d$ is accompanied by a near unit root in the autoregressive polynomial.

In panel A, the realized variance data set, the $(0, d, 3)$ model gets the best (lowest) AIC criterion and the highest likelihood value, the third-order MA parameter $\theta_3$ gets a $t$-statistic of 2.78, and the Ljung-Box statistic is by far the lowest in the panel. In principle, the likelihood ratio test does not reject reduction in the model, and the residuals actually appear serially uncorrelated in some of the other models estimated, but we hesitate to discard a statistically significant parameter ($\theta_3$). This is even more so because estimators in oversimplified models are inconsistent, whereas the only cost of overparameterization is inefficiency, but not inconsistency. We thus prefer the
table 4.—parameter estimates

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$\alpha$</th>
<th>$\rho_f$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poterba and Summers (1986)</td>
<td>0.06%</td>
<td>0.035%</td>
<td>0.087%</td>
</tr>
<tr>
<td>Realized variance 1988:1–2002:12</td>
<td>0.4777%</td>
<td>0.2046%</td>
<td>0.0926%</td>
</tr>
<tr>
<td>VIX implied variance 1990:1–2005:2</td>
<td>0.5011%</td>
<td>0.1436%</td>
<td>0.1306%</td>
</tr>
</tbody>
</table>

Note: Estimates of the mean equity risk premium ($\alpha$), risk-free interest rate ($\rho_f$), and growth rate of expected dividends ($\gamma$) are reported for the data period for which we have high-frequency realized variance observations and the data period for which we have VIX implied variance observations. Standard errors are in parentheses. For comparison, the parameter estimates from Poterba and Summers (1986) are reported in the first row.

ARFIMA(0, $d$, 3) model for realized volatilities. We also estimated higher-order models (results not reported), but they were clearly inferior to the chosen model.

In panel B, the implied variance data, model selection is less obvious. The (1, $d$, 0) model gets the best (lowest) SIC criterion and the second-best AIC in the panel. The first-order AR coefficient $\phi_1$ is estimated to 0.45, with a $t$-ratio of 3.27, and the Ljung-Box statistic shows no sign of misspecification. From the table, adding parameters such as $\phi_2$ or $\theta_0$ yields $t$-tests below 1. Based on these observations we select the ARFIMA(1, $d$, 0) model for VIX implied volatilities.

In the chosen (0, $d$, 3) and (1, $d$, 0) models, the long-memory parameter $d$ takes the values 0.30 and 0.31 and is strongly significant. Hence, the results in table 3 confirm the long-memory property of volatility found in the literature, and our estimates of the long-memory parameter are in line with those in, for example, Andersen et al. (2001) and Andersen et al. (2003).4 In particular, they are well within the stationary long-memory region $0 < d < 1/2$.

The long-memory property of volatility changes the implications for the level of the stock market. We have seen the first indication of this in table 1, for the simple case of a (0, $d$, 0) model for volatility. We now examine this issue in more detail for the empirically relevant volatility models from table 3. Table 4 shows the parameters needed for the comparison. The first row of the table gives the Poterba and Summers (1986) values for the real risk premium, risk-free rate, and dividend growth rate. The subsequent rows show the corresponding values for our two data sets, estimated from the averages in table 2.

The parameters from table 4 may now be used with proposition 1 to assess the stock market elasticity with respect to volatility changes. To this end, we also need the impulse-responses $\psi_j$ for the volatility process. These are exhibited in figure 1, for each of our two data sets. Initially, we use the impulse-response functions for the estimated ARFIMA models for $\sigma^2_t$ (using realized and implied vola-

tility, respectively, and shown as solid lines), but the figure also shows the impulse-responses for the corresponding ARMA models, that is, for $\Delta^d\sigma^2_t$ (shown as dotted lines), which we will need below.

