Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Which GARCH Model for Option Valuation?

Peter Christoffersen, Kris Jacobs,

To cite this article:
Peter Christoffersen, Kris Jacobs, (2004) Which GARCH Model for Option Valuation?. Management Science 50(9):1204-1221.
https://doi.org/10.1287/mnsc.1040.0276

MANAGEMENT SCIENCE
Vol. 50, No. 9, September 2004, pp. 1204–1221
ISSN 0025-1909 | EISSN 1526-5501 | 04 | 5009 | 1204

DOI 10.1287/mnsc.1040.0276

Which GARCH Model for Option Valuation?

Peter Christoffersen, Kris Jacobs
Faculty of Management, McGill University, Montréal, Québec, Canada, and CIRANO
{peter.christoffersen@mcgill.ca, kris.jacobs@mcgill.ca}

Characterizing asset return dynamics using volatility models is an important part of empirical finance. The existing literature on GARCH models favors some rather complex volatility specifications whose relative performance is usually assessed through their likelihood based on a time series of asset returns. This paper compares a range of GARCH models along a different dimension, using option prices and returns under the risk-neutral as well as the physical probability measure. We judge the relative performance of various models by evaluating an objective function based on option prices. In contrast with returns-based inference, we find that our option-based objective function favors a relatively parsimonious model. Specifically, when evaluated out-of-sample, our analysis favors a model that, besides volatility clustering, only allows for a standard leverage effect.

Key words: option pricing; GARCH; risk-neutral pricing; parsimony; forecasting; out-of-sample
History: Accepted by Ravi Jagannathan, finance; received December 5, 2002. This paper was with the authors $9\frac{1}{2}$ months for 2 revisions.

1. Introduction

The family of GARCH volatility models has become an important tool kit in empirical asset pricing and financial risk management. Following the work of Engle (1982) and Bollerslev (1986), a voluminous econometric literature has developed on volatility estimation and forecasting.[1] Important papers in empirical finance that apply GARCH models include Bollerslev et al. (1988), Campbell and Hentschel (1992), French et al. (1987), Glosten et al. (1993), Maheu and McCurdy (2004), Pagan and Schwert (1990), and Schwert (1989).

Motivated equally by the success of GARCH models in fitting asset returns and by the failure of deterministic volatility models in fitting option prices,[2] researchers have extended the GARCH model into the domain of option valuation. Duan (1995) characterizes the transition between the physical and risk-neutral probability distributions if the dynamic of the primitive security is given by a GARCH process, and thus establishes the foundation for the valuation of European options. Theoretical aspects of hedging in the GARCH option-pricing model are further discussed in Garcia and Renault (1998), and Ritchken and Trevor (1999) construct trinomial trees to price American options under GARCH. Duan et al. (2004) extend the standard GARCH option-valuation model to include jumps.

A few papers have investigated certain aspects of the empirical performance of specific GARCH option-pricing models, including Amin and Ng (1993), Engle and Mustafa (1992), Duan (1996), Hardle and Hafner (2000), and Heston and Nandi (2000). While these papers make important contributions, many issues remain unexplored. This paper attempts to fill two gaps in this literature.

First, the literature that estimates GARCH processes using time series on asset returns contains a wealth of evidence on specifications that generalize the fairly simple GARCH processes that are used in the option-valuation literature. It seems natural that the large amount of work researchers have put into modeling the volatility of foreign exchange, bond, equity, and index returns should also be relevant for fitting the prices of options written on these underlying assets. Curiously, however, no paper so far has assessed the relative performance of different GARCH models in valuing options across strike prices and times to maturity, while estimating parameters using information on option prices. Doing so, we apply the GARCH specifications of Ding et al. (1993) and Hentschel (1995). These specifications summarize the differences between different GARCH models in terms of differences in the news impact function (Pagan and Schwert 1990, Engle and Ng 1993). Through judicious choice of the specification of the news impact function, these models manage to nest

[1] See the overview in Bollerslev et al. (1992).

[2] Motivated by the poor empirical performance of the seminal Black and Scholes (1973) model, Dupire (1994) and Derman and Kani (1994) have proposed deterministic volatility models whose performance in turn has been questioned by Dumas et al. (1998). Dumas et al. found that option prices from a simple OLS regression of implied volatility on a polynomial in strike price and maturity (the so-called Practitioner's Black-Scholes) outperformed models whose diffusion term is a deterministic function of the strike price and maturity.

a number of existing GARCH specifications. Adopting the nesting specifications therefore allows us to investigate the pricing performance of a number of existing models while keeping the empirical analysis manageable.

Second, not much empirical work has been done estimating GARCH parameters by combining information on option prices and return dynamics. Only a few papers estimate implied GARCH parameters from option prices under the risk-neutral probability measure. Amin and Ng (1993) and Hardle and Hafner (2000) estimate GARCH parameters using time series of returns under the physical distribution and subsequently use these estimates to compute option prices. Duan (1996) estimates GARCH parameters under the risk-neutral distribution using a single day of option prices and investigates the potential of this model to explain the smile, and Engle and Mustafa (1992) repeatedly estimate such GARCH parameters using one day of options at a time. The approach in this paper is closest to that of Heston and Nandi (2000), who estimate a GARCH model simultaneously using many days of options and data on the underlying asset returns. Heston and Nandi (2000) formulate a particular GARCH specification that yields an analytical solution and provide an empirical analysis of the model. They convincingly demonstrate that the inclusion of a leverage effect as well as volatility clustering are of great importance in improving valuation performance.[3] It must also be noted that an important literature exists that estimates continuous-time stochastic volatility models for option valuation.[4] This paper does not compare the option-valuation performance of discrete-time GARCH models with continuous-time stochastic volatility models. Doing so is an important topic for future research.

When estimating and evaluating the valuation of different GARCH option-pricing models, we get surprising results. While most of the GARCH literature that investigates returns data using maximum likelihood favors relatively complex models, we find that for the purpose of option valuation, one should not look beyond a simple GARCH model that allows for volatility clustering and a leverage effect.[5] While the more elaborate models we investigate achieve a better fit when estimated using returns data and when evaluated on option prices in-sample, they perform worse than the parsimonious model out-of-sample. Moreover, the improvement in in-sample fit is small in some cases, even though the extra parameters are usually significantly estimated. We conclude that in our application relying on statistical measures of significance for parameter estimation and model selection can be highly misleading if the true object of interest is a more economically based criterion such as option valuation.

The remainder of the paper is organized as follows. In §2 we discuss the estimation methodology under the physical as well as the risk-neutral measure. Section 3 introduces the data set, §4 presents the empirical results, §5 explores the results further and assesses the physical and risk-neutral densities implied by models and data. Section 6 concludes and points out directions for future work.

2. Methodology

In this section we discuss the different methodologies used to estimate parameters for the option-valuation models. We first outline the methodology used to estimate the parameters of the volatility process under the physical probability measure. This methodology exclusively uses the returns on the underlying asset. Subsequently, we discuss the estimation of the parameters of the volatility process under the risk-neutral probability measure. We estimate the risk-neutral measure by combining data on returns with data on option prices. To clarify the relationship between the physical and risk-neutral probability measures, we indicate the relationship between the parameters obtained under the physical and risk-neutral volatility dynamics.

2.1. Estimation Using Asset Returns Only

We investigate the standard class of GARCH models pioneered by Engle (1982) and Bollerslev (1986). GARCH models are discrete-time models that have been used to estimate a variety of financial time series such as stock returns, interest rates, and foreign exchange rates. See Bollerslev et al. (1994) for an overview of GARCH models and Bollerslev et al. (1992) and Campbell et al. (1997) for an overview of the use of GARCH models for financial time series. We assume that the logarithm of stock returns under the physical probability measure P follows the dynamic

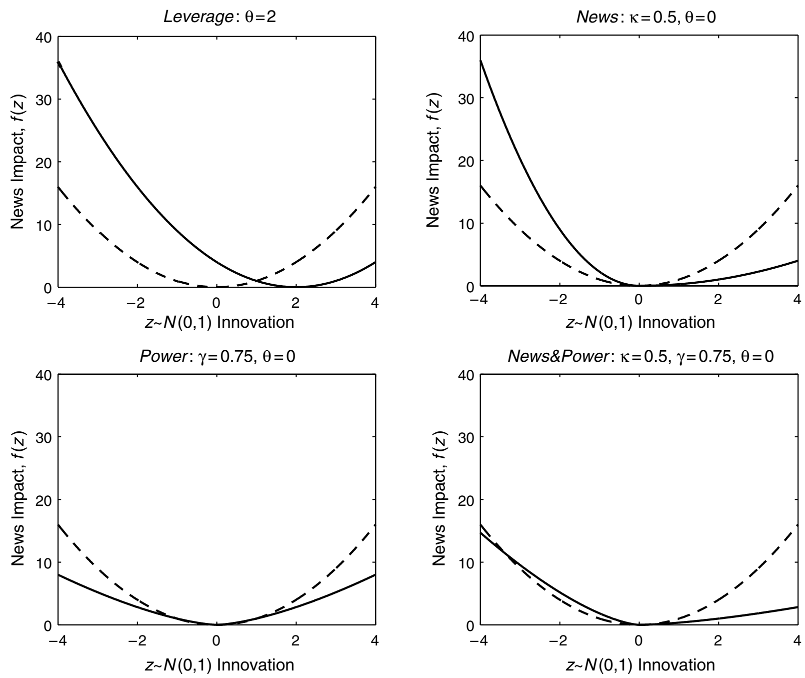
$$\ln(S_t/S_{t-1}) \equiv R_t = r + \lambda\sqrt{h_t} - \tfrac{1}{2}h_t + \sqrt{h_t}z_t, \quad z_t \sim N(0,1), \tag{1}$$

[3] Several studies have also investigated the relationship between volatility implied in option prices and the conditional volatility in GARCH models estimated from returns. See Day and Lewis (1992), Jorion (1995), and Lamoureux and Lastrapes (1993). Engle et al. (1993) and Noh et al. (1994) evaluate the usefulness of GARCH forecasts for trading profits.

[4] See, for example, Andersen et al. (2002), Bakshi et al. (1997), Benzoni (1998), Chernov and Ghysels (2000), Bates (1996), Eraker (2004), Jiang (1998), Jones (2003), and Pan (2002).

[5] As always in empirical work, our findings are conditional on the data set we use. We analyze a large and very standard data set of S&P 500 index options with up to one year of maturity. Longer-maturity options such as the LEAPS investigated by, for example, Bollerslev and Mikkelsen (1996, 1999), may suggest the use of more complex models such as long-memory GARCH models.

