Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility

Michael J. DUEKER
Federal Reserve Bank of St. Louis, St. Louis, MO 63166

This article introduces four models of conditional heteroscedasticity that contain Markov-switching parameters to examine their multiperiod stock-market volatility forecasts as predictions of options-implied volatilities. The volatility model that best predicts the behavior of the options-implied volatilities allows the Student-t degrees-of-freedom parameter to switch such that the conditional variance and kurtosis are subject to discrete shifts. The half-life of the most leptokurtic state is estimated to be a week, so expected market volatility reverts to near-normal levels fairly quickly following a spike.

KEY WORDS: Asset-price volatility; Conditional heteroscedasticity; Kurtosis.

Volatility clustering is a well-documented feature of financial rates of return: Price changes that are large in magnitude tend to occur in bunches rather than with equal spacing. A natural question is how long financial markets will remain volatile because volatility forecasts are central to calculating optimal hedging ratios and options prices. Indeed we can study the behavior of options-implied stock-market volatilities to find stylized facts that parametric volatility models should aim to capture. Two stylized facts that conventional volatility models, notably generalized autoregressive conditional heteroscedasticity [GARCH, Bollerslev (1986)], find hard to reconcile are that (1) conditional volatility can increase substantially in a short amount of time at the onset of a turbulent period and (2) the rate of mean reversion in stock-market volatility appears to vary positively and nonlinearly with the level of volatility. In other words, stock-market volatility does not remain persistently two to three times above its normal level in the same way it can persist at 30–40% above normal.

Hamilton and Susmel (1994) and Lamoureux and Lastrapes (1993) highlighted the forecasting difficulties of conventional GARCH models by showing that they can provide worse multiperiod volatility forecasts than constant-variance models. In particular, multiperiod GARCH forecasts of the volatility are too high in a period of above-normal volatility. Friedman and Laibson (1989) addressed the forecasting issue by not allowing the conditional variance in a GARCH model to respond proportionately to "large" and "small" shocks. In this way, the conditional variance is restrained from increasing to a level from which volatility forecasts would be undesirably high. One drawback of this approach is that in such a model the conditional volatility might understate the true variance by not responding sufficiently to large shocks and thereby never be pressed to display much mean reversion. Thus, such "threshold" models do not necessarily address the two stylized facts listed previously—sharp upward jumps in volatility, followed by fairly rapid reversion to near-normal levels. This article endeavors to craft a volatility model that can address these two stylized facts from within the class of GARCH models with Markov-switching parameters. Markov-switching parameters ought to enable the volatility to experience discrete shifts and discrete changes in the persistence parameters.

Partly in response to Lamoureux and Lastrapes (1990), who observed that structural breaks in the variance could account for the high persistence in the estimated conditional variance, Hamilton and Susmel (1994) and Cai (1994) introduced Markov-switching parameters to autoregressive conditional heteroscedasticity (ARCH) models, and I extend the approach to GARCH models because the latter are more flexible and widely used. Section 1 presents tractable methods of estimating GARCH models with Markov-switching parameters. Section 2 describes four specifications that are estimated and provides in-sample and out-of-sample goodness-of-fit test results. Section 3 uses the estimated models to generate multiperiod forecasts of stock-market volatility and compares the forecasts with options-implied volatilities to see which of the GARCH/Markov-switching models best explains the two stylized facts described previously.

1. GARCH/MARKOV–SWITCHING VOLATILITY MODELS

Each of the volatility-model specifications will assume a student-t error distribution with n_t df in the dependent variable y:

$$y_t = \mu_t + \varepsilon_t, \tag{1}$$

$\varepsilon_t \sim$ student-t (mean $= 0, n_t, h_t$), $n_t > 2$. In all of the models, the conditional mean, μ_t, is allowed to switch according to a Markov process governed by a state variable, S_t: $\mu_t = \mu_l S_t + \mu_h(1 - S_t)$, $\quad S_t \in \{0, 1\} \quad \forall t$,

$$\begin{aligned} \Pr(S_t = 0 | S_{t-1} = 0) &= p \\ \Pr(S_t = 1 | S_{t-1} = 1) &= q. \end{aligned} \tag{2}$$

The unconditional probability of $S_t = 0$ equals $(1 - q)/(2 - p - q)$. The variance of ε_t is denoted σ_t^2 and is a function of n_t and h_t in all of the models considered such that

Journal of Business & Economic Statistics
January 1997, Vol. 15, No. 1

$\sigma_t^2 = f(n_t, h_t)$, where the specific function f varies across the models. In all cases, however, h is assumed to be a GARCH(1, 1) process with Markov-switching parameters also governed by S so that a general form for h is

$$h_t(S_t, S_{t-1}, \ldots, S_0) = \gamma(S_t) + \alpha(S_{t-1})\varepsilon_{t-1}^2 + \beta(S_{t-1})h_{t-1}(S_{t-1}, \ldots, S_0). \quad (3)$$

Note that the presence of lagged h on the right side of (3) causes the GARCH variable to be a function of the entire history of the state variable. If h were an ARCH(p) process, then h would depend only on the p most recent values of the state variable, as in the work of Cai (1994) and Hamilton and Susmel (1994). Here I discuss how methods described by Kim (1994) can be applied to make estimation feasible for GARCH processes subject to Markov switching.

Clearly it is not practical to examine all of the possible sequences of past values of the state variable when evaluating the likelihood function for a sample of more than 1,000 observations because the number of cases to consider exceeds 1,000 by the time $t = 10$. Kim (1994) addressed this problem by introducing a collapsing procedure that greatly facilitates evaluation of the likelihood function at the cost of introducing a degree of approximation that does not appear to distort the calculated likelihood by much. The collapsing procedure, when applied to a GARCH process, calls for treating the conditional dispersion, h_t, as a function of at most the most recent M values of the state variable S. For the filtering to be accurate, Kim noted that, when h is p-order autoregressive, then M should be at least $p + 1$. In the GARCH(1, 1) case, $p = 1$, so we would have to keep track of M^2 or four cases, based on the two most recent values of a binary state variable. Thus, h_t is treated as a function of only S_t and S_{t-1}: $h_t^{(i,j)} = h_t\ (S_t = i, S_{t-1} = j)$.

