GARCH Option Pricing Models: Evidence from Joint Likelihood Estimations

I. Papantonis*
papantonis@econ.auth.gr
Aristotle University of Thessaloniki
February 7, 2015

Abstract
In this paper we utilize the affine-GARCH option-pricing framework of Heston and Nandi (2000) combined with the recently-introduced variance-dependent pricing kernel of Christoffersen, Heston and Jacobs (2013) (henceforth CHJ) that has been proved to explain some commonly observed option-pricing irregularities. In our empirical analysis we verify the validity of the implications of this kernel specification by employing a somewhat different estimation approach. We build a joint likelihood function that incorporates both spot and forward-looking information, but instead of working with the wide cross-section of index options as in CHJ, we propose to use the daily variance dynamics inferred from the VIX index to approximate the implied risk-neutral variance process. Minimizing the implied variance errors in the objective function is technically equivalent to minimizing a cross-section of vega-weighted option-valuation errors. This technique is straightforward and computationally more efficient and, as we show, it can produce similar results. We find strong evidence that support the hypothesis of priced volatility risk, since the model fits the data much better after allowing for this more flexible pricing kernel process. Our findings indicate a U-shaped pattern for the logarithmic kernel, which can be very helpful in explaining option-pricing puzzles and is also in line with the semi-parametric evidence documented in CHJ.

1 Introduction
There is a bulk of practical and theoretical studies in recent finance literature that strongly indicate a significant divergence of risk-neutral distributions from their physical counterparts in the post-crash ’87 period. Following Bates (1996), numerous stylized facts for option prices have been repeatedly reported, highlighting the necessity of more complex models for capturing the dynamics of returns and variance more accurately. Starting from the observation that risk-neutral volatility smiles have become much steeper than those in the pre-crash period, this dramatic change pointed out a research direction for option pricing away from the standard Black-Scholes model. The BS model has been proved inadequate to fit market option prices, especially for out-of-the-money put options, and this is primarily due to its restrictive requirement of a constant variance across time and moneyness and its implied log-linear pricing kernel that fails to reproduce the observed option price structure. However, providing a consistent link between empirical and option-implied densities has been proved to be an obscure task.

*Contact details: email: papantonis@econ.auth.gr, tel. +30 6948 944 560.
A sizable stream in the literature has focused on trying to provide explanations to several empirical option-pricing anomalies that are indeed quite prevalent in the index option market. The most commonly observed pattern is that of implied variance from options exceeding the future realized variance of the underlying. This robust empirical puzzle can be at least partially explained by a significant negative variance risk premium reported in recent studies. In an attempt to quantify this premium, Bakshi and Kapadia (2003) form delta-hedged portfolios of options and find that they are quite sensitive to changes in the underlying variance. Their findings are supportive of a consistently negative premium for variance risk which also seems to increase in more volatile periods, and these results persist even after accounting for jumps. Carr and Wu (2009) also document a significant variance risk premium of similar magnitude by capturing the difference between realized variances and variance swap rates synthesized from options. Intuitively, this negative premium suggests that investors are actually willing to bear lower excess returns in order to hedge themselves against unexpected volatility increases. Bollerslev, Tauchen, and Zhou (2009) and Ang et al. (2006) underline that volatility risk can be a factor of utmost importance in explaining the cross-section of equity returns.

From this puzzling behavior of variance it directly becomes evident that option-implied forward-looking variance cannot give an unbiased estimation of the subsequent realized variance of the underlying. Especially during the post-crash period, this bias has been the basic motivation for speculators to exploit profitable opportunities by writing options. Coval and Shumway (2001) implement option investment strategies and show that returns may vary with the strike price, while short positions on at-the-money straddles can produce on average economically significant profits. Bondarenko (2004) also finds evidence of negative and priced market volatility risk that is able to explain hedge-fund performance. In an equilibrium framework Driessen and Maenhout (2007) analyze the reasons why a utility-maximizing investor is always better-off when selling out-of-the-money puts or at-the-money straddles. Generally, researchers have reached consensus to some extent regarding the forecasting ability of implied volatilities. Findings support the idea that there is valuable information embedded in options that can be used to improve forecasts of the subsequent volatility. However, these forecasts are almost with certainty upwardly biased, suggesting the existence of a significant systematic volatility risk factor that appears to be priced in options.