Figure 1 Stylized News Impact Functions, $f(z)$: Various Models



Notes. The innovation function from each model (solid line) is superimposed on the symmetric squared innovation function from the *Simple* model (dashed line). The innovation function conveys the impact on volatility from a particular standard normal innovation, which is given on the horizontal axis.

where r is the risk-free rate, and λ is the price of risk. Notice that it follows that the conditional expectation of gross returns is

$$E^P[\exp(R_t) \mid \Omega_{t-1}] = \exp(r + \lambda\sqrt{h_t}). \quad (2)$$

The specification of returns in (1) is common to all GARCH models we investigate. The differences between the models concern the specification of the volatility dynamic h_t. Ding et al. (1993) and Hentschel (1995) provide very general specifications of the volatility dynamic that nest most existing work. Motivated by these studies, we first write the volatility dynamic h_t as follows[6]

$$h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 h_{t-1} f(z_{t-1}) \quad (3)$$

where $z_t \sim N(0, 1)$. Different GARCH models are mainly characterized by differences in the innovation functions f.

We consider the following specifications of f

$$\begin{aligned} \textit{Simple}:\quad & f(z_{t-1}) = z_{t-1}^2 \\ \textit{Leverage}:\quad & f(z_{t-1}) = (z_{t-1} - \theta)^2 \\ \textit{News}:\quad & f(z_{t-1}) = \{|z_{t-1} - \theta| - \kappa(z_{t-1} - \theta)\}^2 \\ \textit{Power}:\quad & f(z_{t-1}) = (z_{t-1} - \theta)^{2\gamma} \\ \textit{News\&Power}:\quad & f(z_{t-1}) = \{|z_{t-1} - \theta| - \kappa(z_{t-1} - \theta)\}^{2\gamma}. \end{aligned}$$

[6] The vast majority of papers in the literature find little or no support for higher-order GARCH models; thus, we restrict attention to first-order models here.

We then generalize to allow for nonlinear volatility dynamics as follows

$$\textit{Box-Cox}: h_t^{\psi} = \beta_0 + \beta_1 h_{t-1}^{\psi} + \beta_2 h_{t-1}^{\psi} f(z_{t-1}), \quad \text{with } f(z_{t-1}) = (z_{t-1} - \theta)^{2\psi}. \quad (4)$$

Notice that the *News&Power* model nests the preceding models, while the *News*, *Power*, and *Box-Cox* models nest the *Leverage* and *Simple* models. To appreciate the importance of the innovation function, Figure 1 plots four of these innovation functions for certain (arbitrarily chosen) parameter combinations. All panels plot the innovation function $f(z_{t-1})$ as a function of the i.i.d. "shock" z_{t-1}. In every panel the broken line depicts the symmetric innovation function associated with the *Simple* model, and the solid line depicts the innovation function for the alternative model, which has extra parameters. It can be seen that the *Leverage* parameter θ "shifts" the innovation function, the *News* parameter κ "tilts" the innovation function, and the *Power* parameters γ and ψ flatten or steepen the innovation function. Similar effects are described in Pagan and Schwert (1990), Engle and Ng (1993), Ding et al. (1993), and Hentschel (1995).

The persistence of the models under consideration can be written as

$$\textit{Persistence} = \beta_1 + \beta_2 E\{f(z_{t-1})\}, \quad (5)$$

where $E\{f(z_{t-1})\}$ can be calculated analytically for the *Simple* model to be 1, for the *Leverage* model to be $1+\theta^2$, and for the *News* model to be

$$(1+\theta^2)(1+\kappa^2)+2\kappa[2\phi(\theta)+(1+\theta^2)(2\Phi(\theta)-1)],$$

where $\phi(\theta)$ and $\Phi(\theta)$ denote the probability density function and the cumulative distribution function for the standard normal. For the other three models, *Power*, *News&Power*, and *Box-Cox*, the persistence measure can be computed easily using simulation. Notice that in the *Box-Cox* model, persistence is with respect to the power of volatility, h_t^{ψ}, rather than of volatility itself. It is the case for all models that restricting the persistence to be below unity guarantees covariance stationarity of the model.

One way to price options under GARCH dynamics for returns is to proceed in two steps. In a first step, one estimates the parameters of (1) and (3) under the physical probability measure using asset returns. In a second step, one maps these parameters into the parameters for the risk-neutral distribution. For GARCH models, this approach is followed, for instance, by Amin and Ng (1993), Bollerslev and Mikkelsen (1996, 1999), and Hardle and Hafner (2000). This second step is described in more detail below. To implement the first step, a convenient approach is to estimate the model parameters using maximum likelihood. We maximize the following log-likelihood function (conditioning on the first observation)

$$\ln L = -(T-1)\ln(2\pi)/2 - \sum_{t=2}^{T}\ln(h_t)\Big/2 - \sum_{t=2}^{T}\left(R_t - r - \lambda\sqrt{h_t} + \tfrac{1}{2}h_t\right)^2\Big/(2h_t) \qquad (6)$$

where h_t denotes the volatility dynamic (3).

To use the parameters obtained under the physical probability measure for option valuation, we have to be explicit about the relationship between the physical and risk-neutral dynamic. This relationship amounts to a choice of pricing kernel or, equivalently, a choice of the utility function of the representative agent. In a GARCH context, Duan (1995), building on the work of Rubinstein (1976) and Brennan (1979), provides a locally risk-neutral valuation relationship (*LRNVR*), which is satisfied by a measure Q if

$$E^Q[\exp(R_t)\mid\Omega_{t-1}] = \exp(r), \qquad (7)$$

and

$$\mathrm{Var}^Q[R_t\mid\Omega_{t-1}] = \mathrm{Var}^P[R_t\mid\Omega_{t-1}] = h_t. \qquad (8)$$

The *LRNVR* implies that under the risk-neutral measure Q, the return process evolves according to

$$R_t = r - \tfrac{1}{2}h_t + \sqrt{h_t}z_t^*, \qquad z_t^* \sim N(0,1). \qquad (9)$$

The volatility process (3) becomes, under the risk-neutral measure,[7]

$$h_t = \beta_0 + \beta_1 h_{t-1} + \beta_2 h_{t-1} f(z_{t-1}^* - \lambda). \qquad (10)$$

To provide more insight into this result, notice that solving for z_{t-1}^* from the risk-neutral dynamic (9) yields

$$z_{t-1}^* = \left(R_{t-1} - r + \tfrac{1}{2}h_{t-1}\right)\Big/\sqrt{h_{t-1}}, \qquad (11)$$

whereas solving for z_{t-1} from the physical dynamic (1) we get

$$\begin{aligned} z_{t-1} &= \left(R_{t-1} - r - \lambda\sqrt{h_{t-1}} + \tfrac{1}{2}h_{t-1}\right)\Big/\sqrt{h_{t-1}} \\ &= \left[\left(R_{t-1} - r + \tfrac{1}{2}h_{t-1}\right)\Big/\sqrt{h_{t-1}}\right] - \lambda. \end{aligned} \qquad (12)$$

Therefore, it is clear from comparing (11) and (12) that for a general innovation function f, we have

$$f(z_{t-1}^* - \lambda) = f(z_{t-1}), \quad \forall f. \qquad (13)$$

Given the risk-neutral dynamic in (9) and (10), the price of a European call option can be computed as

$$C(h_t) = \exp(-r(T-t))E^Q[\max(S_T - K, 0)\mid\Omega_{t-1}]. \qquad (14)$$

However, we do not have an analytical expression for this option price. Because the multiperiod distribution of the GARCH process is unknown, we need to compute the conditional expectation using Monte Carlo simulation.[8]

As mentioned above, we can implement the option-valuation models in two steps. In a first step, we obtain estimates under the physical probability measure using a time series of returns as described above. In a second step, we use these parameters in the risk-neutral dynamic (10) and compute the option price using (14). The fit of a volatility model can then be assessed using standard loss functions. Following much of the existing literature, we use mean-squared dollar errors ($MSE),[9]

$$\$MSE = \frac{1}{N}\sum_i (C_i - C_i(h_t))^2, \qquad (15)$$

where C_i is the market price of option i, $C_i(h_t)$ is the model price for option i, and N is the total number of contracts in the sample. The sample can consist of a number of contracts on a given day, or a number of contracts on different days.

[7] This specification only covers the *Simple*, *Leverage*, *News*, *Power*, and *News&Power* models. The transformation for the *Box-Cox* model in (4) is similar.

[8] The details of the Monte Carlo simulation are discussed in the online appendix (available at http://mansci.pubs.informs.org/ecompanion.html).

[9] The choice of loss function is an important and often ignored topic. See Christoffersen and Jacobs (2004) and Renault (1997).

2.2. Estimation Using Option Prices and Asset Returns

In this section we describe an alternative approach to estimating the parameters of the volatility process and testing the option-valuation models. The approach described above, which estimates the parameters under the physical probability measure, is conceptually straightforward and computationally easy. However, for the purpose of option valuation, it may be preferable to estimate the parameters directly using (a different set of) option prices. One possible alternative is to use just one day of option prices to estimate the parameters. For instance, if we use a mean-squared dollar objective function, we can obtain model parameters by minimizing

$$\$MSE = \frac{1}{N_t}\sum_i (C_{i,t} - C_{i,t}(h_t))^2, \tag{16}$$

where N_t is the number of option prices present in the sample at time t. Alternatively, we can use more than one day of option prices to estimate the parameters. The objective, again, is to minimize the sum of squared option-valuation errors in

$$\$MSE = \frac{1}{N_T}\sum_{t,i} (C_{i,t} - C_{i,t}(h_t))^2, \tag{17}$$

where $N_T = \sum_{t=1}^{T} N_t$ and T is the total number of days included in the sample. In this case the implementation requires linking the volatility on different dates using the time series of stock returns.[10]

We have described two approaches to estimating parameters for use in option-valuation models. The first approach consists of using the time series of asset returns to estimate the parameters under the physical probability measure. Subsequently, these parameters are mapped into the risk-neutral parameters, and plugged into the option-valuation formulas. The second approach is to estimate the risk-neutral parameters minimizing the option-pricing errors directly. If we use just one day of option prices, this approach is relatively straightforward. If we use multiple days, we have to use a volatility updating rule.

A priori, we would expect the approach that estimates the risk-neutral parameters directly from option prices to work better than the approach based on the time series of asset returns described in the previous section, for several reasons. First, option prices contain forward-looking information over and beyond historical returns, and thus using option prices to find parameters can have an important advantage simply from the perspective of the data used. Second, when using maximum likelihood to estimate parameters under the physical measure, it is clear that the loss function is quite different from an out-of-sample loss function, which could be something like the mean-squared dollar errors in (15).

3. Data

We conduct our empirical analysis using four years of data on S&P 500 call options. First, the three-year period between June 1, 1988, and May 31, 1991, which we denote Sample A, is used exclusively for in-sample estimation. This sample closely corresponds to the data used by Bakshi et al. (1997).[11] We subsequently use a fourth year of data covering the period from June 1, 1991, to May 31, 1992, which we refer to as Sample B. We apply several filters to Sample B that are identical to the ones used in Bakshi et al. (1997), and we refer the reader to that study for the details.