Denoting φ_t as the information available through time t, I keep the number of cases to four by integrating out S_{t-1} before plugging lagged h into the GARCH equation:

$$h_t^{(i)} = \sum_{j=0}^{1} \Pr(S_{t-1} = j | S_t = i, \varphi_t) h_t^{(i,j)}. \quad (4)$$

This method of collapsing of $h_t^{(i,j)}$ onto $h_t^{(i)}$ at every observation gives us a tractable GARCH formula, which is approximately equal to the exact GARCH equation from Equation (3):

$$h_t^{(i,j)} = \gamma\ (S_t = i) + \alpha\ (S_{t-1} = j)(\varepsilon_{t-1}^{(j)})^2 + \beta\ (S_{t-1} = j) h_{t-1}^{(j)}. \quad (5)$$

Note that the collapsing procedure integrates out the first lag of the state variable, S_{t-1}, from the GARCH function, h_t, at the right point in the filtering process to prevent the conditional density from becoming a function of a growing number of past values of the state variable.

From this general framework, I choose specifications that differ according to the parameters that switch and the relationship between the GARCH process, h, and the variance σ^2. In several specifications, the GARCH processes are functions of lagged values of the state variable but not the contemporaneous value, S_t. For these, I treat h_t as a function of only S_{t-1}, so I only need to keep track of two cases: $h^{(j)} = h\ (S_{t-1} = j)$. Furthermore, after integrating out S_{t-1}, I am left with a scalar in the collapsing process:

$$\hat{h}_t = \Pr(S_{t-1} = 0 | \varphi_t) h_t^{(0)} + \Pr(S_{t-1} = 1 | \varphi_t) h_t^{(1)}. \quad (6)$$

A tractable GARCH equation is then an even simpler version of Equation (5):

$$h_t^{(j)} = \gamma + \alpha\ (S_{t-1} = j)(\varepsilon_{t-1}^{(j)})^2 + \beta\ (S_{t-1} = j)\hat{h}_{t-1}. \quad (7)$$

Another feature of this GARCH/Markov-switching framework is that the state variable implies a connection between the mean stock return and the variance and possibly kurtosis. If the mean stock return is lower in the high-volatility state, then the model can explain negatively skewed distributions, both unconditional and conditional on available information. The student-t distributions have zero skewness only when conditional on particular values of the state variables, which are unobservable.

2. FOUR SPECIFICATIONS AND ESTIMATION RESULTS

The first specification is a GARCH analog to Cai's (1994) ARCH model with Markov switching in γ. The variance is assumed to follow a GARCH process so that $\sigma_t^2 = h_t$ and the only parameter in h_t subject to Markov switching is γ. This type of switching is tantamount to allowing shifts in the unconditional variance because the unconditional variance of the ordinary, constant-parameter GARCH(1, 1) process is $\gamma/(1 - \alpha - \beta)$. For this model, the GARCH variance takes the form

$$h_t^{(i,j)} = \gamma\ (S_t = i) + \alpha(\varepsilon_{t-1}^{(j)})^2 + \beta h_{t-1}^{(j)}, \quad (8)$$

with constant α and β. We denote this model as the GARCH-UV model for GARCH with switching in the unconditional variance. In practice, we parameterize $\gamma(S_t)$ as $g(S_t)\gamma$, where $g\ (S = 1)$ is normalized to unity.

The second specification is a GARCH analog to Hamilton and Susmel's (1994) ARCH model with Markov switching in a normalization factor g, where the variance $\sigma_t^2 = g_t h_t$. In this case, the GARCH Equation (5) takes the form

$$h_t^{(j)} = \gamma + \frac{\alpha}{g\ (S_{t-1} = j)}\ (\varepsilon_{t-1}^{(j)})^2 + \beta \hat{h}_{t-1}, \quad (9)$$

where γ and β are constant and $g\ (S = 1)$ is normalized to unity. I denote this model as the GARCH-NF model for GARCH with switching in the normalization factor, g. Note that in the GARCH-NF model the GARCH process in Equation (9) is not a function of S_t, so estimation is somewhat simplified.

The third specification is a Markov-switching analog to Hansen (1994), in which the variance follows a GARCH process ($\sigma_t^2 = h_t$) and the student-t degrees-of-freedom parameter is allowed to switch. Hansen (1994) introduced a model in which the student-t degrees-of-freedom parameter, n_t, is allowed to vary over time as a probit-type function of variables dated up to time $t-1$. Because Hansen's (1994) specification is not conducive to multiperiod forecasting, however, I chose to make n_t follow a Markov process governed by S_t: $n_t = n_l S_t + n_h(1-S_t)$. Although n_t does not enter the GARCH Equation (7) in this specification, the GARCH process is still a function of the state variable because state-switching in the mean implies that ε is a function of the state variable:

$$h_t^{(j)} = \gamma + \alpha(\varepsilon_{t-1}^{(j)})^2 + \beta \hat{h}_{t-1}. \tag{10}$$

Because the kurtosis of a student-t random variable equals $3(n_t-2)/(n_t-4)$ and is uniquely determined by n_t, we call this the GARCH-K model for GARCH with switching in the conditional kurtosis.

The fourth specification is similar to the GARCH-K model except the variance is assumed to be

$$\sigma_t^2 = h_t n_t/(n_t-2) \tag{11}$$

rather than $\sigma_t^2 = h_t$. In this model, the GARCH process h_t scales the variance of ε_t for a given value of the shape parameter n_t. Here it is convenient to define $v_t = 1/n_t$ so that $(1-2v_t) = ((n_t-2)/n_t)$ and the GARCH Equation (7) becomes

$$h_t^{(j)} = \gamma + \alpha(1-2v_{t-1}^{(j)})(\varepsilon_{t-1}^{(j)})^2 + \beta \hat{h}_{t-1}. \tag{12}$$

I denote this specification as the GARCH-DF model for GARCH with switching in the degrees-of-freedom parameter. As in the GARCH-NF and GARCH-K models, h is a function of S_{t-1}, but not S_t, in the GARCH-DF model. The GARCH-DF model shares two features with the GARCH-NF model: The variance is subject to discrete shifts, and the lagged squared residuals are endogenously downweighted in states in which σ^2/h is large. With the GARCH-K model, the GARCH-DF model shares the feature of time-varying conditional kurtosis so that conditional fourth moments are not assumed to be constant.