What are the main drivers that give rise to this significant volatility risk premium still remains a major issue that requires further investigation. Carr and Wu (2009) and more recently Christoffersen, Heston, and Jacobs (2013) have suggested that the negative price of variance risk can be decomposed into two distinct components originating from different sources. The first source is attributed to the negative correlation between returns and volatility that has been heavily reported in the literature. This negative correlation is considered to be responsible for generating a pronounced asymmetric effect in the behavior of volatility. Theoretically, this

---

1 This model-free approach to replicate the variance swap rate from observed option prices was pioneered by Demeterfi, Derman, Kamal, and Zou (1999) and Britten-Jones and Neuberger (2000) and later extended by Jiang and Tian (2005) to include jumps in the asset price process.

asymmetric volatility nature has been linked to the “leverage” and/or the “volatility feedback” effect hypotheses, but unfortunately early empirical studies have yielded inconclusive results regarding the sign of the variance premium.\(^3\) The second source of a variance risk premium is due to the stochastic nature of volatility. This implies the existence of a significant component that can arise separately as an independent variance risk factor.

A primary objective of both academics and practitioners has been the development of models that can adequately fit the observed dynamics of returns and variance. Broadly speaking, there are two main categories of models. The first one seems to include models that belong in the GARCH-class. In this category, there is only one innovation that drives both returns and conditional volatility, and the asymmetric property can emerge by considering a GARCH specification that can introduce a correlation structure between returns and variance.\(^4\) The second category refers to the so-called stochastic volatility models, where returns and volatilities are considered to be driven by separate sources of innovations. Asymmetry can arise either by allowing for the two shock terms to be correlated or purely endogenously through the pricing kernel. These stochastic volatility diffusion processes could be further augmented with jumps in returns and volatility in order to produce richer statistical properties. However, reconciling and adequately specifying the implied return dynamics is not a trivial task, especially when it comes to applying these processes for the purposes of more accurate option-valuation.

In the context of stochastic volatility, Hull and White (1987) and Wiggins (1987) were among the first to develop an option-pricing model, followed by Melino and Turnbull (1990) and the seminal affine-model of Heston (1993). Bates (1996, 2000, 2006), Bakshi, Cao, and Chen (1997) and Andersen, Benzoni, and Lund (2002) among many others have demonstrated the importance of capturing jumps in returns for fitting option prices. Since the jump risks in financial time-series can be very significant, Duffie, Pan, and Singleton (2000), Pan (2002), Eraker, Johannes, and Polson (2003), Eraker (2004) and Broadie, Chernov, and Johannes (2007) have suggested that allowing for jumps in the stochastic process of volatility can further improve the understanding of the evolution of option prices.\(^5\)

Even though ARCH-type processes have dominated the literature of econometric time-series analyses, yet their implications for option-valuation have been discussed in only a quite limited number of papers until recently. Duan (1995) was the first to develop an equilibrium preference-based argument for GARCH processes under which a locally risk-neutral valuation relation (henceforth LRNVR) holds and European options can be priced. Kallsen and Taqqu (1998) obtain a similar result for ARCH processes in a continuous-time no-arbitrage setting that ensures market completeness. Duan, Gauthier, and Simonato (1997) utilize the LRNVR to analytically approximate option prices under non-linear asymmetric GARCH dynamics, later extending this approach for the GJR and the E-GARCH models as well (Duan, Gauthier, Simonato, 1997).


\(^4\)See Engle and Ng (1993) for a review on asymmetric GARCH models.