In both samples, we only use options data for Wednesdays.[12] If Wednesday is a holiday, we use the next trading day. Using only Wednesday data allows us to study a fairly long time series, which is useful considering the highly persistent volatility processes. An additional motivation for only using Wednesday data is that, following the work of Dumas et al. (1998), several studies have used this setup (see, for instance, Heston and Nandi 2000). For our empirical work, this data selection criterion leaves us with 156 days of options data in the June 1, 1988, to May 31, 1991, period, and 52 days of options data in the June 1, 1991, to May 31, 1992, period.

Table 1 presents the number of contracts used in the empirical work for Sample A and B by moneyness and maturity. It can be seen that the relative importance of the different cells is fairly similar across periods. Tables 2 and 3 present average prices and implied volatilities for both sample periods. Given the differences between option prices in different cells in Table 2, it is clear that different options will receive different weights when using the mean-squared dollar objective function (15). Table 3 indicates that in both sample periods we find implied volatility patterns across maturity and moneyness that are comparable to those found in other studies. In-the-money calls (and therefore also out-of-the-money puts) are expensive relative to the Black-Scholes model. Figure 2 presents a plot of the average implied volatilities extracted from the options on a week-by-week basis. It can be seen that

[10] This procedure is described in detail in the online appendix.

[11] The three years of data were graciously provided to us by Gurdip Bakshi.

[12] This choice is to some extent motivated by constraints. Because we cannot compute option prices analytically, the optimization problems are fairly time-intensive, and limiting the number of options reduces the computational burden.

Table 1 Number of Contracts Across Moneyness and Maturity

	Sample A					Sample B			
	DTM < 60	60 < DTM < 180	180 < DTM	Total		DTM < 60	60 < DTM < 180	180 < DTM	Total
S/X < 0.94	164	641	515	1,320	S/X < 0.94	49	222	173	444
0.94 < S/X < 0.97	436	413	207	1,056	0.94 < S/X < 0.97	170	201	104	475
0.97 < S/X < 1.00	558	400	202	1,160	0.97 < S/X < 1.00	233	193	82	508
1.00 < S/X < 1.03	523	357	172	1,052	1.00 < S/X < 1.03	214	187	97	498
1.03 < S/X < 1.06	474	312	137	923	1.03 < S/X < 1.06	199	160	63	422
1.06 < S/X	1,015	985	570	2,570	1.06 < S/X	346	332	179	857
Total	3,170	3,108	1,803	8,081	Total	1,211	1,295	698	3,204

Notes. We report descriptive statistics on our two subsamples of call options on the S&P 500 index quoted at the closing of every Wednesday. Sample A denotes June 1, 1988–May 31, 1991; and Sample B denotes June 1, 1991–May 31, 1992. Each statistic is reported for three maturity bins and six moneyness bins.

implied volatility varies considerably over time. Interestingly, the level of implied volatility in Sample B is lower than in Sample A. Also, it is evident from Figure 2 that there is substantial clustering in implied volatilities.[13]

4. Empirical Results

In this section we discuss our main empirical results. First, we present the results from maximum likelihood estimation of the parameters under the physical probability measure. Second, we discuss the performance of the option-valuation models using those physical parameters. Third, we discuss the estimation of parameters using options from Sample A (June 1, 1988, to May 31, 1991) and returns. We also discuss the performance of the different option-valuation models for this in-sample exercise. Fourth, we use the parameters estimated using Sample A options and asset returns to price options in Sample B (June 1, 1991, to May 31, 1992).

4.1. Maximum Likelihood Estimation Using Returns

Tables 4 and 5 show parameter estimates obtained by maximization of the log-likelihood function (6). Robust standard errors are computed according to White (1982). For the risk-free interest rate, we have assumed a constant 5% yearly rate leading to a daily rate of $0.05/365 = 0.000137$. It is well known that it is difficult to estimate GARCH parameters precisely from returns data unless long time series are used. For comparable models, Hentschel (1995) and Ding et al. (1993) use more than 50 years of daily data. Table 4 presents results obtained using a 12-year period beginning at the start of our options data set and ending on December 31, 1999, the last date for which CRSP data were available. Table 5 investigates the robustness of these results by reporting results obtained using an 18-year period beginning on June 1, 1981, and ending on December 31, 1999.[14]

The results are comparable across both samples and also comparable with standard findings in the literature on GARCH processes. First, consider the *Simple* and *Leverage* models. Parameter estimates for β_1 and β_2 are very precise and roughly of the same order of magnitude as in the existing literature. For the *Leverage* model the estimate of the parameter θ is positive, indicating negative skewness.[15] The persistence of the process implied by these parameter estimates in both samples is 0.9823 and 0.9835, indicating a very persistent process, again in accordance with the literature. Finally, the standard deviation of returns implied by the model parameters in both samples is 0.1786 and 0.1672, which is reasonable.

While implied persistence and standard deviation for the three more richly parameterized models *News*, *Power*, and *News&Power* are very similar to the *Leverage* model, other estimation results differ between the models. While in the *News* model the additional parameter κ is not estimated significantly, the power parameter γ is estimated to be significantly smaller than the one in the *Power* and *News&Power* models. Interestingly, the parameter κ is estimated significantly in the *News&Power* model, but the estimate of the θ parameter in this model is very different from the other models and no longer significant. The combination of κ and γ is apparently able to capture part of the leverage effect previously estimated by θ. Note also that the *Box-Cox* model has a power parameter ψ which is about two standard deviations below 1.

[13] The online appendix contains a time series plot of the underlying index returns and their absolute values.

[14] We also experimented with shorter sample periods, which yielded roughly comparable point estimates but lower t-statistics. Notice that the return sample includes the time period in which we compute option prices.

[15] This estimate also indicates a negative relationship between shocks to returns and volatility, labeled the "leverage effect" by Black (1976). This effect has been documented by a large number of studies that estimate stock returns. Using option prices the presence of this effect has been confirmed among others by Benzoni (1998), Chernov and Ghysels (2000), Eraker (2004), Heston and Nandi (2000), and Nandi (1998).

Table 2 **Average Quoted Price Across Moneyness and Maturity**

	Sample A					Sample B			
	DTM < 60	60 < DTM < 180	180 < DTM	All		DTM < 60	60 < DTM < 180	180 < DTMs	All
S/X < 0.94	1.52	5.05	10.20	6.62	S/X < 9.94	0.82	3.36	9.40	5.43
0.94 < S/X < 0.97	2.63	9.69	18.93	8.59	0.94 < S/X < 0.97	1.81	7.80	17.80	7.85
0.97 < S/X < 1.00	5.35	14.97	24.93	12.08	0.97 < S/X < 1.00	4.86	13.50	24.00	11.23
1.00 < S/X < 1.03	11.14	21.38	31.84	18.00	1.00 < S/X < 1.03	11.80	20.90	31.90	19.13
1.03 < S/X < 1.06	18.58	28.35	37.01	24.62	1.03 < S/X < 1.06	20.60	29.00	40.20	26.71
1.06 < S/X	40.35	50.39	62.65	49.14	1.06 < S/X	40.00	48.30	60.30	47.46
All	18.92	25.53	33.53	24.72	All	18.12	22.78	31.33	22.88

Notes. We report descriptive statistics on our two subsamples of call options on the S&P 500 index quoted at the closing of every Wednesday. Sample A denotes June 1, 1988–May 31, 1991; and Sample B denotes June 1, 1991–May 31, 1992. Each statistic is reported for three maturity bins and six moneyness bins.

Finally, and probably most importantly, the likelihood ratio tests indicate that any model is preferred to the *Simple* model. When testing against the *Leverage* model, the *News&Power* model is found to be significantly better, and the *Power* model to be marginally better.[16]

4.2. Option Valuation with Maximum Likelihood Estimates of Physical Processes

At the bottom of Tables 4 and 5, we present results obtained by using the parameters obtained in §4.1 to price options in Sample A (June 1, 1988, to May 31, 1991) as well as in Sample B (June 1, 1991, to May 31, 1992). To assess the models' fit, we present the mean-squared dollar error loss as well as its square root, which expresses the valuation error in dollar terms

$$\$RMSE = \sqrt{\frac{1}{N_T}\sum_{i,t}(C_{i,t} - C_{i,t}(h_t))^2}, \tag{18}$$

where $C_{i,t}$ is the market price of contract i on day t, $C_{i,t}(h_t)$ is the corresponding model price, and N_T denotes the total number of contracts available. The evidence can be summarized very briefly. The more complex models usually perform worse than the *Leverage* model, and sometimes even worse than the *Simple* model. Remember that both the *Simple* and the *Leverage* model are nested in the more complex models. In light of its inferior performance using the maximum likelihood (ML) criterion above, it is perhaps surprising that the *Leverage* model performs so well in terms of option valuation. We conclude that inference procedures from ML criteria are not reliable to rank the models' performance in option valuation.[17]

4.3. Nonlinear Least-Squares Estimation Using Options and Returns

In this section we present estimation results obtained using 156 Wednesdays of options data from Sample A. For all models, the objective function used is the (square root) dollar mean-squared error loss function (18). Model option prices are obtained using (9), (10), and (14), and volatility is updated using the volatility dynamics.[18] Robust standard errors are computed according to White (1982).

Table 6 presents the results for the nonlinear least-squares (NLS) estimation of the risk-neutral parameters. A number of results are noteworthy. First, the persistence for all models is extremely high and roughly the same across models. Persistence is higher than the persistence for the physical process in Tables 4 and 5, even when transforming the physical parameters using the risk-neutralization procedure. The implied standard deviation of the four models is roughly comparable and reasonable.

To understand these results, consider the simple *Leverage* model. Leverage under the physical measure is simply θ, but under the risk-neutral measure it is $\theta+\lambda$, which is greater if the price of risk is positive. Volatility persistence under the physical measure is $\beta_1+\beta_2(1+\theta^2)$ and under the risk-neutral measure it is $\beta_1+\beta_2(1+(\theta+\lambda)^2)$, which is higher than the physical persistence for a positive θ (implying negative skewness) and a positive price of risk λ.

When comparing the estimated parameter values across models in Table 6, it is of interest to remember that in many cases it is combinations of parameters

[16] It must be noted that the *Simple* model in this paper sets the price of risk λ equal to zero. Allowing for a nonzero λ in the *Simple* model will induce a leverage effect under the risk-neutral measure even when none exists under the physical measure. To avoid any type of leverage effect under any measure for the *Simple* model, we implement the model with $\lambda=\theta=0$. Our estimates therefore imply a symmetric news impact function under both measures.