I also report results on the usual GARCH(1, 1) model with Markov switching in the mean and a model of switching ARCH with a leverage effect (SWARCH-L), as in the work of Hamilton and Susmel (1994). The SWARCH-L model has switching in a normalizing factor in the variance: $\sigma_t^2 = g_t h_t$, where h_t follows an ARCH(2) process with a leverage effect:

$$h_t^{(j,k)} = \gamma + \frac{(\alpha_1 + \delta D_{t-1}^{(j)})}{g\ (S_{t-1}=j)}\ (\varepsilon_{t-1}^{(j)})^2 + \frac{\alpha_2}{g\ (S_{t-2}=k)}\ (\varepsilon_{t-2}^{(k)})^2, \tag{13}$$

where $D_{t-1}^{(j)}$ is a dummy variable that equals 1 when $\varepsilon\ (S_{t-1}=j)_{t-1} < 0$. The leverage effect posits that negative stock returns increase debt-to-equity ratios, making firms riskier initially. Hence the leverage-effect parameter δ is expected to have a positive sign.

The log-likelihood function for the GARCH-DF model, for example, is

$$\ln L_t^{(i,j)} = \ln\Gamma(.5(n_t^{(i)}+1)) - \ln\Gamma(.5n_t^{(i)}) - .5\ln(\pi n_t^{(i)} h_t^{(j)}) - .5(n_t^{(i)}+1) \times \ln\left(1 + \frac{(\varepsilon_t^{(i)})^2}{h_t^{(j)} n_t^{(i)}}\right), \tag{14}$$

where $i \in \{0,1\}$ corresponds with $S_t \in \{0,1\}, j \in \{0,1\}$ corresponds with $S_{t-1} \in \{0,1\}$, and Γ is the gamma function. The function maximized is the log of the expected likelihood or

$$\sum_{t=1}^{T} \ln\left(\sum_{i=0}^{1}\sum_{j=0}^{1} \Pr(S_t=i, S_{t-1}=j|\varphi_{t-1}) L_t^{(i,j)}\right) \tag{15}$$

as in the work of Hamilton (1990).

Estimation Results

The four GARCH/Markov-switching volatility models, the usual GARCH(1, 1) model, and the SWARCH-L model are applied to daily percentage changes in the S&P 500 index from January 6, 1982, to December 31, 1991. Observations from the post-1991 period are reserved for evaluation of the out-of-sample fit.

Some interpretation of the parameter estimates in Table 1 follows. The GARCH-DF model shows switching in the student-t degrees-of-freedom parameter between the values of 2.64 and 8.28. This implies that conditional fourth moments do not exist in one state, whereas the conditional kurtosis is $3(n-2)/(n-4) = 4.4$ in the other state. The weight given to lagged squared residuals in the GARCH process is shown to be $\alpha(1-2v(S_{t-1}))$ in Equation (12), and this weight shifts with the state variable between .009 and .027. In this way, shocks drawn from the low degree-of-freedom state do not affect the persistent GARCH dispersion process proportionately. Most importantly, shifts in the degrees-of-freedom parameter bring potentially large discrete shifts in the variance. A shift out of the low degree-of-freedom state causes the variance to decrease by about 68%, holding the dispersion constant:

$$\left(\frac{\frac{n_h}{n_h-2}}{\frac{n_l}{n_l-2}}\right) = .32.$$

The unconditional probability of being in the low degree-of-freedom state is about 10% with a half-life of five trading days. The unconditional value for the degrees-of-freedom parameter is about 6.8. The GARCH-DF model also suggests that stock returns are negatively skewed because the mean stock return is below normal in the high-volatility state when $S_t = 0$. In fact, all of the models find negative skewness except the conventional GARCH model.

The GARCH-NF model finds an estimate of the variance inflation factor $g\ (S=0) = 12.59$ with a large standard error. The effective sample from which to estimate this parameter is small because the unconditional probability of $S_t = 0$ is only about 1%.

The factor $g\ (S=0)$ raises γ by a significant multiple of 5.7 in the GARCH-UV model, but the unconditional prob-

Table 1. GARCH/Markov-Switching Models Applied to Daily Percentage Changes in S&P 500 Index

Parameter	GARCH-DF	GARCH-NF	GARCH-UV	GARCH-K	GARCH	SWARCH-L
Log-likelihood	−3294.3	−3294.1	−3292.7	−3295.3	−3301.0	−3311.5
μ (S_t = 0)	.0107	−1.333	−.0971	.0158	.0542	.0366
	(.1431)	(1.487)	(.1287)	(.0219)	(.1287)	(.0884)
μ (S_t = 1)	.0619	.0576	.0636	.0803	.0556	.0585
	(.0207)	(.0164)	(.0169)	(.0224)	(.0890)	(.0168)
v (S_t = 0)	.3787	.1478	.1762	.0931	.1860	.1833
	(.0480)	(.0209)	(.0090)	(.0323)	(.0195)	(.0198)
v (S_t = 1)	.1208	.1478	.1762	.2393	.1860	δ = .041
	(.0272)	(.0209)	(.0090)	(.0270)	(.0195)	(.025)
γ	.0105	.0109	.0124	.0233	.0228	.6912
	(.0035)	(.0038)	(.0033)	(.0066)	(.0064)	(.0372)
α	.0360	.0334	.0138	.0328	.0344	α_1 = 1.3E-4
	(.0076)	(.0060)	(.0046)	(.0074)	(.0077)	(.0083)
β	.9466	.9537	.9554	.9307	.9394	α_2 = .0192
	(.0102)	(.0082)	(.0090)	(.0323)	(.0121)	(.0139)
g (S_t = 0)	n.a.	12.59	5.703	n.a.	n.a.	3.782
		(6.414)	(1.882)	n.a.	n.a.	(.4780)
g (S_t = 1)	n.a.	1	1	n.a.	n.a.	1
p	.8544	.7479	.9602	.9978	.9144	.9849
	(.0961)	(.1644)	(.0173)	(.0021)		(.0076)
q	.9842	.9980	.9950	.9986	.9420	.9977
	(.0148)	(.0017)	(.0018)	(.0013)		(.0012)

NOTE: Standard errors are in parentheses.

ability of being in that state is only 11%. The state with g $(S = 1)$ is extremely persistent with $q = .995$.

The GARCH-K model estimates that the degrees-of-freedom parameter switches between 10.7 and 4.2, with an unconditional value of about 6. Both states are highly persistent with nearly identical transition probabilities. Two states for the mean stock return are better defined in the GARCH-K model than in the conventional GARCH model with switching in the mean. Table 1 shows that in the usual GARCH(1, 1) model the mean stock return, μ, is virtually identical in both states. Hence the two states are not well identified and the calculation of standard errors for the transition probabilities failed.