\(^5\)Todorov (2009) also documents the implications of stochastic volatility and jumps for the evolution of the variance risk premium dynamics.
& Sasseville, 2006). Härdle and Hafner (2000) also show that asymmetric threshold-GARCH models that allow for more flexible variance dynamics by capturing leverage effects are much better in explaining observed option prices. Ritchken and Trevor (1999) build a lattice algorithm that can be used to price both European and American-style options under generalized GARCH dynamics that also appear to have numerous bivariate stochastic volatility diffusions as limiting cases (a property also demonstrated in Duan (1997)). Heston and Nandi (2000) significantly contribute to this field by proposing an asymmetric affine GARCH model that also allows for a European-option pricing formula to be derived in closed-form. Duan and Simonato (2001) and Duan, Gauthier, Sasseville, and Simonato (2003) also discuss the case of approximating the prices of American options, either numerically or by finite-state Markov-Chain techniques, when the underlying follows a GARCH process. There are also some papers focusing on empirically comparing the out-of-sample performance of the GARCH variants when applied in option-valuation. Lehar, Scheicher, and Schittenkopf (2002) demonstrate that the GARCH option pricing models can provide superior results in terms of Value at Risk forecasts when compared to stochastic volatility models. Christoffersen and Jacobs (2004) compare several risk-neutral GARCH dynamics when fitted on market prices of options and display evidence in favor of more parsimonious approaches. Hsieh and Ritchken (2005) and, very recently, Kanninen, Lin, and Yang (2014) also support a similar finding, since they show that a simple asymmetric (non-affine) N-GARCH model of Engle and Ng (1993) can produce smaller out of sample option-valuation errors. Other authors have tried to improve the model’s fit by allowing for more realistic error distributions. Lehnert (2003) supports that an exponential E-GARCH (Nelson, 1991) model with a GED is better for explaining option prices (see also Christoffersen, Dorion, Jacobs, and Wang (2010)). Christoffersen, Heston, and Jacobs (2006) and Stentoft (2008) consider the case of a Normal Inverse Gaussian distribution in order to capture conditional skewness and fat-tails. Barone-Adesi, Engle, and Mancini (2008) propose a more advanced GARCH option-pricing methodology with filtered historical innovations in order to allow for the physical and risk-neutral densities to be different and, hence, better match observed option prices. Christoffersen, Jacobs, Ornthanalai, and Wang (2008) build a new GARCH process that comprises of long-run and short-run variance components in order to enhance the flexibility of the GARCH models to explain option prices in practice.

However, all these studies rely on the implicit assumption that variance risk is not priced. Only recently Christoffersen, Heston, and Jacobs (2013) (henceforth CHJ) developed a new pricing kernel that incorporates variance risk. Based on equilibrium arguments they propose an affine-GARCH framework for pricing European options in closed-form that includes the approach of Heston and Nandi (2000) as a special case. They argue that with a variance-dependent kernel the model can produce much richer dynamics that are able to partially resolve several empirical anomalies that are crucial for option-pricing. We should also underline here that the availability of a closed-form solution for European option prices, which is usually an advantage of affine processes, is of primary importance since it significantly improves the computational efficiency of the model by avoiding complicated and time-demanding price approximations with simulation procedures.

Therefore, in this paper we choose to utilize the affine-GARCH process of Heston and Nandi
in order to capture the premium for variance risk and improve our ability to reconcile the implied physical and risk-neutral dynamics. In order to capture the properties of the pricing kernel we incorporate both spot and forward-looking information in a joint maximum likelihood estimation procedure. We implement the model using returns of the S&P500 index, but instead of simultaneously working with the wide cross-section of index options as in CHJ, we propose to use the daily variance dynamics inferred from the VIX index in order to approximate the implied risk-neutral variance process. Since the VIX index is estimated on a daily basis from the properly weighted cross-section of option prices written on the index, this technique is equivalent to using options and can produce similar results, while at the same time being more efficient and straightforward from a computational perspective. We find strong evidence that support the hypothesis of priced volatility risk, since the model fits the data much better after allowing for this more flexible pricing kernel process. Our findings indicate a U-shaped pattern for the logarithmic kernel, which can be very helpful in explaining option-pricing puzzles after projecting the impact of volatility risk on the return process. This is also in line with the semi-parametric evidence documented in CHJ regarding the implied shape of the kernel. An option-pricing calibration is carried out to demonstrate how these results can be used to generate a more realistic option-price structure across moneyness and maturity.

The paper is organized as follows. In the next section we present the model framework. We start by describing the properties of the physical and risk-neutral dynamics under standard preference assumptions and then we explain how this approach is enhanced by introducing variance-risk in the pricing kernel specification. We analyze the characteristics of the model and the intuition behind each parameterization for describing empirical anomalies. In section 3 we present the closed-form option-pricing methodology under the affine-GARCH dynamics, and in section 4 we show how to estimate the model by employing a joint maximum likelihood. Next, we discuss on the obtained results and we underline the importance of the more-flexible kernel process for reconciling the underlying joint dynamics of the data. We also demonstrate an option-pricing calibration of the model under the different parameter restrictions. In the last section we summarize and conclude, leaving some directions for future research.