[17] It is not straightforward to relate these results to existing ones in the literature. Amin and Ng (1993) perform a similar analysis using data on single stock options. Their results favor Nelson's (1990) EGARCH model. Hardle and Hafner (2000) estimate model parameters under the objective distribution but use relative valuation errors. Bollerslev and Mikkelsen (1996) analyze leaps. Their results are therefore difficult to compare to the ones in this paper. The online appendix contains further analysis of this issue.

[18] Details on the implementation of this procedure and a discussion of the choice of data set are contained in the online appendix.

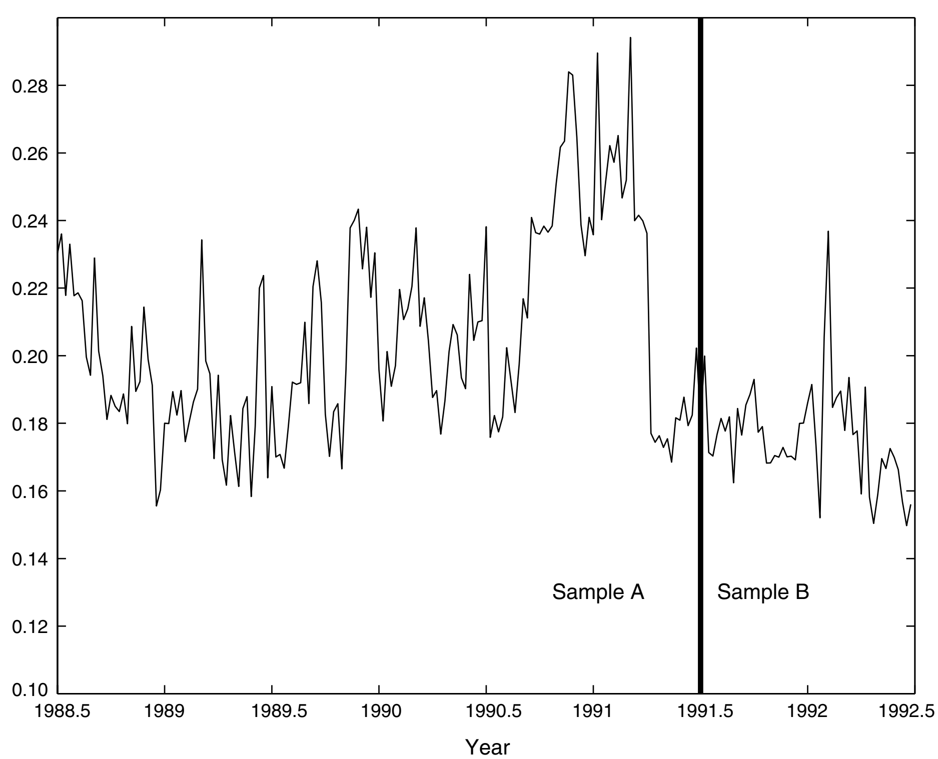
Table 3 Average Implied Volatility Across Moneyness and Maturity

	Sample A					Sample B			
	DTM < 60	60 < DTM < 180	180 < DTM	All		DTM < 60	60 < DTM < 180	180 < DTM	All
S/X < 0.94	0.1786	0.1706	0.1668	0.1701	S/X < 0.94	0.1339	0.1377	0.1506	0.1423
0.94 < S/X < 0.97	0.1652	0.1719	0.1765	0.1700	0.94 < S/X < 0.97	0.1346	0.1461	0.1607	0.1452
0.97 < S/X < 1.00	0.1712	0.1821	0.1858	0.1775	0.97 < S/X < 1.00	0.1445	0.1566	0.1722	0.1536
1.00 < S/X < 1.03	0.1901	0.1941	0.1976	0.1927	1.00 < S/X < 1.03	0.1626	0.1704	0.1781	0.1685
1.03 < S/X < 1.06	0.2172	0.2047	0.1942	0.2096	1.03 < S/X < 1.06	0.1847	0.1824	0.1896	0.1846
1.06 < S/X	0.3131	0.2363	0.2179	0.2626	1.06 < S/X	0.2574	0.2104	0.2037	0.2280
All	0.2262	0.1992	0.1912	0.2080	All	0.1847	0.1707	0.1756	0.1771

Notes. We report descriptive statistics on our two subsamples of call options on the S&P 500 index quoted at the closing of every Wednesday. Sample A denotes June 1, 1988–May 31, 1991; and Sample B denotes June 1, 1991–May 31, 1992. Each statistic is reported for three maturity bins and six moneyness bins.

that determine the model's most important characteristics, and not necessarily the individual parameters. For instance, the model's persistence is an important model characteristic and persistence is determined by a combination of all the model's parameters. Nevertheless, it is tempting to compare individual parameter estimates across models. For instance, the parameter β_0, which determines the model's unconditional volatility, is estimated in a fairly narrow range across the four models. Another important parameter is the $\theta^* \equiv \theta + \lambda$ combination, which indicates the size of the leverage effect. This effect is forced to zero in the *Simple* model. It is seen that θ^* is estimated at 2.9939 for the *Leverage* model, and somewhat lower for the more richly specified models. We therefore conclude that some of the effects captured by the parameters κ and γ are captured by the leverage effect if κ and γ are omitted. It is also noteworthy that when κ and γ are both estimated in the *News&Power* model, their estimates are not very different from the ones obtained in the *News* and *Power* models, respectively. Finally, we see that the power parameter in the

Figure 2 Average Implied Volatility of S&P 500 Index Options



Notes. On each Wednesday, we plot the simple average implied Black-Scholes volatility across the S&P 500 index option contracts observed at the close of trading. The average implied volatilities are shown for each Wednesday from June 1, 1988, through May 31, 1992. The vertical line delimits Sample A, which corresponds to June 1, 1988, through May 31, 1991, from Sample B, which corresponds to June 1, 1991, through May 31, 1992.

Table 4 Maximum Likelihood Estimates of Physical Processes

	Simple	*Leverage*	*News*	*Power*	*News&Power*	*Box-Cox*
r	1.37E-04	1.37E-04	1.37E-04	1.37E-04	1.37E-04	1.37E-04
λ		0.0452	0.0455	0.0463	0.0459	0.0464
		0.0185	0.0185	0.0186	0.0185	0.0186
β_0	1.84E-06	2.24E-06	2.30E-06	2.07E-06	1.98E-06	8.46E-06
	2.27E-07	1.82E-07	1.85E-07	1.94E-07	1.86E-07	5.46E-06
β_1	0.8873	0.8524	0.8601	0.8389	0.8528	0.8620
	0.0054	0.0052	0.0101	0.0107	0.0126	0.0050
β_2	0.0984	0.0867	0.0771	0.1222	0.1884	0.0988
	0.0026	0.0053	0.0102	0.0187	0.0261	0.0084
θ		0.7061	0.5605	0.6327	−0.1059	0.6379
		0.0845	0.1509	0.0754	0.0728	0.0779
κ			0.1267		0.7025	
			0.1093		0.0840	
γ				0.8515	0.5335	
				0.0626	0.0481	
ψ						0.8532
						0.0700
Statistical properties						
Persistence*	0.9857	0.9823	0.9822	0.9879	0.9931	0.9831
Annual Std. Dev.	0.1799	0.1786	0.1802	0.2075	0.2691	0.1876
Loglikelihood	1,0590.3	10,639.0	10,639.6	10,640.8	10,651.3	10,639.7
LR P-value (*Simple*)		0.0000	0.0000	0.0000	0.0000	0.0000
LR P-value (*Leverage*)			0.2690	0.0557	0.0000	0.2331
Option \$RMSEs						
Sample A	1.8126	1.4254	1.4315	1.6744	2.7158	1.4889
Sample B	1.5961	1.1940	1.1530	1.9294	3.3571	1.5422

Notes. We estimate the six GARCH models using maximum likelihood on daily S&P 500 returns from June 1, 1987, through December 31, 1999, for a total of 3,182 observations. Standard errors from White (1982) appear below each estimate. The persistence and the annualized standard deviation implied by each model is reported at the bottom of the table. (*) For the *Box-Cox* model, persistence refers to the power of volatility rather than volatility itself. The LR P-values refer to likelihood ratio tests of the null hypothesis that a particular model fits the return data no better than the *Simple* GARCH model or the *Leverage* GARCH model, respectively. The option-pricing fit (\$RMSE) implied from the risk-neutral transformation of the GARCH MLEs are reported for the two samples described in Tables 1–3.

Box-Cox model is much higher compared to the estimate in Tables 4 and 5, but still roughly two standard deviations below 1.

The most important conclusion from Table 6 is obtained by comparing the estimation results with those of Tables 4 and 5, obtained under the physical probability measure. For each model, the resulting two sets of parameter estimates are related by virtue of the precise form of the risk neutralization discussed in §2.2. It can be seen that implied parameter estimates are quite different. The most important differences are the following. First, for all models, the estimates of β_0 are not just different, but of a different order of magnitude. Second, the estimate of $\theta^* = \theta + \lambda$ implied by Tables 4 and 5 is positive for all models except for the *News&Power* model. However, it is always much smaller than the corresponding estimate in Table 6, implying a much smaller leverage effect even after adding λ for proper comparison. One potential explanation is that λ is poorly estimated under the physical measure where it only enters the conditional mean equation. Third, the estimates of κ in Tables 4 and 5 have a different sign than the ones in Table 6, even though the parameter is not significantly estimated for the *News* model. On the other hand, the estimates of γ in Tables 4 and 5 are not too different from those in Table 6.

4.4. Option Valuation with NLS Estimates of Risk-Neutral Processes

When comparing the six models in-sample, the most important criterion is the value of the minimized objective function. These values are presented in the last row of Table 6 and repeated with more detail in the first panel of Table 7. First, we note that all GARCH models outperform the Black-Scholes model as well as the Practitioner's Black-Scholes (PBS) model from Dumas et al. (1998) in Sample A.[19] Table 7A presents two versions of the PBS model,

[19] The PBS model simply assumes a second-order polynomial in strike price and maturity for the implied Black-Scholes volatility. It must be noted that the implementation of the PBS model reported on in Table 6 is different from the one in Dumas et al. (1998). The coefficients in the polynomial used for Table 6 are constant across the sample to ensure comparability of the results with the implementation of the GARCH models in this table. In Dumas et al. (1998), the coefficients of the polynomial are constant for a given day only. We present an empirical analysis that is identical to the one in Dumas et al. (1998) below.