Table 5 shows the resulting elasticities based on proposition 1. Panel A shows elasticities using the parameter values from the first row of table 4, and panel B shows elasticities using the parameter estimates from our data (last two rows of table 4). In each panel of table 5, elasticities are shown for the preferred model for each of our two data sets. The first column of elasticities is based on the ARFIMA impulse-responses from figure 1 and the risk premium link (3). All four estimated elasticities are of the same order of magnitude as those reported for short-memory autoregressive models by Poterba and Summers (1986, table 2) in their empirical section ranging from $-0.014$ to $-0.048$, whereas their reported elasticities for integrated moving-average models are an order of magnitude larger, at $-0.175$ to $-0.225$. The largest of our elasticity estimates, at $-0.047$, is obtained using implied volatility.

As discussed in section II, use of the risk premium link (3) imposes long memory on returns, which is not empirically warranted. Thus, the next column of elasticities in table 5 is based on the alternative risk premium link (13). Calculation of elasticities in this case amounts to substituting the impulse-responses for the corresponding ARMA models (that is, the models for $\Delta^d\sigma^2_t$) into equation (11). These impulse-response functions are also exhibited in figure 1 (dotted lines). In all cases, the resulting elasticities are even lower than in the previous column, although the parameters from table 4 are unchanged within each row of table 5. A glance at figure 1 shows that, as expected, the impulse-responses for each of the ARMA models are much lower than for the associated ARFIMA model for the same volatility series. It is these lower values for $\psi_j$ that feed into equation (11) and yield lower elasticities.

If we turn to risk premium link (14), thus allowing for martingale difference returns, the elasticity is given simply as $-\alpha$, shown in the last column of table 5. This produces estimates that are slightly lower but of the same order of magnitude as in the second-to-last column. Thus, both risk premium links that do not force long memory upon returns yield elasticity estimates of $-0.01$ or less, that is, smaller than in the first column of elasticities and smaller than the estimates in panel A. All in all, our results suggest that when allowing for long memory in volatility, the stock market elasticity with respect to volatility is small, probably no larger than about $-0.01$, and smaller than the estimates reported in Poterba and Summers (1986), which in light of the evidence on long memory appear to have been exaggerated.

The elasticities considered so far reveal something, but not everything, about the relation between volatility and the level of the stock market. To gain further insight into the dynamics of this relation, we next consider the explicit

4 We also estimated the same models for realized and implied standard deviations, as opposed to variances, following Andersen et al. (2001) and Andersen et al. (2003), and got slightly higher $d$-values, much in line with results in these studies. Applying the alternative semiparametric $d$ estimator of Geweke and Porter-Hudak (1983) using bandwidth $m = 20$ to the variance measurements yields $d = 0.40$ for VIX and $d = 0.29$ for realized, both with a standard error of 0.18.
simultaneous system of changes in both, that is, the multivariate model for stock market returns and volatility changes. This development is unique to the present paper. As volatility is governed by a fractionally integrated process, the volatility changes considered in the multivariate system are fractional differences. Thus, we implement in turn the various system specifications of section III for the bivariate process $X_t$. Here, the components of $X_t$ are the deviations from the unconditional means of the returns $r_t$ and the volatility changes $\Delta^d \sigma_t^2$. The unconditional means are 0.0090 and 0.0006, respectively, for the realized volatility data set, and 0.0085 and 0.0005, respectively, for the data set with implied volatilities.

We first apply the general unrestricted VARMA model (1) to the data set with realized volatilities. Based on the univariate analysis, the orders of the vector AR and MA polynomials are set at 1 and 3, respectively; that is, we implement an unrestricted VARMA(1, 3) model for $X_t = (r_t - \bar{r}, \Delta^d (\sigma_t^2 - \bar{\sigma}^2))'$. The autoregressive term accommodates the leverage effect from part B of propositions 3 and 4. In part A, the presence of this autoregressive term generates an additional four zero restrictions, since $p = 0$ in the preferred univariate model for realized volatility. Estimation results appear in table 6, panel A. From proposition 3, part A, asset pricing theory implies eight restrictions on the sixteen coefficients in the AR and MA polynomials. In particular, with $(p, q) = (0, 3)$, there are $2(0 + 3) = 6$ zero restrictions and max $\{ 0 + 3 - 1, 0 \} = 2$ cross-equation restrictions. Adding the four zero restrictions on the vector AR matrix yields a total of twelve restrictions. For part B of proposition 3, there is one less restriction, due to the inclusion of the additional leverage effect. We also estimate the model as restricted under part A and part B, separately, and the results are shown in panels B and C. In each panel, the first two columns of estimates correspond to the first-order AR matrix $A_1$, with coefficients for the return equation in the first row, and coefficients for the volatility equation in the second row. The following columns correspond to the MA matrices. The last two columns show the $2 \times 2$ variance-covariance matrix (times 1,000) of the VARMA residuals $u_t$. The estimation method is Gaussian maximum likelihood, and standard errors based on the sum of outer products of the score contributions are in parentheses. Strictly, Gaussianity is not needed for consistency and asymptotic normality of the parameter estimates in stationary VARMA models (Lütkepohl, 1991), and we get similar