Using daily data, the weights attached to lagged squared residuals are not significant in the SWARCH-L model, with the borderline exception of the leverage-effect parameter, δ. The normalizing factor, g, is estimated to raise the variance by a multiple of 3.78 in the high-volatility state, which has unconditional probability .13. The high degree of persistence of both states suggests that low- and high-volatility states constitute regimes, as opposed to short-lasting episodes. The GARCH-DF model, on the other hand, finds relatively short-lasting low-degree-of-freedom states.

If we were certain that significant state-switching occurred in the mean, then likelihood ratio tests of state-switching in the degrees-of-freedom parameters and g would be appropriate. But, the GARCH model suggests that switching in the mean cannot be taken for granted, so likelihood ratio tests cannot assume that the transition probabilities are identified under the null of no state switching in v or g. Hansen (1992) discussed simulation methods to derive critical values for such likelihood ratio tests with nonstandard distributions. The critical values are computationally burdensome to calculate, however, so I do not pursue that strategy here. Instead, I follow Vlaar and Palm (1993) by using a goodness-of-fit test that is valid for data that are not identically distributed. I perform the test over the in-sample period (1982–1991) and an out-of-sample period (1992–September 1994). I divide the observations into 100 groups based on the probability of observing a value smaller than the actual residual. If the model's time-varying density function fits the data well, these probabilities should be uniformly distributed between 0 and 1. Following Vlaar and Palm (1993),

$$n_i = \sum_{t=1}^{T} I_{it},$$

where

$$\begin{aligned} I_{it} &= 1 \quad \text{if } \frac{(i-1)}{100} < EF(\varepsilon_t, \hat{\theta}) \leq \frac{i}{100} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{16}$$

The expected value of the cumulative density function, F, is taken across the states that might have held at each time. The goodness-of-fit test statistic equals $100/T \sum_{i=1}^{100} (n_i - T/100)^2$ and is distributed χ^2_{99} under the null.

Table 2 provides results from the goodness-of-fit tests. Only the GARCH-DF model is not rejected on an in-sample basis, with a .57 probability value. All six models are rejected out of sample, however.

To examine the source of failure in models other than GARCH-DF in the goodness-of-fit test, Figures 1 and 2 plot the distribution of the in-sample observations across the 100 groups. Figure 1 shows that the GARCH-DF observations are roughly uniformly distributed across the groups, whereas the GARCH-NF observations have a hump-shaped distribution in Figure 2. Too many GARCH-NF residuals are near the center of the cumulative density, which implies that the model's conditional densities are overly peaked—that is, are too leptokurtic. By not allowing the

Table 2. Chi-squared Goodness-of-fit Tests for GARCH/Markov-Switching Models: In-Sample Period: 1982–1991, Out-of-Sample: 1992–September 1994

Model	In-sample	Out-of-sample
GARCH-DF	96.65	152.1
	(.577)	(4.8E-4)
GARCH-NF	193.7	208.1
	(.000)	(.000)
GARCH-UV	136.6	188.4
	(.007)	(.000)
GARCH-K	235.4	294.3
	(.000)	(.000)
GARCH	140.0	228.4
	(.004)	(.000)
SWARCH-L	231.0	307.9
	(.000)	(.000)

NOTE: Probability values are in parentheses.

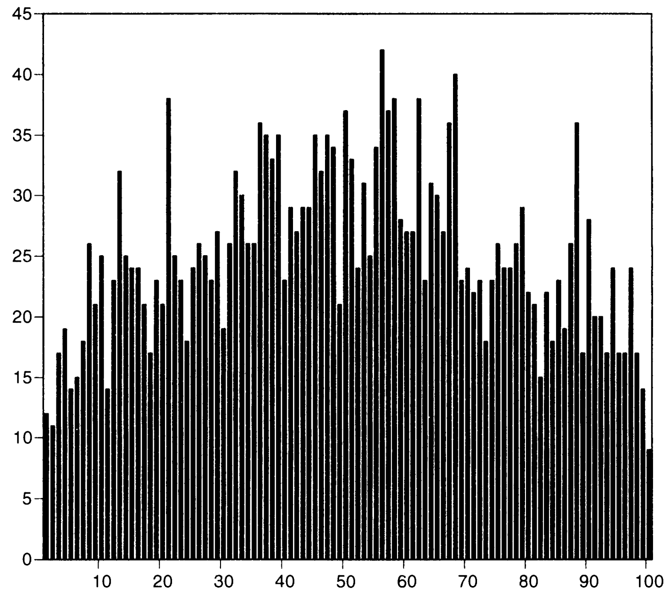

Figure 2. Distribution of GARCH-NF Residuals Into 100 Groups Based on Cumulative Density Function.

conditional kurtosis to change, the GARCH-NF model apparently fits a constant conditional kurtosis that is too high. If time-varying kurtosis is an important feature of stock returns, then it worth studying the distribution of the observations in the GARCH-K model also. Figure 3 shows that the GARCH-K model also provides conditional densities that are too leptokurtic on average, despite its provision for time-varying kurtosis. The reason might be that the GARCH-K model has a very persistent state in which fourth moments do not exist because $q = .9986$. It is possible that the GARCH-K model overstates the persistence of periods of fat-tailed stock-returns distributions: They might be better described as episodes than regimes, as the GARCH-DF model suggests.

3. PREDICTING OPTIONS-IMPLIED VOLATILITIES

As an economic test of the GARCH/Markov-switching models, I use them to predict the next day's opening level of the volatility index (VIX) market compiled by the Chicago Board Options Exchange. The VIX is derived from an options-pricing model and is not a direct observation of market expectations. Nevertheless, many financial-market participants are interested in options-implied volatilities in their own right. The VIX attempts to represent, as closely as possible, the implied volatility on a hypothetical at-the-money option on the Standard & Poor (S&P) 100 with 30 calendar days (22 trading days) to expiration. Details on the construction of the VIX from near-the-money options prices were given by Whaley (1993). The implied volatility on an option reflects beliefs about average volatility over the life of the option. Thus, the constant 22-day horizon of the VIX implies that we must use the GARCH/Markov-switching volatility models to create