Table 5 Maximum Likelihood Estimates on Long Return Sample

	Simple	*Leverage*	*News*	*Power*	*News&Power*	*Box-Cox*
r	1.37E-04	1.37E-04	1.37E-04	1.37E-04	1.37E-04	1.37E-04
λ		0.0434	0.0436	0.0437	0.0442	0.0437
		0.0150	0.0150	0.0150	0.0150	0.0150
β_0	1.37E-06	1.83E-06	1.89E-06	1.69E-06	1.56E-06	4.80E-06
	1.69E-07	1.55E-07	1.58E-07	1.57E-07	1.59E-07	2.64E-06
β_1	0.9151	0.8847	0.8867	0.8785	0.8812	0.8901
	0.0037	0.0041	0.0067	0.0075	0.0089	0.0039
β_2	0.0729	0.0718	0.0692	0.0890	0.1335	0.0778
	0.0017	0.0035	0.0061	0.0117	0.0194	0.0058
θ		0.6131	0.5220	0.5835	−0.0329	0.5840
		0.0687	0.1190	0.0659	0.0840	0.0666
κ			0.0659		0.4975	
			0.0717		0.0750	
γ				0.8947	0.6277	
				0.0530	0.0533	
ψ						0.8920
						0.0584
Statistical Properties						
Persistence*	0.9880	0.9835	0.9828	0.9862	0.9918	0.9840
Annual Std. Dev.	0.1698	0.1672	0.1661	0.1754	0.2196	0.1680
Loglikelihood	1,5612.8	15,663.6	15,663.9	15,664.8	15,671.5	15,664.2
LR P-value (*Simple*)		0.0000	0.0000	0.0000	0.0000	0.0000
LR P-value (*Leverage*)			0.4719	0.1274	0.0004	0.2722
Option $RMSEs						
Sample A	1.8663	1.4653	1.4652	1.3844	1.9363	1.4083
Sample B	1.6412	0.9988	1.0130	1.1104	2.2444	1.0528

Notes. We estimate the six GARCH models using maximum likelihood on daily S&P 500 returns from June 1, 1981, through December 31, 1999, for a total of 4,700 observations. Standard errors from White (1982) appear below each estimate. The persistence and the annualized standard deviation implied by each model is reported at the bottom of the table. (*) For the *Box-Cox* model, persistence refers to the power of volatility rather than volatility itself. The LR P-values refer to likelihood ratio tests of the null hypothesis that a particular model fits the return data no better than the *Simple* GARCH model or the *Leverage* GARCH model, respectively. The option-pricing fit ($RMSE) implied from the risk-neutral transformation of the GARCH MLEs are reported for the two samples described in Tables 1–3.

where all parameters are kept constant throughout the in-sample period. The PBS model is estimated using either OLS on implied volatility or NLS-minimizing $MSE.[20]

It is clear that the extra parameters in the *News*, *Power* and *Box-Cox* models do not improve the fit of the model, as the $RMSE is only slightly lowered from 1.0445 to 1.0410, 1.0400, and 1.0440 respectively. However, the combination of κ and γ in the *News&Power* model lowers the $RMSE to 1.0106.[21]

The evidence discussed above pertains to in-sample evaluation of the option valuation models. However, from a practitioner's perspective out-of-sample valuation performance is much more important. The second panel of Table 7A investigates out-of-sample valuation performance by using the model parameters estimated in Table 6 (using Sample A) to evaluate the models' performance in Sample B. The most important conclusion is that the ranking of the models is reversed compared to the in-sample exercise. While the most richly parameterized model (*News&Power*) has the best in-sample performance, it has the worst out-of-sample performance except for the *Simple* model. Conversely, the most parsimonious model (*Leverage*) has the worst in-sample performance, but performs better than all the other models, bar *News*, out-of-sample. It is interesting to compare this pattern with the one found in Tables 4 and 5 when estimating under the physical probability measure. In Tables 4 and 5, the change in the log likelihood for the *News* and *Box-Cox* models is very small, and the *News&Power* model yields a bigger change in the log likelihood. However, when using the *News&Power* model for (out-of-sample) option valuation in Tables 4 and 5, it performs poorly.

It is also interesting to compare the fit from the NLS estimates with the fit obtained using the (risk-neutralized) ML estimates in Tables 4 and 5. The $RMSE differs somewhat between Tables 4 and 5. For example, for Sample B the valuation error for the *Leverage* model based on ML estimates is much higher in Table 4 ($1.1940) than in Table 5 ($0.9988). However, the $RMSE of $0.9988 is very close to the valua-

[20] The PBS model is traditionally implemented using an OLS setup (see Dumas et al. 1998). Christoffersen and Jacobs (2004) demonstrate that the pricing errors can be lowered considerably by using an NLS setup.

[21] The online appendix reports MSEs by moneyness and maturity and contains further discussion of the results.

Table 6 NLS Estimates of Risk-Neutral Processes: Sample A

	Simple	*Leverage*	*News*	*Power*	*News&Power*	*Box-Cox*
r	1.37E-04	1.37E-04	1.37E-04	1.37E-04	1.37E-04	1.37E-04
β_0	4.89E-07	5.92E-07	5.79E-07	5.92E-07	5.66E-07	9.64E-07
	1.75E-08	9.74E-09	9.63E-09	9.75E-09	9.82E-09	2.65E-07
β_1	0.9699	0.8629	0.8666	0.8649	0.8253	0.8735
	0.0007	0.0037	0.0041	0.0030	0.0038	0.0058
β_2	0.0250	0.0133	0.0209	0.0291	0.1832	0.0163
	0.0007	0.0003	0.0245	0.0023	0.0166	0.0017
$\theta+\lambda$		2.9939	2.9190	2.3321	2.4186	2.7391
		0.0737	0.0790	0.0784	0.0525	0.1381
κ			−0.1933		−0.6217	
			0.4760		0.0285	
γ				0.8306	0.7193	
				0.0184	0.0157	
ψ						0.9466
						0.0294
Persistence*	0.9950	0.9959	0.9962	0.9961	0.9962	0.9953
Annual Std. Dev.	0.1880	0.2302	0.2346	0.2347	0.2347	0.2032
$MSE	2.6028	1.0910	1.0836	1.0817	1.0214	1.0900
$RMSE	1.6133	1.0445	1.0410	1.0400	1.0106	1.0440

Notes. We estimate the risk-neutral dynamics for each GARCH model directly by fitting the observed option prices using a nonlinear least-squares routine to minimize $MSE. Only options in Sample A (June 1, 1988–May 31, 1991; 8,081 contracts) are used in estimation. Robust standard errors are reported below each parameter estimate. The bottom of the table reports the risk-neutral volatility persistence and the risk-neutral annualized standard deviation implied by the GARCH parameters. (*) For the *Box-Cox* model, persistence refers to the power of volatility rather than volatility itself. We also report the $MSE and $RMSE at the parameter optima.

tion error in Table 7A based on NLS estimates under risk neutrality, which is $0.9888. It therefore appears that one can obtain a fairly satisfactory fit using ML estimates. This is perhaps somewhat surprising, given the differences between risk-neutralized ML estimates and risk-neutral NLS estimates noted earlier. A potential explanation is that both skewness and kurtosis affect the price of in- and out-of-the-money options (see, for example, Heston 1993). While θ^* (determining skewness) is smaller in Table 6, the parameter that determines kurtosis (β_2) is larger. It is therefore possible that risk-neutral and physical estimates lead to similar valuation results by emphasizing different moments.

Table 7B compares the accuracy of the models using the Diebold and Mariano (1995) test on the weekly $RMSEs from Sample B, keeping the parameters fixed at their Sample A values. We test each model against the *Leverage* model. Not surprisingly, the *Simple* model is rejected, as are the Black-Scholes and the PBS(OLS) models. None of the alternative GARCH models is significantly different from the *Leverage* model.

The following broad conclusions emerge from the results in Table 7: First, even when using relatively large samples (for example, Sample A contains more than 8,000 option contracts) and for the parsimonious models (3–7 parameters) studied here, results from in-sample estimation do not carry over to out-of-sample experiments. Second, when using options to estimate the parameters, the *Simple* model without leverage effect and price of volatility risk performs poorly, but it is difficult to improve on the *Leverage* model. Third, all GARCH models (except for the *Simple* model) outperform the Black-Scholes model as well as the Practitioner Black-Scholes model from Dumas et al. (1998) with constant parameters throughout the period.

5. Further Exploration of the Results

In this section we present additional empirical results and we provide further interpretation of some of the results presented before. First, we evaluate the performance of the option-valuation models in Sample B by reestimating the models every week and valuing the options one week out-of-sample. Subsequently, we use our empirical results to further explore the relationship between the physical and risk-neutral densities. We use these results to comment on the quality of the risk-neutralization procedure and the likely source of error in the class of models under study.

5.1. Option Valuation with Weekly Updating

The conclusion from the out-of-sample exercise is that one should use the parsimonious *Leverage* model rather than the more richly parameterized models. However, it may be argued that the out-of-sample exercise we conduct is very different from the way these models are typically used by practitioners. It may be unrealistic to assume that the model's parameters are constant over the entire sample period. We now investigate the robustness of our conclusions in a situation where we allow the model parameters to change over time. We work exclusively with the data in Sample B (June 1, 1991, to May 31, 1992) and

Table 7A Fit of Option Prices from NLS Estimates: Sample A Estimates

Sample A Fit	*Simple*	*Leverage*	*News*	*Power*	*News&Power*	*Box-Cox*	Black-Scholes	PBS(OLS)	PBS(NLS)
Sample $MSE	2.6028	1.0910	1.0836	1.0817	1.0214	1.0900	3.7351	3.6157	2.9307
Sample $RMSE	1.6133	1.0445	1.0410	1.0400	1.0106	1.0440	1.9326	1.9015	1.7119
Sample B Fit									
Sample $MSE	2.4250	0.9777	0.9684	1.0197	1.1571	0.9825	4.0112	3.5982	1.4511
Sample $RMSE	1.5573	0.9888	0.9841	1.0098	1.0757	0.9912	2.0028	1.8969	1.2046

Table 7B Comparing Weekly Predictive Accuracy with the Leverage Model

Sample B	*Simple*	*Leverage*	*News*	*Power*	*News&Power*	*Box-Cox*	Black-Scholes	PBS(OLS)	PBS(NLS)
Mean weekly $RMSE	1.5169	0.9032	0.9002	0.9278	1.0020	0.9062	1.9535	1.7662	1.1298
DM-test value	6.9936		−1.4427	1.5817	1.3143	1.2179	15.2805	6.3207	1.7858
P-value (2-sided)	0.0000		0.1491	0.1137	0.1887	0.2233	0.0000	0.0000	0.0741

Notes. We compute the $MSE and $RMSE of the option prices estimated in Table 6 on both Sample A and Sample B. The Sample A numbers for the six GARCH models are identical to those reported in Table 6. We also report Sample A fits for three other models: First, the standard Black-Scholes with volatility estimated using nonlinear least-squares minimizing $MSE. Second, the Practitioner Black-Scholes, PBS(OLS), from Dumas et al. (1998), who regress implied volatility on a second-order polynomial in strike price and time to maturity. Third, PBS(NLS), which estimates the PBS polynomial using nonlinear least squares to minimize $MSE. We keep all parameters constant across the 156 weeks in Sample A. We also report the out-of-sample fits from the nine models in Sample B, using the parameter estimates from Sample A. In Table 7B we calculate the average of the weekly $RMSE across the 52 weeks in Sample B. Due to the concavity of the square-root function, the average weekly $RMSEs are lower than the sample $RMSE calculated over all the 52 weeks. We use the weekly $RMSE sequences to test the significance of the difference in fits across models by applying the Diebold-Mariano (DM) (1995) test. The DM test is implemented, allowing for autocorrelation of up to four weeks in the $RMSE difference sequence. We take the *Leverage* GARCH model to be the benchmark in the pairwise DM tests. The DM test has a standard normal distribution and we report the two-sided *P*-values.

conduct the following exercise: For each model we estimate different parameter values every Wednesday and use these parameter values to price the options the next Wednesday. In addition to using one day of option prices, we also use a volatility updating rule for the 250 days predating the Wednesday used in the estimation exercise.