| Table 5.—Elasticities Based on Estimated Parameter Values |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Panel A: Using parameter estimates from Poterba and Summers (1986) |
| Realized variance 1988:1–2002:12 | $(0, d, 3)$ | $-0.020477$ | $-0.007623$ | $-0.006$ |
| VIX implied variance 1990:1–2005:2 | $(1, d, 0)$ | $-0.028339$ | $-0.010175$ | $-0.006$ |
| Panel B: Using parameter estimates from our data sets |
| Realized variance 1988:1–2002:12 | $(0, d, 3)$ | $-0.030403$ | $-0.006391$ | $-0.004777$ |
| VIX implied variance 1990:1–2005:2 | $(1, d, 0)$ | $-0.047471$ | $-0.009031$ | $-0.005011$ |

Note: The table presents values of the elasticities in proposition 1 for both our data sets, using parameter estimates from Poterba and Summers (1986) (panel A) and parameter estimates from table 4 (panel B) along with the risk premium link (3) from Merton (1973, 1980) (elasticity (11)) or the new risk premium links (13) and (14).
Table 6.—Bivariate Estimates Using Realized Volatility 1988:1–2002:12

<table>
<thead>
<tr>
<th>Equation</th>
<th>$A_1$</th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_3$</th>
<th>Cov ($u_t$) × 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Unrestricted estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{1t} = r_t - \bar{r}$</td>
<td>-0.0153</td>
<td>6.6099</td>
<td>-0.0000</td>
<td>-4.5319</td>
<td>0.0054</td>
</tr>
<tr>
<td>$X_{2t} = \Delta \ln(a_t^2 - \bar{a}^2)$</td>
<td>0.0098</td>
<td>-0.1097</td>
<td>-0.0233</td>
<td>0.0935</td>
<td>-0.0057</td>
</tr>
<tr>
<td>Panel B: Restricted (Prop. 3) estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{1t} = r_t - \bar{r}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.5306</td>
<td>0</td>
</tr>
<tr>
<td>$X_{2t} = \Delta \ln(a_t^2 - \bar{a}^2)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1194</td>
<td>0</td>
</tr>
<tr>
<td>Panel C: Restricted (Prop. 3 with leverage) estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{1t} = r_t - \bar{r}$</td>
<td>-0.0619</td>
<td>0</td>
<td>0</td>
<td>0.0646</td>
<td>0</td>
</tr>
<tr>
<td>$X_{2t} = \Delta \ln(a_t^2 - \bar{a}^2)$</td>
<td>-0.0133</td>
<td>0</td>
<td>0</td>
<td>0.0139</td>
<td>0</td>
</tr>
<tr>
<td>Panel D: Restricted (Prop. 4) estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{1t} = r_t - \bar{r}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1539</td>
<td>0</td>
</tr>
<tr>
<td>$X_{2t} = \Delta \ln(a_t^2 - \bar{a}^2)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0548</td>
<td>0</td>
</tr>
<tr>
<td>Panel E: Restricted (Prop. 4 with leverage) estimates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_{1t} = r_t - \bar{r}$</td>
<td>-0.0049</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_{2t} = \Delta \ln(a_t^2 - \bar{a}^2)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0022</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Restricted and unrestricted FIML parameter estimates of the bivariate VARMA model $A(L)X_t = B(L)u_t$, for returns and variances from equation (21) are presented for the realized variance data set. Asymptotic standard errors are in parentheses.

inferences from using robust (sandwich-type) standard errors.