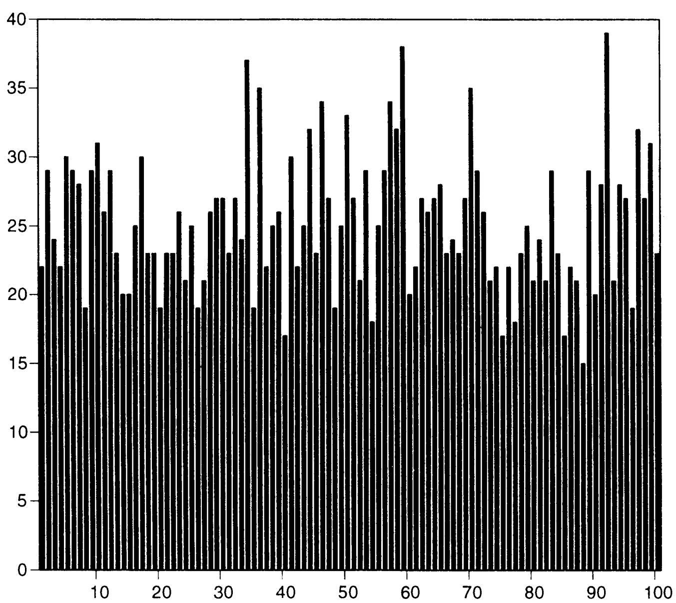

Figure 1. Distribution of GARCH-DF Residuals Into 100 Groups Based on Cumulative Density Function.

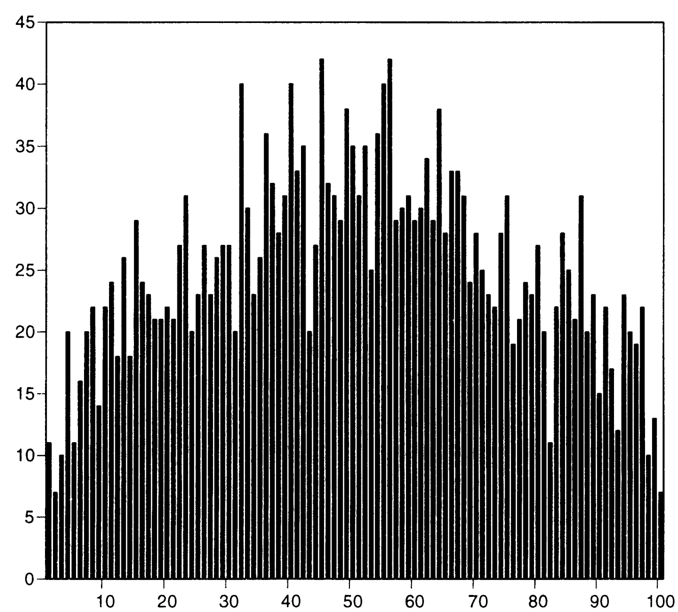

Figure 3. Distribution of GARCH-K Residuals Into 100 Groups Based on Cumulative Density Function.

multiperiod forecasts of volatility for all periods between one and 22 days ahead. In other words, to predict the VIX well, the GARCH/Markov-switching models need to provide good multiperiod forecasts for a full range from 1 to 22 trading days ahead.

Daily data on the VIX were available from 1986–1992. Because the VIX data are based on the S&P 100 and the stock-market data are S&P 500 returns, the mean of the VIX index is slightly higher than the average volatility forecast from the GARCH models. The broader S&P 500 index is somewhat less volatile than the S&P 100. For this reason, I normalize each volatility measure with its 1986–1992 sample mean. Hence a value of 1.5 means that volatility is expected to be one-and-a-half times its normal level in the coming month. Details on the construction of multiperiod forecasts from the GARCH/Markov-switching models are in the Appendix.

I use a minimum-forecast-error variance criterion to measure the closeness of the model-implied and options-implied monthly volatilities. If I denote the options-implied volatility as VIX and the monthly average of the model-predicted volatilities as $\bar{\sigma}$, then the criterion is

$$\frac{1}{T}\sum_{t=1}^{T}(\bar{\sigma}_t - \mathrm{VIX}_t)^2.$$

Note that $\bar{\sigma}_t$ for a Wednesday, for example, is calculated using information available through Tuesday, whereas VIX_t is the data from Wednesday's opening quotes. In this sense, I am using the GARCH/Markov-switching models to predict the options-implied volatilities.

Table 3 shows that only the GARCH-DF and GARCH-K models predict the options-implied volatility index better than the conventional GARCH model and the GARCH-DF model achieves a notable 14% reduction in the forecast-error variance.

Figures 4–6 depict the 22-day average volatility forecasts for all the models and the VIX volatility in the aftermath of the October 1987 stock-market crash. As described by Schwert (1990), for several days after October 19, 1987, options markets became very thin and the options written contained extremely large risk premia—that is, implied

Table 3. Predicting Options-Implied Volatility Index With GARCH/Markov-Switching Models: 1986–1992

Model	Forecast-error variance
GARCH-DF	.0365 (.86)
GARCH-NF	.0548 (1.33)
GARCH-UV	.0589 (1.43)
GARCH-K	.0399 (.97)
GARCH	.0413 (1.00)
SWARCH-L	.0956 (2.31)

NOTE: Size of forecast-error variance relative to GARCH model in parentheses.

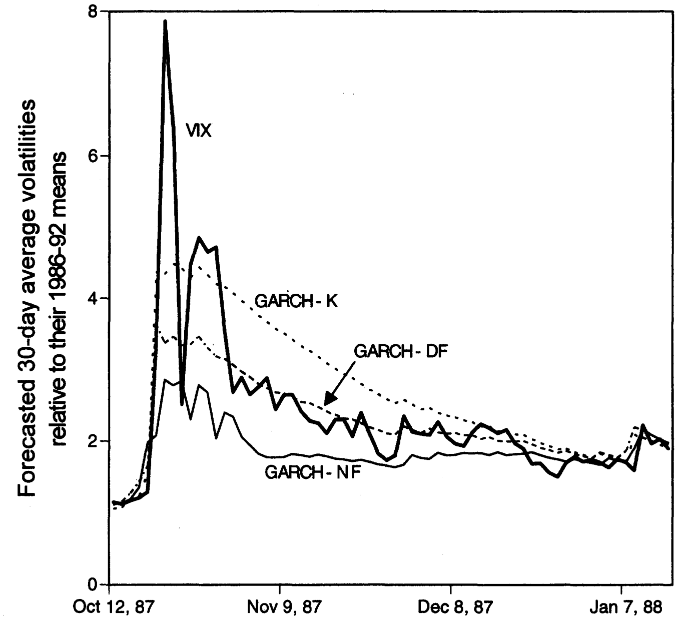

Figure 4. VIX Options-Implied and 3 Model-Implied Volatilities, Autumn 1987.

volatilities. Figures 4–6 show that the VIX reached about eight times its normal level immediately following the crash but returned to less than two times normal by the end of October 1987. The GARCH-DF model best predicts the VIX throughout November and early December 1987. The switch to $n\,(S_t = 0) = 2.64$ led to a downweighting, from .027 to .009, of the lagged squared residuals in the persistent GARCH process. Furthermore, the conditional variance temporarily shifted discretely upward for as long as $n\,(S_t = 0)$ was expected to persist.