Table 8 contains the results of the weekly estimation analysis. The most important part of the table are Panels B and C, which report the results of the one-week-ahead exercise. For each of the 52 Wednesdays, we evaluate the fit of each day using parameters estimated from the previous Wednesday and returns data for the 250 days preceding it. A first important conclusion is that the $RMSEs in Table 8 are much lower than the corresponding numbers in Table 7. For example, the Sample B average dollar error for the *Leverage* model is $0.7081 with weekly updating, compared to $0.9777 when parameter estimates are not updated. When testing the average weekly $RMSEs in Panel C of Table 8, the conclusions

Table 8 Fit of Option Prices from Weekly NLS Estimates on Sample B

	Simple	*Leverage*	*News*	*Power*	*News&Power*	*Box-Cox*	Black-Scholes	PBS(OLS)	PBS(NLS)
A. Current Week									
Sample $RMSE	1.4639	0.4223	0.3963	0.3882	0.3802	0.4197	1.4859	0.9257	0.3557
B. One Week Ahead									
Sample $RMSE	1.5540	0.7081	0.7348	0.7135	0.7454	0.6978	1.5656	1.0952	0.6351
C. Comparing Weekly Predictive Accuracy with the Leverage Model									
Mean weekly $RMSE	1.5295	0.5838	0.5980	0.6104	0.6558	0.5875	1.5429	0.8764	0.5395
DM test value	17.2292		1.5564	6.6388	1.9602	0.1746	20.5065	3.8026	−1.6054
P-value (2-sided)	0.0000		0.1196	0.0000	0.0500	0.8614	0.0000	0.0001	0.1084

Notes. On every Wednesday in Sample B, we estimate a new set of GARCH parameters for each model based only on options data from that particular day and returns from the past 250 days. We then report the Sample $RMSE for all the current Wednesdays (current week) as well as the Sample $RMSE for the same Wednesdays using the previous Wednesdays' GARCH estimates (one week ahead). We also report Sample $RMSE for three other models: First, the standard Black-Scholes model with volatility estimated using nonlinear least-squares minimizing $MSE. Second, the Practitioner Black-Scholes, PBS(OLS), from Dumas et al. (1998), who regress implied volatility on a second-order polynomial in strike price and time to maturity. Third, PBS(NLS) which estimates the PBS polynomial using nonlinear least squares to minimize $MSE. In these three models, the parameters are reestimated weekly as well. In Table 8C we calculate the mean of the weekly $RMSE across the 52 weeks in Sample B. Due to the concavity of the square-root function, the average weekly $RMSEs are lower than the sample $RMSEs calculated over all the 52 weeks. We use the weekly $RMSE sequences to test the significance of the difference in fits across models by applying the Diebold-Mariano (DM) (1995) test. The DM test is implemented, allowing for autocorrelation of up to four weeks in the $RMSE difference sequence. We take the *Leverage* GARCH model to be the benchmark in the pairwise DM tests. The DM test has a standard normal distribution, and we report the two-sided *P*-values.

from the out-of-sample analysis in Table 7 are reinforced. The parsimonious *Leverage* model is not outperformed by any of the alternative models, and it outperforms several other GARCH models as well as Black-Scholes and the PBS(OLS) model, which is now updated every week. Remarkably, the *Leverage* model is not significantly outperformed by the PBS(NLS) model, which is estimated minimizing $MSE.

For completeness, Panel A of Table 8 presents the in-sample fits using parameters that are updated weekly. In addition, Figure 3 presents the $RMSEs for each of the 52 Wednesdays in Sample B, both for the in-sample and out-of-sample analysis. It can be seen that out of sample, the relative performance of the four models can differ substantially from week to week.

In summary, the empirical out-of-sample results with weekly updating confirm the earlier out-of-sample results. When considering a family of GARCH models that includes a large number of existing models, there seems to be no good reason to look beyond a model that allows for volatility clustering and a leverage effect. While in-sample analysis on options data and estimation on returns data may suggest more richly parameterized models, the out-of-sample performance of those models is rather disappointing. Conversely, the comparisons between the GARCH option-valuation models and the PBS models can be interpreted as being rather positive for GARCH models. All GARCH models, bar the *Simple* one, outperform the PBS(OLS) model in- and out-of-sample in our paper. The equally simple but tougher competitor PBS(NLS) is only marginally better than the best GARCH model.

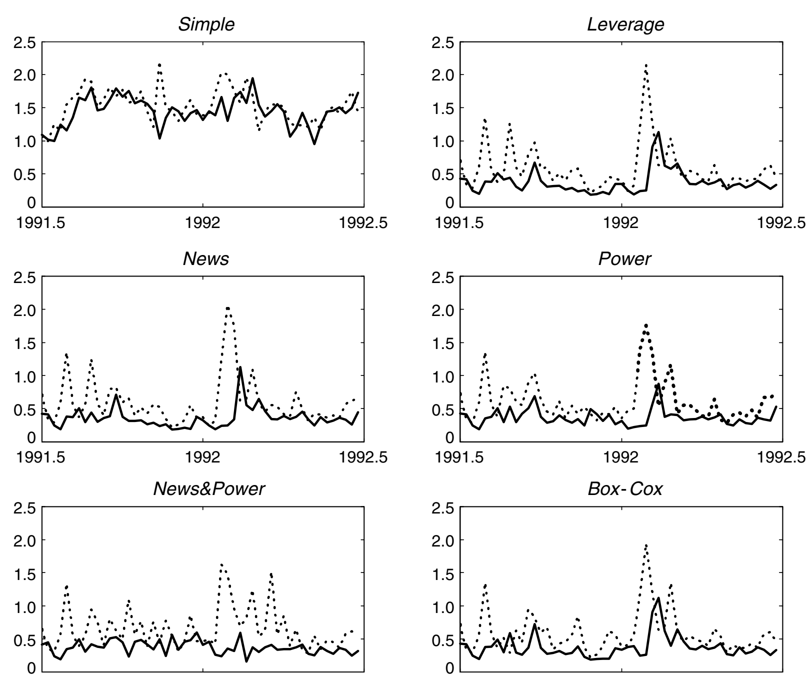
5.2. Discussion

These empirical results raise a number of important questions. The first question concerns the relationship between the estimation results under the physical and risk-neutral measures. In particular, do the empirical results empirically support the risk-neutralization procedure? Comparing Tables 4 and 5 with Table 6, the answer is no, because the parameter values under the two measures differ considerably after the risk neutralization. However, as mentioned before, it is interesting that the parameter estimates are very different under both measures, but that the resulting valuation errors are in some cases not dramatically different. This deserves further investigation.

A second question concerns the interpretation of the results. Given that we do not find a model that improves on the *Leverage* model, is this model adequate? What types of shortcomings does it have? Answers to these questions and a characterization of the model's weaknesses should prove valuable to guide the search for better models.

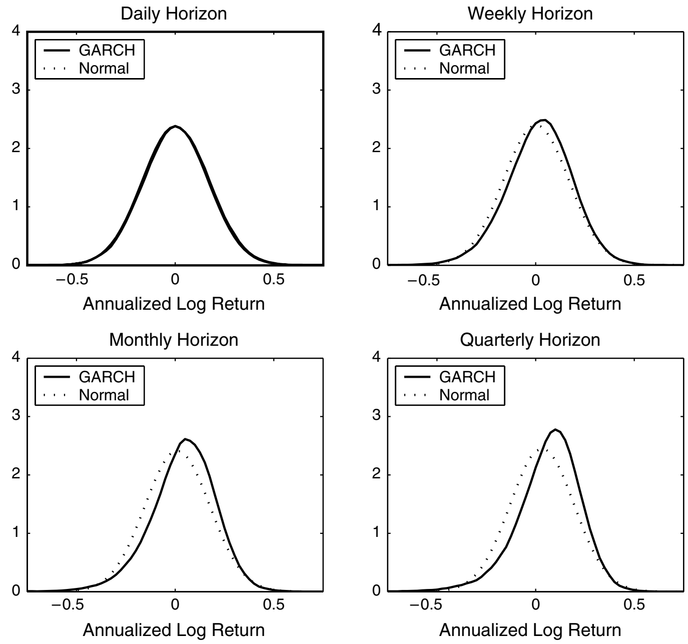
To provide answers to these questions, Figures 4 through 7 present estimates of a number of differ-

Figure 3 **In- and Out-of-Sample $RMSE: Weekly Estimates**



Notes. On every Wednesday in Sample B, we estimate a new set of GARCH parameters for each model based only on options data from that particular day and returns from the past 250 days. We then plot the $RMSE for the current Wednesday (in-sample) as well as the $RMSE for the same Wednesday using last Wednesday's GARCH estimates (out-of-sample). The solid lines denote in-sample $RMSE for each model and the dashed lines denote out-of-sample $RMSE.

Figure 4 Physical GARCH Return Densities vs. the Normal



Notes. Using standard normal innovations, we simulate a long sample of daily returns from the *Leverage* GARCH model starting from the unconditional variance and aggregating the daily data to various horizons. The GARCH densities (solid lines) are estimated using a normal kernel with a bandwidth that is optimal asymptotically for normal data. The dotted line denotes the normal p.d.f. with mean and variance matching the GARCH model at each horizon.

ent densities. Each figure contains four panels, which represent densities at different horizons. Figure 4 presents GARCH densities for the *Leverage* model based on the ML estimates in Table 5. The GARCH densities are compared with a normal density with a mean and variance equal to the mean and variance of the simulated data. The figure clearly illustrates that the unconditional GARCH density has excess skewness and kurtosis compared to the normal.