The unrestricted estimates in panel A are all insignificant, suggesting that the model is overfitted, but three of the six estimates in the restricted model in panel B (corresponding to proposition 3A) are significant. Note that in panel B, six estimates are reported in the eight columns of VARMA coefficients, but these correspond to only four distinct parameters, due to the two cross-equation restrictions. Proposition 3A shows how the restricted parameter estimates relate to the parameters from the univariate analysis in table 3. For example, the lower-right corner of $B_1$ corresponds to $\theta_1$. In particular, the parameter $\theta_1$ that was significant at the 5% level in the univariate analysis remains significant in the multivariate analysis, although the point estimate is somewhat lower than the univariate estimate. In panel C (restricted model based on proposition 3B), the lower left corner of $A_1$ estimates the additional leverage parameter $\lambda$, which is negative as expected, at $-0.013$, and strongly significant, with a $t$-statistic of about $-4.0$. This clearly shows the empirical relevance of the leverage effect, inducing an inverse relation between returns and future volatility.

The theory restrictions from proposition 3 are based on the risk premium link (13). Working within the VARMA framework, we have already abandoned the link (3), which has the empirically unrealistic triangular fractional cointegration implication; see equations (18)–(19). The possibility remains that returns should be modeled as martingale differences. This is allowed under the alternative risk premium link (14). From proposition 4A, there are nine asset pricing theory restrictions on the VARMA coefficients in this case, in addition to the four zero restrictions on $A_1$. The estimated model thus restricted appears in panel D of table 6. Here, two of the three restricted coefficient estimates are significant. Panel E shows estimates for the restricted model corresponding to proposition 4B, including the leverage effect, so a total of twelve restrictions are imposed. The estimated leverage effect $\lambda$ is very similar to that in panel C, at $-0.015$, with a $t$-statistic in excess of 5 in magnitude. Finally, the estimated residual covariance matrices are nearly identical in all five panels, indicating that not much information is lost by imposing the restrictions.

We also carry out the similar analysis using the data set with VIX implied volatilities. Table 7 shows the results. Based on the univariate results, we estimate a VARMA(1,0) model. Unlike in the realized volatility data set, three of the four coefficients in the unrestricted model (panel A) are now significant, and the residual covariance matrices for the restricted models without leverage (panels B and D) are quite different from those of the unrestricted model (panel A) and the restricted models including the leverage effect (panels C and E). Proposition 3A imposes two zero restrictions on the model, and proposition 3B imposes a single cross-equation restriction, so the four coefficient estimates in panel C correspond to only three parameters. Proposition 4A imposes three zero restrictions, and proposition 4B imposes two. Throughout the table, all VARMA coefficients in the restricted models are significant. The point estimates of the leverage coefficient $\lambda$ are similar to those in table 6, and the $t$-statistics now exceed 10 in magnitude.

Table 8 shows the results of formal hypothesis testing within the VARMA framework. Panel A shows results for the data set with realized volatilities, and panel B for the data set with implied volatilities. The first column shows the
maximized log likelihood values. The second column shows the likelihood ratio tests of the restrictions from propositions 3 and 4, respectively, against the appropriate unrestricted model. The degrees of freedom and p-values based on the asymptotic χ²-distribution are given in the next two columns. The following three columns show the estimated standard deviations of the expectations revisions and volatility innovations and their correlation. The final two columns show the estimates of the risk-return tradeoff and the leverage effect λ (asymptotic standard errors in parentheses). Note that γ and the structural parameters related to the expectations revisions and volatility innovations are only reported for the model corresponding to the proposition 3 restrictions, which is where they are identified without imposing ω_{μμ} = 0 (the estimated correlation between ϵ, and μ, ranges between −0.24 and −0.82 in the table). Thus, using proposition 3, the reported γ-estimate is calculated off the entries in the leading AR coefficient matrix, B_1, as γ = B_{1,12}/B_{1,22} in the data set with realized volatilities. In the data set with implied volatilities, γ is calculated off the entries in the leading AR coefficient matrix, A_1, as γ = A_{1,12}/A_{1,22}. From these expressions, asymptotic standard errors are calculated by the delta method.