The volatility implied by the GARCH-K model in Figure 4, in contrast, overpredicts the VIX for about six weeks, beginning at the end of October 1987. The variance in the

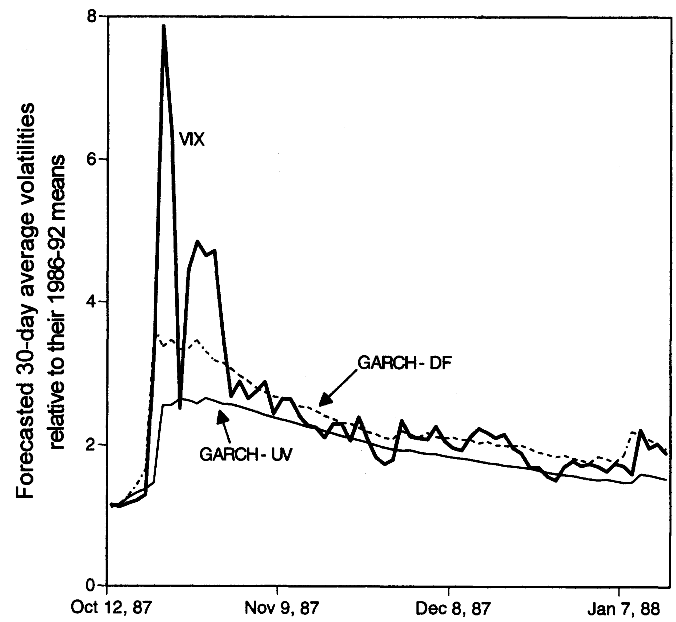

Figure 5. VIX Options-Implied and 2 Model-Implied Volatilities, Autumn 1987.

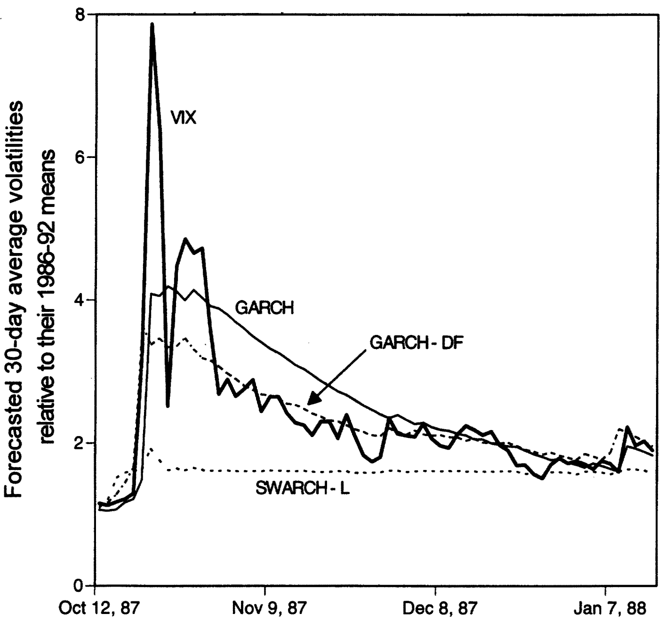

Figure 6. VIX Options-Implied and Model-Implied Volatilities With SWARCH, Autumn 1987.

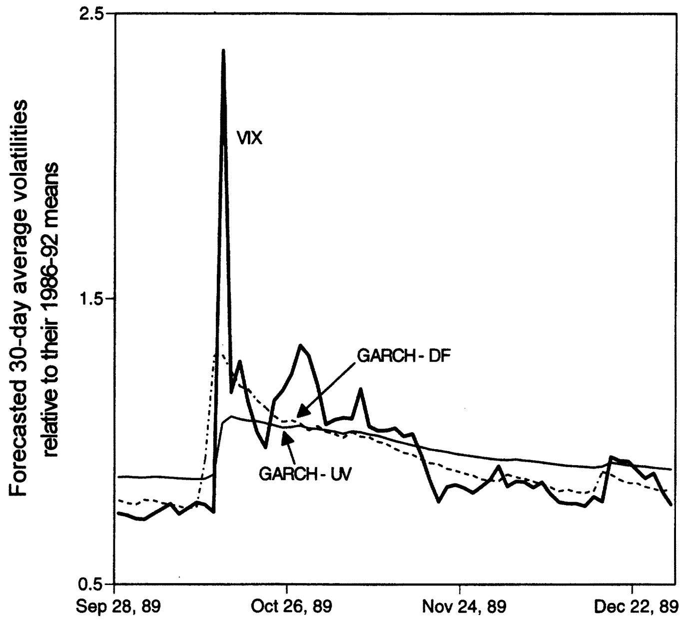

Figure 8. VIX Options-Implied and 2 Model-Implied Volatilities, Autumn 1989.

GARCH-K model is a GARCH process, so it displays the same overpersistence that characterizes the conventional GARCH model, shown in Figure 6. In fact, the forecasts from the conventional GARCH model and the GARCH-K model look very similar. The GARCH-NF model in Figure 1, on the other hand, underpredicts volatility following the crash. The GARCH-NF model quickly switched to the state in which $g\ (S_t = 1) = 12.59$, so the squared residuals were given little weight in the GARCH process and h did not increase much. The variance $\sigma_t^2 = g_t h_t$ did increase with $g = 12.59$, but the increase was never projected to last long with $p = .75$. Consequently, the forecasted average volatility for the month never increased to more than three times the normal level in the GARCH-NF model. In Figure 5, the GARCH-UV model badly underpredicts the VIX in late October 1987 but does fairly well in November and December 1987. The GARCH-UV model estimates a constant and relatively low weight, α, on the lagged squared residuals in the GARCH process, so the conditional variance never increases to more than three times normal, in contrast to the spike in the VIX. In this sense, the GARCH-UV does not necessarily describe the rate of mean reversion in stock-market volatility well because it does not capture the initial volatility spike. In Figure 6, the SWARCH-L model shows a good deal of persistence but does not put enough weight on lagged squared residuals to lift the conditional variance