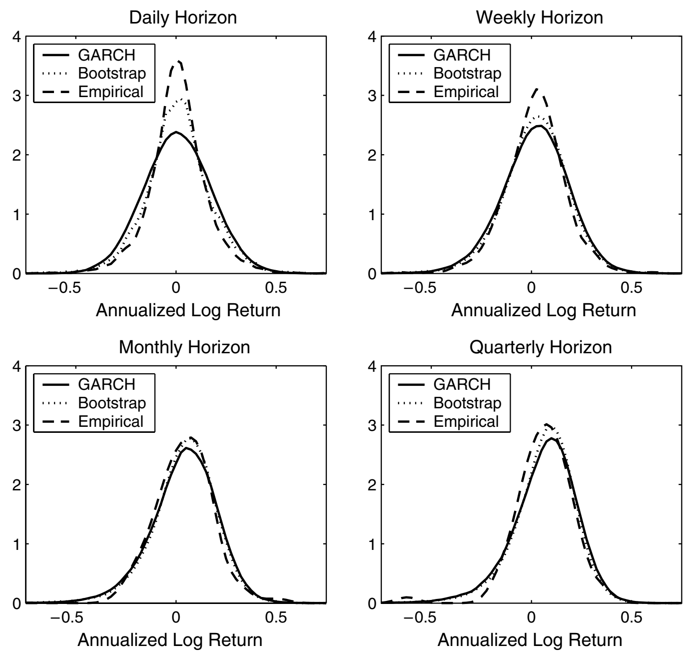
Figure 5 presents the same GARCH density, but now compares it with two other densities. The most important comparison is with the empirical density which is estimated nonparametrically. Even though the GARCH density has excess kurtosis compared to the normal (see Figure 4), it does not display enough kurtosis to match the data. The dotted lines in Figure 5 represent a density obtained by replacing the conditional normal innovation in the GARCH model with an innovation bootstrapped from returns. The comparison between the three densities shows that, on the one hand, the *Leverage* model is inadequate because the assumption of a conditional normal innovation is incorrect. In addition, even if the innovation is data driven, the model does not match the data density, indicating that the volatility dynamic is also misspecified. Interestingly, the differences between the three densities are most pronounced in the upper-left panel. This panel contains the densities for the daily horizon, which is not very relevant for option valuation. Model densities at the more relevant monthly and quarterly horizons are much more in agreement with the data.

Figure 6 addresses the differences between the ML and NLS estimates. For different horizons, it presents densities estimated nonparametrically on risk-neutral returns simulated from the *Leverage* GARCH model. One density is obtained using the NLS estimates, and the other one using the risk-neutralized ML estimates. It can clearly be seen that there are dramatic differences between the two densities. The density obtained using ML displays much more kurtosis and much less skewness than the NLS-based density. This is not surprising given the differences between the estimates in Tables 4 and 5 on the one hand, and Table 6 on the other hand. The estimate of β_2 is much higher in Tables 4 and 5, while the sum of θ and λ from Tables 4 and 5 is much lower than the estimate of $\theta + \lambda$ in Table 6.

Finally, Figure 7 investigates how closely the risk-neutral density implied by the NLS estimates matches the density implied from the observed option prices. This question cannot be answered using the technique used to generate Figure 6. Instead, we use the semiparametric technique advocated by Ait-Sahalia and Lo (1998).[22] In each panel, the dotted line serves as a reference point; it represents a normal density with mean equal to the average risk-free rate and

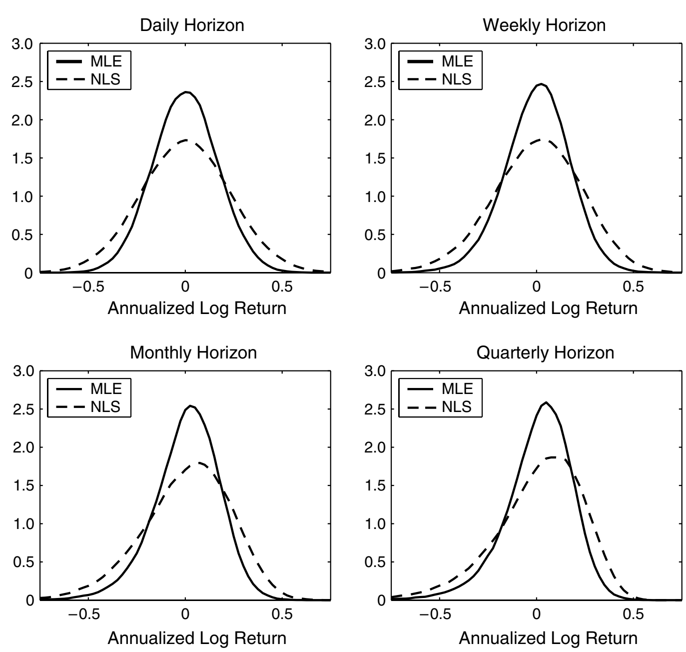
[22] The details of our implementation of their approach are described in the online appendix.

Figure 5 Physical GARCH Return Densities vs. the Empirical



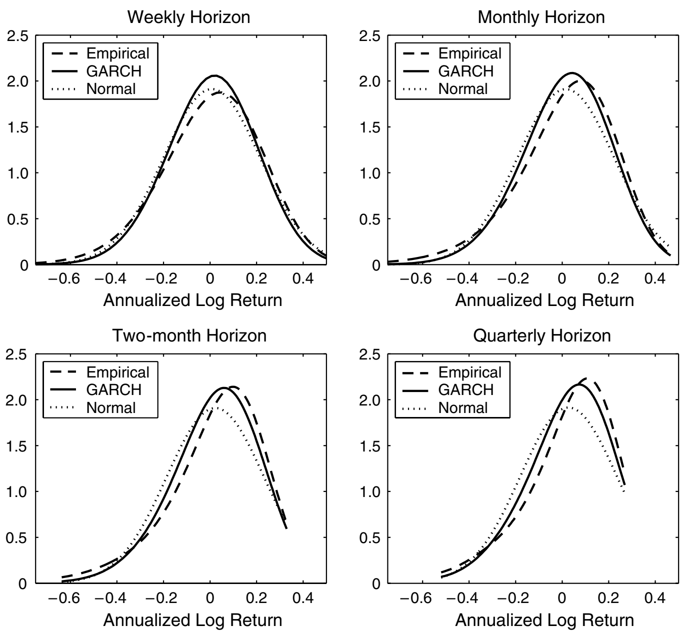
Notes. We simulate a long sample of daily returns from the *Leverage* GARCH model starting from the unconditional variance and aggregating to various horizons. The densities are estimated using a normal kernel with a bandwidth that is optimal asymptotically for normal data. The solid line uses innovations drawn from the standard normal distribution. The dotted line bootstraps innovations from observed returns. The dashed line denotes the empirical density calculated directly on nonoverlapping S&P 500 returns from June 1, 1981, through December 31, 1999.

Figure 6 Risk-Neutral Densities from Simulation



Notes. We simulate a long sample of daily risk-neutral returns from the *Leverage* GARCH model starting from the unconditional variance and aggregating to various horizons. The densities are estimated using a normal kernel with a bandwidth that is optimal asymptotically for normal data. The solid line uses the return-based ML estimates from Table 5. The dashed line uses the option-based NLS estimates from Table 6.

Figure 7 Semiparametric Risk-Neutral Densities from Option Prices



Notes. The dashed line shows the risk-neutral density estimates from observed option prices calculated using Ait-Sahalia's and Lo's (1998) semiparametric estimation method. See the online appendix for details on the calculation. The solid line represents the corresponding density using the GARCH option values rather than the market data. The GARCH estimates used are the return-based MLEs from Table 5. The dotted line corresponds to the normal distribution. The longer-horizon densities are truncated due to the availability of strike prices in the options data.

variance equal to the average Black-Scholes volatility. The model apparently does not contain enough skewness to match the data. Interestingly, this shortcoming of the model is especially relevant at longer horizons. This finding is very different from that for the physical densities in Figure 5, where the differences between model and data were less pronounced for longer horizons.[23]

We therefore arrive at some interesting conclusions. Modeling of dynamics richer than those of the *Leverage* model does not improve the fit of the model, because the implied densities are not sufficiently different.[24] Figure 5 suggests that it may be possible to improve on these models by modeling innovations that are different from the conditional model, but is not clear if this will be helpful at horizons that are relevant for valuation. Most importantly, for some reason GARCH models cannot capture the degree of skewness in the data. This is clearly seen from Figure 7. A potential explanation is that the functional specification that governs the skewness is not rich enough. One likely reason for this is that the risk neutralization, which partially determines skewness, is incorrectly specified. This conclusion is confirmed by Table 8, by comparison of the ML and NLS parameter estimates, and by inspection of Figures 5 and 7. While the GARCH density quite accurately captures the physical density at the quarterly horizon, it does not capture the skewness in the risk-neutral distribution. Interestingly, even though the risk-neutral densities implied by the ML and NLS estimates are quite different, pricing using ML estimates is fairly satisfactory. The ML estimates generate a fit far superior to Black-Scholes by picking a high kurtosis, while the NLS estimates do so by picking a higher skewness.

There is an interesting parallel between our findings on skewness and the findings in the continuous-time literature. Modelling skewness has also proven difficult for continuous-time models. The stochastic volatility model of Heston (1993) is able to generate negative skewness, but when using estimated parameters, not enough skewness is generated by the model. The proposed solution has been to add jumps in returns and volatility; see, e.g., Bates (1996), Eraker (2004), and Pan (2002), but the out-of-sample results from adding these jumps have arguably been mixed. Adequately modelling skewness has thus remained a challenge in the option-valuation literature.

[23] Note that the horizons in Figures 5 and 7 are different. The reason is that one cannot construct risk-neutral densities at daily horizons, because those option data are not available. The absence of options with relevant strike prices explains why the densities are truncated in Figure 7. With the method used in Figure 6, which is based on return simulation, this is not the case.

[24] It may, of course, be the case that the different dynamics do not perform very differently because the models are all misspecified in a more fundamental way. Bondarenko (2003) and Driessen and Maenhout (2003) argue that it is difficult to explain option prices with models of rational behavior.

6. Summary and Directions for Future Work

In the existing literature, different volatility models are typically judged by comparing the likelihoods from time series of asset returns. We instead investigate the performance of GARCH models explicitly for the purpose of option valuation. We compare a number of models using the families of specifications in Ding et al. (1993) and Hentschel (1995). We find that a comparison of the in-sample fit (based on dollar-squared option-valuation errors) favors the most richly parameterized model, as does a traditional likelihood-based comparison that uses only the time series of asset returns. However, when assessing the performance of the GARCH models out-of-sample (again using dollar-squared option valuation errors), the data favor the more parsimonious model that only contains simple volatility clustering and a leverage effect.

The consequences of these findings for the GARCH literature are potentially far reaching. The advantage of the specifications of Ding et al. (1993) and Hentschel (1995) is that they nest a large number of existing GARCH models, although of course not all. Our results suggest that it is dangerous to make inference about the potential for volatility models to fit option prices based only on statistical analysis on the underlying asset returns. Unfortunately, existing volatility model comparisons are often made in-sample, and most often based on the model's likelihood value, which may not have an obvious relationship with more economically motivated objective functions such as those based on option valuation.