In the data set with realized volatilities (panel A), we reject the restrictions from Propositions 3A and 4A at the 5% level, but not at the 1% level. Strikingly, there is no evidence against the models including the leverage effect (propositions 3B and 4B), which get p-values of 55% and 56%. Consistent with the results in table 6 (parameter significance and analysis of error covariance), we prefer the VARMA(1,3) model restricted according to proposition 4B for this data set. The pattern is similar but even more striking in the data set with implied volatilities (panel B), where the restrictions from propositions 3A and 4A are overwhelmingly rejected, whereas the restrictions from propositions 3B and 4B are not rejected at the 10% level. Thus, the results from both data sets confirm the empirical importance of the leverage effect. We prefer the model corresponding to proposition 3B for the data set with implied volatilities, since the two VARMA parameters dropped when moving from panel C to panel E in table 7 are significant at conventional levels.

The risk-return tradeoff parameter γ from equation (13) is estimated to values between 4.7 and 21.2 in the two data sets, and gets t-values of 2.70 and 1.94 in the data set with implied volatilities. Based on this evidence, γ appears significantly positive, reflecting a positive risk-return relation. Volatility increases are compensated through increased risk premia. The leverage parameter λ is strongly signifi-

### Table 7.—Bivariate Estimates Using Implied Volatility 1990:1–2005:2

<table>
<thead>
<tr>
<th>Equation</th>
<th>A_1</th>
<th>Cov (ω_t) × 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Unrestricted estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_{it} = r_t - ̂μ_t</td>
<td>−0.0156</td>
<td>5.2332</td>
</tr>
<tr>
<td>X_{it} = \Delta ϵ_t(σ_t^2 - ̂μ_t^2)</td>
<td>−0.0131</td>
<td>0.4653</td>
</tr>
<tr>
<td>Panel B: Restricted (Prop. 3) estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_{it} = r_t - ̂μ_t</td>
<td>0</td>
<td>5.2376</td>
</tr>
<tr>
<td>X_{it} = \Delta ϵ_t(σ_t^2 - ̂μ_t^2)</td>
<td>(1.9023)</td>
<td>0.4601</td>
</tr>
<tr>
<td>Panel C: Restricted (Prop. 3 with leverage) estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_{it} = r_t - ̂μ_t</td>
<td>−0.0819</td>
<td>3.0607</td>
</tr>
<tr>
<td>X_{it} = \Delta ϵ_t(σ_t^2 - ̂μ_t^2)</td>
<td>(0.0426)</td>
<td>0.4774</td>
</tr>
<tr>
<td>Panel D: Restricted (Prop. 4) estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_{it} = r_t - ̂μ_t</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X_{it} = \Delta ϵ_t(σ_t^2 - ̂μ_t^2)</td>
<td>0</td>
<td>0.4815</td>
</tr>
<tr>
<td>Panel E: Restricted (Prop. 4 with leverage) estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_{it} = r_t - ̂μ_t</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X_{it} = \Delta ϵ_t(σ_t^2 - ̂μ_t^2)</td>
<td>−0.0132</td>
<td>0.4788</td>
</tr>
</tbody>
</table>

Note: Restricted and unrestricted FIML parameter estimates of the bivariate VARMA model A(L)X_t = B(L)\epsilon_t, for returns and variances from equation (21) are presented for the VIX implied variance data set. Asymptotic standard errors are in parentheses.