2 The Model

Our analysis uses as a core model the affine GARCH representation introduced by Heston and Nandi (2000). One of the most important features of this model is that the affine dynamics allow for pricing European call and put options in closed-form. Furthermore, the model is able to reproduce several stylized properties of variance that are of utmost importance when valuing options. To be more specific, the parameters of the model are able to generate variance asymmetry through a negative conditional correlation structure between returns and variance, at the same time inducing negative risk-neutral skewness in the distribution of returns. The model also nests the symmetric case which is equivalent to the standard GARCH. The original
The HN–GARCH process has the following form:

\[
\begin{align*}
\ln(S_{t+1}) &= \ln(S_t) + r + \left(\mu - \frac{1}{2}\right)h_{t+1} + \sqrt{h_{t+1}}z_{t+1} \\
h_{t+1} &= \omega + \beta h_t + \alpha(z_t - \gamma \sqrt{h_t})^2
\end{align*}
\] (1) (2)

where \(r\) is the daily continuously compounded rate of return and \(z\) is a disturbance term assumed to have a standard normal distribution. In this setting, the parameter \(\gamma\) accounts for the asymmetric effect of shocks by introducing negative skewness and negative correlation between returns and volatility and \(\alpha\) controls the kurtosis of the distribution by driving the variance of variance. The parameter \(\mu\) captures the equity premium (per unit of risk) that must be priced and \(-\frac{1}{2}h_{t+1}\) is the Jensen’s adjustment term.

It is crucial for our analysis to study the statistical characteristics of this process. Taking unconditional expectations we can work-out the long-run level of variance:

\[
E[h_{t+1}] \equiv \bar{h} = \frac{\omega + \alpha}{1 - \beta - \alpha \gamma^2}
\] (3)

which allows us to express the expected conditional variance as a function of the unconditional variance:

\[
E_t[h_{t+2}] = (\beta + \alpha \gamma^2)h_{t+1} + (1 - \beta - \alpha \gamma^2)\bar{h} = \rho h_{t+1} + (1 - \rho)\bar{h}
\] (4)

where \(\rho = \beta + \alpha \gamma^2\) is the first-order autocorrelation of the variance process. In order for variance to be a stationary mean-reverting process with finite first and second moment, it must be the case that the persistence is less than unity, i.e. \(\rho < 1\). The conditional variance of variance is also linear in the current variance state:

\[
\text{var}_t(h_{t+2}) = 2\alpha^2 + 4\alpha^2 \gamma^2 h_{t+1}
\] (5)

It is obvious here that \(\alpha\) is the main driver of kurtosis and in the unlikely case of \(\alpha\) being zero the model suggests that variance is deterministic.

It is very important to understand the mechanism that generates asymmetry in the behavior of volatility. Let us first denote the daily log-returns as \(R_{t+1} = \ln\left(\frac{S_{t+1}}{S_t}\right)\). The conditional covariance between returns and variance is found to be:

\[
cov_t(R_{t+1}, h_{t+2}) = -2\alpha \gamma h_{t+1}
\] (6)

Since the parameters \(\omega, \alpha\) and \(\beta\) of the variance process are considered to be strictly positive in order for variance to be well-defined, this implies that the negative correlation between returns and variance that is prevalent in the stock markets can only arise with positive values of \(\gamma\). In other words, \(\gamma\) controls the magnitude of this volatility asymmetry or the “leverage effect”, which directly affects the skewness of the risk-neutral density.\(^6\) In this framework the asymmetry arises in a slightly different way compared to other asymmetric volatility models in the GARCH literature such as the E-GARCH of Nelson (1991) or the GJR-GARCH and the T-GARCH of Glosten et al. (1993) and Zakoian (1994). These models consider that volatility increases in

\(^6\)The “leverage effect” hypothesis was first documented by Black (1976).
both positive and negative shocks to returns, but with higher intensity for the negative shocks. However, we would say that the HN-GARCH is more closely related to the N-GARCH and the V-GARCH of Engle and Ng (1993). In this type of processes, volatility is not only affected in an asymmetric way by positive/negative shocks, but also is allowed to decrease when small positive shocks occur. It is easier to understand this asymmetric mechanism through a graphical representation of the news impact curves in figure 5.

It also follows from the formulas for the conditional variance and covariance that the conditional correlation of returns and variance is:

\[ \text{corr}_t(R_{t+1}, h_{t+2}) = \frac{-2\gamma\sqrt{h_{t+1}}}{\sqrt{2 + 4\gamma^2 h_{t+1}}} \]  

which again is negative for positive values of \( \gamma \). Notice here that the correlation in this particular GARCH model is also time-varying, driven by the conditional level of variance. It is also easy to verify the fact that the standard symmetric GARCH is nested in this model. When the asymmetry parameter \( \gamma \) is zero, the model assumes no correlation between returns and variance. This means under symmetry both positive and negative shocks have the same positive impact on variance. Furthermore, the variance of variance becomes constant since it is not affected by the current variance state any more. In our analysis we will refer to the symmetric-GARCH model as \( G \) and to the asymmetric Heston-Nandi affine GARCH as \( HN \), which will be more convenient for illustration purposes.