Looking forward, our results suggest that most is to be gained by more radical departures from the modeling assumptions used in a traditional GARCH option-pricing model. For instance, modeling deviations from normality in the innovations process may prove instructive and is pursued in Christoffersen et al. (2003). Also, all the models investigated in this paper share the same specification for the price of risk. Our comparison of the risk-neutral and physical model densities suggest that it may be worthwhile to model the price of risk differently. We are currently pursuing work along these lines.

While our paper is set in discrete time, a significant part of the literature specifies latent stochastic volatility factors in continuous time. Early work includes Hull and White (1987), Johnson and Shanno (1987), Melino and Turnbull (1990), Scott (1987), and Wiggins (1987). Whereas the results of Nelson and Foster (1994), Duan (1997), and Heston and Nandi (2000) indicate that certain continuous-time models can be seen as the limits of discrete-time GARCH models, Corradi (2000) demonstrates that continuous-time models with deterministic volatility can also be retrieved as the limits of discrete-time GARCH processes. Therefore, neither at the theoretical nor empirical level is the relationship between GARCH processes and stochastic volatility models straightforward, and the implications for option valuation are not obvious either. We plan to address the relative performance of discrete- and continuous-time models in future work.

An online appendix to this paper is available at http://mansci.pubs.informs.org/ecompanion.html.

Acknowledgments

The authors thank David Bates, Jerome Detemple, Jin Duan, Rob Engle, Rene Garcia, Steve Heston, Nour Meddahi, and Eric Renault for helpful comments. John Kozlowski, Ines Chaieb, and Vadim di Pietro provided expert research assistance. They are grateful to FCAR, IFM^2, and SSHRC for financial support. Any remaining inadequacies are theirs alone.

References

Ait-Sahalia, Y., A. Lo. 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *J. Finance* **53** 499–547.

Amin, K., V. Ng. 1993. ARCH processes and option valuation. Working paper, University of Michigan, Ann Arbor, MI.

Andersen, T., L. Benzoni, J. Lund. 2002. An empirical investigation of continuous-time equity return models. *J. Finance* **57** 1239–1284.

Bakshi, C., C. Cao, Z. Chen. 1997. Empirical performance of alternative option pricing models. *J. Finance* **52** 2003–2049.

Bates, D. 1996. jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Rev. Financial Stud.* **9** 69–107.

Benzoni, L. 1998. Pricing options under stochastic volatility: An econometric analysis. Manuscript, University of Minnesota, Minneapolis, MN.

Black, F. 1976. Studies of stock price volatility changes. *Proc. 1976 Meetings Bus. Econom. Statist. Section.* American Statistical Association, Washington, D.C., 177–181.

Black, F., M. Scholes. 1973. The pricing of options and corporate liabilities. *J. Political Econom.* **81** 637–659.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327.

Bollerslev, T., H. O. Mikkelsen. 1996. Modeling and pricing long-memory in stock market volatility. *J. Econometrics* **73** 151–184.

Bollerslev, T., H. O. Mikkelsen. 1999. Long-term equity anticipation securities and stock market volatility dynamics. *J. Econometrics* **92** 75–99.

Bollerslev, T., R. Chou, K. Kroner. 1992. ARCH modelling in finance: A review of the theory and empirical evidence. *J. Econometrics* **52** 5–59.

Bollerslev, T., R. Engle, D. Nelson. 1994. ARCH models. R. Engle, D. McFadden, eds. *Handbook of Econometrics*, Vol. IV. Elsevier, Amsterdam, The Netherlands.

Bollerslev, T., R. Engle, J. Wooldridge. 1988. A capital asset pricing model with time-varying covariances. *J. Political Econom.* **96** 116–131.

Bondarenko, O. 2003. Recovering risk-neutral densities: A new nonparametric approach. Manuscript, University of Illinois at Chicago, Chicago, IL.

Brennan, M. 1979. The pricing of contingent claims in discrete-time models. *J. Finance* **34** 53–68.

Campbell, J., L. Hentschel. 1992. No news is good news: An asymmetric model of changing volatility in stock returns. *J. Financial Econom.* **31** 281–318.

Campbell, J., A. Lo, C. MacKinlay. 1997. *The Econometrics of Financial Markets.* Princeton University Press, Princeton, NJ.

Chernov, M., E. Ghysels. 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation. *J. Financial Econom.* **56** 407–458.

Christoffersen, P., K. Jacobs. 2004. The importance of the loss function in option valuation. *J. Financial Econom.* **72** 291–318.

Christoffersen, P., S. Heston, K. Jacobs. 2003. Option valuation with conditional skewness. Manuscript, McGill University, Montréal, Quebéc, Canada.

Corradi, V. 2000. Reconsidering the continuous time limit of the GARCH(1,1) process. *J. Econometrics* **96** 145–153.

Day, T., C. Lewis. 1992. Stock market volatility and the information content of stock index options. *J. Econometrics* **52** 267–287.

Derman, E., I. Kani. 1994. Riding on the smile. *Risk* **7** 32–39.

Diebold, F. X., R. Mariano. 1995. Comparing predictive accuracy. *J. Bus. Econom. Statist.* **13** 253–263.

Ding, Z., C. Granger, R. Engle. 1993. A long memory property of stock market returns and a new model. *J. Empirical Finance* **1** 83–106.

Driessen, J., P. Maenhout. 2003. A portfolio perspective on option pricing anomalies. Manuscript, INSEAD, Fontainebleau, France.

Duan, J.-C. 1995. The GARCH option pricing model. *Math. Finance* **5** 13–32.

Duan, J.-C. 1996. Cracking the smile. *Risk* **9** 55–59.

Duan, J.-C. 1997. Augmented GARCH(p,q) process and its diffusion limit. *J. Econometrics* **79** 97–127.

Duan, J.-C., P. Ritchken, Z. Sun. 2004. Jump starting GARCH: Pricing and hedging options with jumps in returns and volatilities. Manuscript, University of Toronto, Toronto, Ontario, Canada.

Duan, J.-C., J.-G. Simonato. 1998. Empirical martingale simulation for asset prices. *Management Sci.* **44** 1218–1233.

Dumas, B., J. Fleming, R. Whaley. 1998. Implied volatility functions: Empirical tests. *J. Finance* **53** 2059–2106.

Dupire, B. 1994. Pricing with a smile. *Risk* **7** 18–20.

Engle, R. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* **50** 987–1008.

Engle, R., C. Mustafa. 1992. Implied ARCH models from options prices. *J. Econometrics* **52** 289–311.

Engle, R., V. Ng. 1993. Measuring and testing the impact of news on volatility. *J. Finance* **48** 1749–1778.

Engle, R., C.-S. Hong, A. Kane, J. Noh. 1993. Arbitrage valuation of variance forecasts with simulated options. *Adv. Futures Options Res.* **6** 393–415.

Eraker, B. 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *J. Finance* **59** 1367–1403.

French, K., G. W. Schwert, R. Stambaugh. 1987. Expected stock returns and volatility. *J. Financial Econom.* **19** 3–30.

Garcia, R., E. Renault. 1998. A note on hedging in ARCH and stochastic volatility option pricing models. *Math. Finance* **8** 153–161.

Glosten, L., R. Jagannathan, D. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48** 1779–1801.

Hardle, W., C. Hafner. 2000. Discrete time option pricing with flexible volatility estimation. *Finance Stochastics* **4** 189–207.

Hentschel, L. 1995. All in the family: Nesting symmetric and asymmetric GARCH models. *J. Financial Econom.* **39** 71–104.

Heston, S. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financial Stud.* **6** 327–343.

Heston, S., S. Nandi. 2000. A closed-form GARCH option pricing model. *Rev. Financial Stud.* **13** 585–626.

Hsieh, K., P. Ritchken. 2000. An empirical comparison of GARCH option pricing models. Manuscript, Case Western Reserve University, Cleveland, OH.

Hull, J., A. White. 1987. The pricing of options with stochastic volatilities. *J. Finance* **42** 281–300.

Jiang, G. 1998. Testing option pricing models with stochastic volatility, random jump and stochastic interest rate—Open the "black box". Working paper, University of Groningen, Groningen, The Netherlands.

Johnson, H. E., D. Shanno. 1987. Option pricing when the variance is changing. *J. Finan. Quant. Anal.* **22** 143–151.

Jones, C. 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. *J. Econometrics* **116** 181–224.

Jorion, P. 1995. Predicting volatility in the foreign exchange market. *J. Finance* **50** 507–528.

Lamoureux, C., W. Lastrapes. 1993. Forecasting stock return variance: Towards an understanding of stochastic implied volatilities. *Rev. Financial Stud.* **5** 293–326.

Maheu, J., T. McCurdy. 2004. News arrival, jump dynamics and volatility components for individual stock returns. *J. Finance* **59** 755-779.

Melino, A., S. Turnbull. 1990. Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45** 239–265.

Nandi, S. 1998. How important is the correlation between returns and volatility in a stochastic volatility model? Empirical evidence from pricing and hedging in the *S&P 500 Index* options market. *J. Banking Finance* **22** 589–610.

Nelson, D. 1990. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59** 347–370.

Nelson, D., D. Foster. 1994. Asymptotic filtering theory for univariate ARCH models. *Econometrica* **62** 1–41.

Noh, J., R. Engle, A. Kane. 1994. Forecasting volatility and option prices of the *S&P 500 Index*. *J. Derivatives* **2** 17–30.

Pagan, A., G. W. Schwert. 1990. Alternative models for conditional stock volatility. *J. Econometrics* **45** 267–290.

Pan, J. 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *J. Financial Econom.* **63** 3–50.

Renault, E. 1997. Econometric models of option pricing errors. D. Kreps, K. Wallis, eds. *Advances in Economics and Econometrics: Theory and Applications: Seventh World Congress*, Vol. 3. Econometric Society Monographs, No. 28. Cambridge University Press, Cambridge, New York, and Melbourne, 223–278.

Ritchken, P., R. Trevor. 1999. Pricing options under generalized GARCH and stochastic volatility processes. *J. Finance* **54** 377–402.

Rubinstein, M. 1976. The valuation of uncertain income streams and the pricing of options. *Bell J. Econom.* **7** 407–425.

Schwert, G. W. 1989. Why does stock market volatility change over time? *J. Finance* **44** 1115–1153.

Scott, L. 1987. Option pricing when the variance changes randomly: Theory, estimators and applications. *J. Financial Quant. Anal.* **22** 419–438.

White, H. 1982. Maximum likelihood specification of misspecified models. *Econometrica* **50** 1–25.

Wiggins, J. B. 1987. Option values under stochastic volatility: Theory and empirical evidence. *J. Financial Econom.* **19** 351–372.