### Table 8.—Bivariate Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>log-lik</th>
<th>LR</th>
<th>d.f.</th>
<th>p-value</th>
<th>ω_ε</th>
<th>ω_μ</th>
<th>ω_{μμ}/(ω_εω_μ)</th>
<th>γ</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>1,213.68</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Restricted (Prop. 3)</td>
<td>1,201.77</td>
<td>23.82</td>
<td>12</td>
<td>0.0215</td>
<td>0.0677</td>
<td>0.0019</td>
<td>−0.8195</td>
<td>21.1876</td>
<td>—</td>
</tr>
<tr>
<td>Restricted (Prop. 3 w. leverage)</td>
<td>1,208.81</td>
<td>9.74</td>
<td>11</td>
<td>0.5539</td>
<td>0.0457</td>
<td>0.0018</td>
<td>−0.5222</td>
<td>4.6559</td>
<td>−0.0133</td>
</tr>
<tr>
<td>Restricted (Prop. 4)</td>
<td>1,199.87</td>
<td>27.63</td>
<td>13</td>
<td>0.0102</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>−0.0149</td>
</tr>
<tr>
<td>Restricted (Prop. 4 w. leverage)</td>
<td>1,208.37</td>
<td>10.64</td>
<td>12</td>
<td>0.5600</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(0.0027)</td>
</tr>
<tr>
<td>Unrestricted</td>
<td>1,347.42</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Restricted (Prop. 3)</td>
<td>1,318.36</td>
<td>58.13</td>
<td>2</td>
<td>0.0000</td>
<td>0.0426</td>
<td>0.0010</td>
<td>−0.3466</td>
<td>11.1657</td>
<td>—</td>
</tr>
<tr>
<td>Restricted (Prop. 3 w. leverage)</td>
<td>1,346.65</td>
<td>1.54</td>
<td>1</td>
<td>0.2141</td>
<td>0.0413</td>
<td>0.0009</td>
<td>−0.2393</td>
<td>6.4113</td>
<td>−0.0128</td>
</tr>
<tr>
<td>Restricted (Prop. 4)</td>
<td>1,316.30</td>
<td>62.25</td>
<td>3</td>
<td>0.0000</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(3.2975)</td>
</tr>
<tr>
<td>Restricted (Prop. 4 w. leverage)</td>
<td>1,345.34</td>
<td>4.17</td>
<td>2</td>
<td>0.1245</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(0.0011)</td>
</tr>
</tbody>
</table>

Note: For each model in tables 6 and 7, this table presents the maximized log-likelihood value (log-lik), value of the LR test of overidentifying restrictions (LR), its degrees of freedom (d.f.) and associated p-value, the standard deviations of the expectation errors (ω_ε) and the volatility innovations (ω_μ) and their correlation (ω_{μμ}/(ω_εω_μ)), the risk-return tradeoff (γ), and the financial leverage effect (λ) for the realized variance data set (panel A) and the VIX implied variance data set (panel B). Asymptotic standard errors are in parentheses.
cant, with $t$-statistics above 4 in magnitude throughout, showing that the leverage effect is indispensable for the model. This is further reinforced by noting that the point estimates vary very little across specifications, from $-0.013$ to $-0.015$. Thus, negative return innovations spur volatility increases. In such bad states, stocks that covary with volatility pay off, and Ang et al. (2006) find evidence of correspondingly lower required premia for these stocks in the cross section. Our aggregate time series evidence is that the overall market risk premium is positive, consistent with, for example, Brandt and Kang (2004) and Ghysels et al. (2005).

We end this section by examining further the preferred model for the data set with VIX implied volatilities, subject to the restrictions of proposition 3B, including both a risk-return tradeoff and the leverage effect. The VARMA form of this model is expressed in terms of the system errors $u_t$. Recasting the model in terms of the original return expectation revisions $\varepsilon_t$ and volatility innovations $\mu_t$, using equation (17), we obtain

$$r_t - \bar{r} = 6.41 \Delta^d (\sigma_t^2 - \bar{\sigma}^2) + \varepsilon_t,$$

(3.30)

$$\Delta^d (\sigma_t^2 - \bar{\sigma}^2) = 0.48 \Delta^d (\sigma_{t-1}^2 - \bar{\sigma}^2) - 0.01 (r_{t-1} - \bar{r}) + \mu_t,$$

(0.02)

(0.001)

with structural error covariance matrix

$$\begin{bmatrix} \omega^2 & \omega_{\varepsilon\mu} \\ \omega_{\mu\varepsilon} & \omega^\mu \end{bmatrix} = \begin{bmatrix} 1.752.0 & -9.421 \\ -9.421 & 0.7951 \end{bmatrix} \times 10^{-6}.$$  

(24)