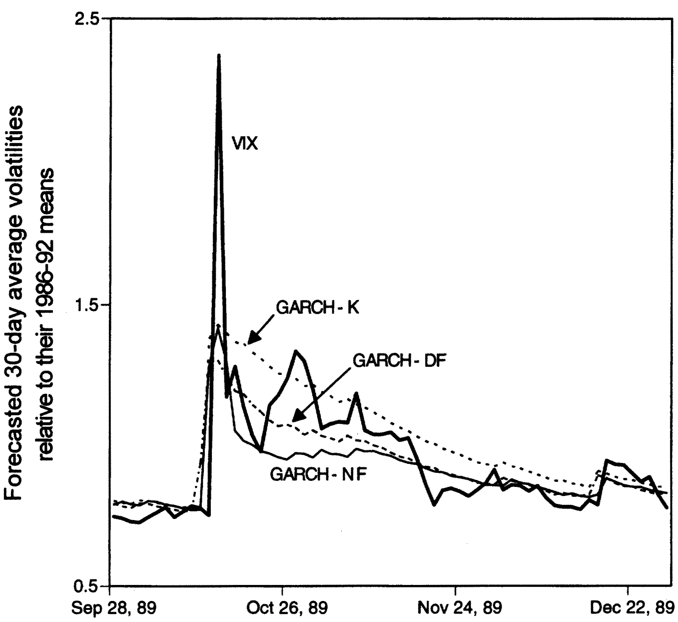

Figure 7. VIX Options-Implied and 3 Model-Implied Volatilities, Autumn 1989.

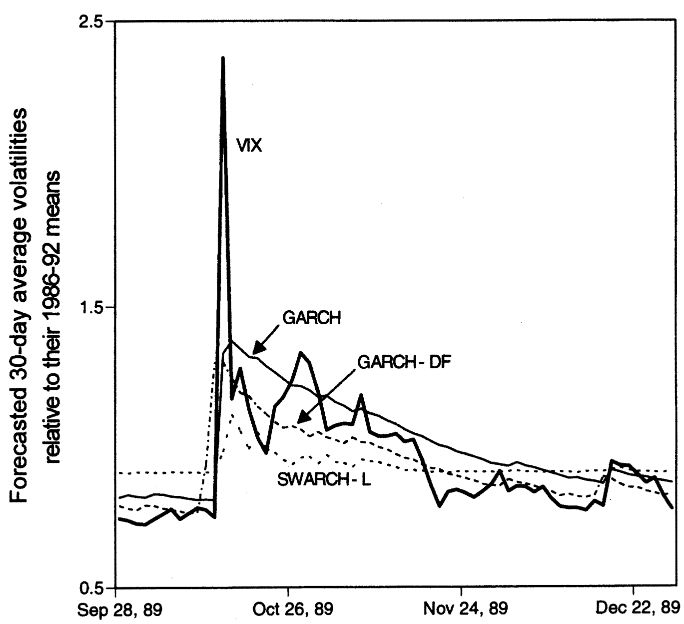

Figure 9. VIX Options-Implied and Model-Implied Volatilities With SWARCH, Autumn 1989.

to the levels necessary to match the spike in the VIX index either.

Figures 7–9 focus on a milder volatility spike in October 1989, when the VIX peaked at about 2.5 times its normal level. Again, the GARCH-DF tends to split through the middle of oscillations of the VIX index better than the other model-implied volatilities, although the improvement is less marked than in Figures 4–6. The same general patterns hold in Figures 7–9 as in Figures 4–6, with the GARCH-K and GARCH models tending to overpredict volatility and the GARCH-UV and SWARCH-L models showing persistence but failing to yield sufficiently dramatic initial increases in volatility.

4. CONCLUSIONS

This article introduces a tractable framework for adding Markov-switching parameters to conditional-variance models. Four different specifications of Markov-switching volatility models are estimated, and the addition of Markov-switching parameters is found to have a variety of effects on the behavior of the conditional volatility, relative to the model without switching. The specification found to predict options-implied expectations of stock-market volatility best is the one in which the student-t degrees-of-freedom parameter switches so as to induce substantial discrete shifts in the conditional variance. This model allows for two sources of mean reversion in the wake of large shocks that are not available in a standard model: A switch out of the fat-tailed state is estimated to induce a 68% decrease in volatility for a given level of dispersion, and the weight given to the most recent shock decreases by two-thirds when the fat-tailed state pertains, thereby reducing the influence and persistence of large shocks.

Another novel feature of this model is that it relates stock returns to the degree of leptokurtosis in the conditional-returns distribution. Traditional models, in contrast, assume constant conditional kurtosis and relate expected returns to the conditional variance. The point estimates support the hypothesis that stock returns are generally lower in the more fat-tailed state.

I also draw economically relevant comparisons between the behaviors of options-implied volatilities and the conditional variances from the volatility models studied. Because options-implied volatilities serve as useful proxies for market expectations of volatility, it is interesting to observe that the conditional variance from one of the switching-in-the-variance models reverts to normal about as quickly as the options-implied volatility following large shocks, such as the stock-market crash of October 1987. The conventional volatility model, in contrast, has a conditional variance that remains above normal with considerably greater persistence. Thus, Markov switching in the variance is shown to add a realistic degree of mean reversion to the conditional variance. In addition, the description of time-varying stock-return skewness and kurtosis provided by these models could prove useful in analyzing options prices on the S&P 500 index.

An interesting extension would be to model the transition probabilities of the Markov process as time-varying functions of conditioning variables to test whether transitions into and out of fat-tailed states could be better predicted using more information.

ACKNOWLEDGMENTS

The content is my responsibility and does not represent official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. I thank James Hamilton and Bruce Hansen for sharing their programs and Barbara Ostdiek and Robert Whaley for the VIX data.