In the estimated version of the restricted model, returns depend positively on volatility changes, which in turn depend positively on lagged volatility changes and negatively on lagged returns. The variance of the expectation revision $\varepsilon_t$ is more than 1,000 times that of the volatility shock $\mu_t$. The estimated covariance between the two, $\omega_{\varepsilon\mu}$, exceeds the variance of the volatility shock $\omega^\mu$ in magnitude. Probably the true covariance is nonzero, implying that it is not appropriate to impose this zero restriction in order to calculate the risk premium parameter $\gamma$ in the unrestricted and proposition 3A and 4A models in table 8. Fortunately, this restriction is not needed for estimation of $\gamma$ in the preferred proposition 3B model for the data set with implied volatility, where the estimate is 6.41 and significant.

The risk-return tradeoff is thus positive (p-value of 5.19%), and the leverage hypothesis is confirmed through the negative values of both $\lambda$ and the covariance $\omega_{\varepsilon\mu}$ between innovations to returns and volatility in equation (24). Obviously, the model implies that returns are not martingale differences. In particular, substituting equation (23) in (22) shows the dependence of returns on lagged returns and volatility changes.

With the estimated multivariate models in hand, it is possible to give a much more detailed, dynamic picture of the relation between volatility and stock prices than that entailed in the simple elasticities considered earlier. In particular, we may solve equation (22) for the impulse-response function for future returns with respect to a unit shock in current volatility. This is not an ordinary univariate impulse-response function, but a response function for the variable in the first equation of the system with respect to a shock in the variable from the second equation, so our multivariate framework is crucial for this approach. The resulting impulse-response function is shown as the dotted line in figure 2. The dynamics inherent in the estimated system is evident from the shape of the impulse-response
function. This is in contrast to the simpler dynamics in the short-memory multivariate system (16) which results from the model (3)–(4). The \( j \)th impulse-response in the simple system is \( \gamma p_{t+j} \), that is, geometrically declining, whereas our impulse-response function in figure 2 is clearly nonmonotonic, a result of the more complicated multivariate model structure.

To better complement the information contained in the elasticities considered earlier, we may in addition examine the impulse-responses of log stock prices with respect to volatility shocks. Expressing log prices as infinite sums of future returns, \( \log P_{t+j} = -\sum_{k=0}^{\infty} r_{t+j+k} \), we calculate the impulse-response function of log prices from the corresponding sums of impulse-responses of returns. The result is shown as the solid line in figure 2. Note that after the initial price drop in reaction to the volatility shock, the market reverts back to the original level, to finally stabilize at the original level after around six months. The initial drop is slightly less than the value of \( \gamma \). These dynamic response patterns reinforce the impression that the true impact of volatility changes on the stock market level, appropriately accounting for long memory in volatility, is quite modest and short lived.

V. Concluding Remarks

Recent empirical literature documents the presence of a long-memory component in volatility. In addition, at least part of the recent literature finds a significantly positive risk-return tradeoff. In conjunction, the two would seem to make for a strong and long-lasting effect on asset values of shocks to volatility. We find that on the contrary the instantaneous as well as dynamic impacts of volatility changes on stock prices are modest and short lived, with an elasticity of stock price with respect to volatility changes of no more than 1%, and the impulse-response function dying out after about six months in our empirically supported model. Thus, the effect is not as first anticipated. We introduce a new VARMA approach which allows for general joint dynamics of returns and volatility, and there is evidence in favor of the theoretical restrictions in our preferred model. The results are consistent with earlier findings of both a positive risk-return tradeoff, long memory in volatility, and a strong financial leverage effect, showing in addition that these interact to yield a perhaps surprisingly low impact of volatility shocks on asset values. Our work complements the literature on realized and implied volatility, the financial leverage effect, and the risk-return relation, and contributes toward an understanding of the roles they play in determining asset values.

REFERENCES


———, “Multipower Variation and Stochastic Volatility” (pp. 73–82), in Albert N. Shiryaev, Maria R. Grossinho, Paulo E. Oliveira, and Manuel L. Esquivel (Eds.), Stochastic Finance (Springer, 2005).


Christensen, Bent J., and Morten Ø. Nielsen, “Asymptotic Normality of Narrow-Band Least Squares in the Stationary Fractional Cointe-
EFFECT OF LONG MEMORY IN VOLATILITY


Appendix: Proofs

Proof of Proposition 1. By the chain rule for differentiation

\[
\frac{dP_t}{d\sigma^2} = E \left( \sum_{j=0}^{\infty} \frac{dP_t}{d\alpha_{t+j}} \frac{d\alpha_{t+j}}{d\sigma^2} \right),
\]

where the first term on the right-hand side is the derivative in an infinite order Taylor expansion of \( P_t \) around \( \alpha \), given by equation (6) of Poterba and Summers (1986) as

\[
\frac{dP_t}{d\alpha_{t+j}} = \psi_j,
\]

where \( \psi_j \) is the impulse-response (12), that is,

\[
\frac{d\alpha_{t+j}}{d\sigma^2} = \gamma,
\]

by equation (3). Collecting the terms we find that

\[
\frac{d \log P_t}{d \log \sigma^2} = \gamma \sum_{j=0}^{\infty} \psi_j \left( 1 + r_t + \alpha \right)^{j-1} (r_t + \alpha - g),
\]

\[
\frac{d \log \sigma^2}{d \log P_t} = \gamma \sum_{j=0}^{\infty} \psi_j \left( 1 + r_t + \alpha \right)^{j-1} (r_t + \alpha - g).
\]

Applying equation (3) and setting the risk premium \( \alpha \) and the dividend yield \( D/P \), equal to their mean values \( \bar{\alpha} \) and \( r_t + \bar{\alpha} - g \), we get the desired result.

Proof of Corollary 2. Follows immediately from proposition 1 and the formula for the impulse-responses, \( \psi_j = \Gamma(j + d)/(\Gamma(d)\Gamma(j + 1)) \).
Proof of Proposition 3. Rewrite equation (20) as

\[ X_{2t} = \Delta(t^2 - \bar{t}) = \lambda X_{1,t-1} + \sum_{i=1}^{p} \phi_i X_{2,t-i} + \theta(L) \mu, \]  

(A1)

Inserting into the definition of the returns (15) and using the risk premium link (13) we find that

\[ X_{1t} = r_t - \bar{r} = \gamma X_{2t} + \epsilon_t, \]  

which implies that

\[ X_{1t} = \gamma \lambda X_{1,t-1} + \gamma \sum_{i=1}^{p} \phi_i X_{2,t-i} + \gamma \theta(L) \mu + \epsilon_t. \]  

(A2)

Combining equations (A1) and (A2) and defining \( u \) as in equation (17) we get the system (21) with the stated restrictions for part B, and part A follows by setting \( \lambda = 0 \).

In part A, there are 2\( p \) zero restrictions and \( p \) cross-equation restrictions in \( A(L) \) and there are 2\( q \) zero restrictions and \( q \) cross-equation restrictions in \( B(L) \). However, one of those is used to identify \( \gamma \) since when \( \omega_{\mu} \neq 0 \), \( \gamma \) cannot be determined from the elements of \( \Omega_{\mu} \) in equation (17) but is determined in either \( A(L) \) or \( B(L) \), which gives the \( p + q \) cross-equation restrictions. If, in addition, we assume \( \omega_{\mu} = 0 \), \( \gamma \) is determined in \( \Omega_{\mu} \) as before, and there is one additional cross-equation restriction.

Proof of Proposition 4. As in the proof of proposition 3, the system follows by inserting \( X_{2t} \) from equation (A1) into the definition of the returns (15) but now using the risk premium link (14). In part A, there are 3\( p \) and 3\( q \) zero restrictions in \( A(L) \) and \( B(L) \), respectively. In part B, there are 3\( \max \{1, p\} - 1 \) and 3\( q \) zero restrictions in \( A(L) \) and \( B(L) \), respectively. As before, \( \gamma \) cannot be determined from the elements of \( \Omega_{\mu} \) when \( \omega_{\mu} \neq 0 \), but if \( \omega_{\mu} = 0 \) then \( \gamma \) is identified as in the proof of proposition 3.