APPENDIX: ON MULTIPERIOD VOLATILITY FORECASTS

Forecasts of the volatility m periods ahead are based on the well-known relationship between GARCH models and autoregressive moving average (ARMA) representations of the squared disturbances. A GARCH(1, 1) process,

$$h_t = \gamma + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{A.1}$$

implies that the squared residuals obey an ARMA(1, 1) process,

$$\varepsilon_t^2 = \gamma + (\alpha+\beta)\varepsilon_{t-1}^2 - \beta(\varepsilon_{t-1}^2 - h_{t-1}) + (\varepsilon_t^2 - h_t), \tag{A.2}$$

where $\varepsilon_t^2 - h_t$ is a mean zero error that is uncorrelated with past information. In forecasting the squared residuals m periods ahead with the GARCH-DF model, for example, I define $H_t = \varepsilon_t^2(1-2v_t)$. In this case H has an ARMA(1, 1) representation,

$$H_t = \gamma + (\alpha+\beta)H_{t-1} - \beta(H_{t-1} - \hat{h}_{t-1}) + (H_t - h_t), \tag{A.3}$$

where

$$E_t[H_{t+1}|H_t] = h_{t+1} = \gamma + \alpha H_t + \beta\hat{h}_t. \tag{A.4}$$

Because the sample size is large, longer-range forecasts can be built from the asymptotic forecasting equation for first-order autoregressive processes so that, for $m > 1$,

$$E_t[H_{t+m}|H_t] = (\alpha+\beta)^{m-1}h_{t+1} + [1-(\alpha+\beta)^{m-1}]\frac{\gamma}{1-\alpha-\beta}. \tag{A.5}$$

It remains to integrate out the unobserved states:

$$E_t\varepsilon_{t+m}^2 = \sum_{i=0}^{1}\sum_{j=0}^{1}\Pr(S_{t+m}=i, S_t=j|\varphi_t) \times E_t[H_{t+m}|S_t=j]\frac{1}{1-2v_{t+m}^{(i)}}, \tag{A.6}$$

where $H_t\ (S_t = j) = (\varepsilon_t^{(j)})^2(1-2v_t^{(j)})$. The expected average variance over the next 22 trading days is then taken as

$$\overline{\sigma_t^2} = \frac{1}{22}\sum_{m=1}^{22}E_t\varepsilon_{t+m}^2. \tag{A.7}$$

Similar forecasts are drawn for the other models with H defined such that $H_t = \varepsilon_t^2 / g_t$ in the GARCH-NF model and $H_t = \varepsilon_t^2 - \gamma_t$ in the GARCH-UV model.

For the SWARCH-L model, the multiperiod forecasts are derived by recursive substitution as in the work of Hamilton and Susmel (1994).

[Received June 1994. Revised September 1995.]

REFERENCES

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307–327.

Cai, J. (1994), "A Markov Model of Switching-Regime ARCH," *Journal of Business & Economic Statistics*, 12, 309–316.

Friedman, B. M., and Laibson, D. I. (1989), "Economic Implications of Extraordinary Movements in Stock Prices," *Brookings Papers on Economic Activity*, 2/1989, 137–189.

Hamilton, J. D. (1990), "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39–70.

Hamilton, J. D., and Susmel, R. (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333.

Hansen, B. E. (1992), "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

——— (1994), "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

Kim, C. J. (1994), "Dynamic Linear Models With Markov Switching," *Journal of Econometrics*, 64, 1–22.

Lamoureux, C., and Lastrapes, W. (1990), "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business & Economic Statistics*, 8, 225–234.

——— (1993), "Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities," *Review of Financial Studies*, 5, 293–326.

Schwert, G. W. (1990), "Stock Volatility and the Crash of '87," *Review of Financial Studies*, 3, 77–102.

Vlaar, P. J., and Palm, F. C. (1993), "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps," *Journal of Business & Economic Statistics*, 11, 351–360.

Whaley, R. D. (1993), "Derivatives on Market Volatility: Hedging Tools Long Overdue," *Journal of Derivatives*, 1, 71–84.

LINKED CITATIONS

You have printed the following article:

Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility
Michael J. Dueker
Journal of Business & Economic Statistics, Vol. 15, No. 1. (Jan., 1997), pp. 26-34.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199701%2915%3A1%3C26%3AMSIGPA%3E2.0.CO%3B2-I

This article references the following linked citations:

References

A Markov Model of Switching-Regime ARCH
Jun Cai
Journal of Business & Economic Statistics, Vol. 12, No. 3. (Jul., 1994), pp. 309-316.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199407%2912%3A3%3C309%3AAMMOSA%3E2.0.CO%3B2-Y

Economic Implications of Extraordinary Movements in Stock Prices
Benjamin M. Friedman; David I. Laibson; Hyman P. Minsky
Brookings Papers on Economic Activity, Vol. 1989, No. 2. (1989), pp. 137-189.
Stable URL:
http://links.jstor.org/sici?sici=0007-2303%281989%291989%3A2%3C137%3AEIOEMI%3E2.0.CO%3B2-T

The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP
B. E. Hansen
Journal of Applied Econometrics, Vol. 7, Supplement: Special Issue on Nonlinear Dynamics and Econometrics. (Dec., 1992), pp. S61-S82.
Stable URL:
http://links.jstor.org/sici?sici=0883-7252%28199212%297%3CS61%3ATLRTUN%3E2.0.CO%3B2-8

Autoregressive Conditional Density Estimation
Bruce E. Hansen
International Economic Review, Vol. 35, No. 3. (Aug., 1994), pp. 705-730.
Stable URL:
http://links.jstor.org/sici?sici=0020-6598%28199408%2935%3A3%3C705%3AACDE%3E2.0.CO%3B2-P

LINKED CITATIONS

Persistence in Variance, Structural Change, and the GARCH Model

Christopher G. Lamoureux; William D. Lastrapes

Journal of Business & Economic Statistics, Vol. 8, No. 2. (Apr., 1990), pp. 225-234.

Stable URL:

http://links.jstor.org/sici?sici=0735-0015%28199004%298%3A2%3C225%3APIVSCA%3E2.0.CO%3B2-P

Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities

Christopher G. Lamoureux; William D. Lastrapes

The Review of Financial Studies, Vol. 6, No. 2. (1993), pp. 293-326.

Stable URL:

http://links.jstor.org/sici?sici=0893-9454%281993%296%3A2%3C293%3AFSVTAU%3E2.0.CO%3B2-0

Stock Volatility and the Crash of '87

G. William Schwert

The Review of Financial Studies, Vol. 3, No. 1, National Bureau of Economic Research Conference: Stock Market Volatility and the Crash, Dorado Beach, March 16-18, 1989. (1990), pp. 77-102.

Stable URL:

http://links.jstor.org/sici?sici=0893-9454%281990%293%3A1%3C77%3ASVATCO%3E2.0.CO%3B2-Z

The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps

Peter J. G. Vlaar; Franz C. Palm

Journal of Business & Economic Statistics, Vol. 11, No. 3. (Jul., 1993), pp. 351-360.

Stable URL:

http://links.jstor.org/sici?sici=0735-0015%28199307%2911%3A3%3C351%3ATMIWER%3E2.0.CO%3B2-T