2.1 Risk-Neutral Dynamics

Following the locally risk-neutral valuation relation (LRNVR) pioneered by Duan (1995) one may express the risk-neutral dynamics of the GARCH process as:

\[
\ln(S_{t+1}) = \ln(S_t) + r - \frac{1}{2} h_{t+1} + \sqrt{h_{t+1}} z^*_{t+1}
\]  

\[ h_{t+1} = \omega + \beta h_t + \alpha (z^*_{t} - \gamma^* \sqrt{h_t})^2 \]  

while by rearranging the original physical process we can see that:

\[ z^*_t = z_t + \mu \sqrt{h_t} \]  

\[ \gamma^* = \gamma + \mu \]  

In this framework and under standard preference assumptions the risk-neutral density of the shock remains standard normal after the change of measure but with a different mean. The distribution is linear in the current volatility state and is shifted by a term that depends on the level of the equity risk-premium \( \mu \). We should underline here that the conditional variance is the same under both physical and risk-neutral measures, which is also a critical point in applying the LRNVR. However, this does not hold for variance forecasts at longer horizons but only applies for time \( t + 1 \) conditional at time \( t \). The model also implies a different degree of asymmetry under the two measures. Since the equity risk-premium is expected to be positive, i.e. \( \mu > 0 \), this means that \( \gamma^* > \gamma \), which produces more negatively skewed risk-neutral
distributions and suggests a stronger negative correlation between returns and variance under risk-neutral expectations. Even if we consider the physical return process to be symmetric, there is still a small degree of asymmetry in the risk-neutral dynamics that arises due to a non-zero equity premium.

Defining the risk-neutral dynamics is of primary importance when trying to value options under risk-neutral expectations. Thus we need an appropriate pricing kernel to change from the physical to the risk-neutral measure. The transformation of the Heston and Nandi (2000) model into its risk-neutral expression implicitly assumes the existence of power pricing kernel of the form:

\[ \tilde{M}_{t+1} = \frac{M_{t+1}}{M_t} = \left( \frac{S_{t+1}}{S_t} \right)^\phi e^\delta \]

where \( \delta \) is a time-preference parameter that controls for subjective discounting and \( \phi \) is the coefficient that captures risk-aversion behavior. For \( \phi < 0 \) the marginal utility is decreasing in returns. This pricing kernel is actually log-linear in returns:

\[ \ln(\tilde{M}_{t+1}) = \delta + \phi R_{t+1} \]

which may be too restrictive when trying to explain empirical puzzles in option prices. This power pricing kernel of Rubinstein (1976) is also used in the GARCH option pricing framework of Duan (1995) and is actually equivalent to the kernel implied by the standard one-period Black-Scholes model.

### 2.2 A Variance-Dependent Pricing Kernel

In order to overcome the limitations of the log-linear pricing kernel and better capture the empirical irregularities in option prices, Christoffersen, Heston, and Jacobs (2013) introduce a new pricing kernel that is able to take into account variance risk premia as well. In discrete-time the pricing kernel can be expressed as:

\[ \tilde{M}_{t+1} = \frac{M_{t+1}}{M_t} = \left( \frac{S_{t+1}}{S_t} \right)^\phi e^{\delta + \eta h_{t+1} + \xi (h_{t+2} - h_{t+1})} \]

where \( \eta \) is another time-preference parameter and \( \xi \) is to capture aversion to variance risk. This means that risk-premia may now arise from two distinct sources. The first source of risk is related to equity risk and is driven by the risk-aversion coefficient \( \phi \) and the other one is related to uncertainty about changes in variance and is captured by \( \xi \). In contrast to equity risk-aversion which suggests that the pricing kernel is decreasing in returns (\( \phi < 0 \)), we expect the sign of variance risk coefficient to be positive (\( \xi > 0 \)), highlighting that the pricing kernel is an increasing function of variance.\(^7\) This is in line with the well-documented negative variance risk premium, which indicates that investors are willing to bear negative excess returns in order to hedge against states of nature with higher variance uncertainty.\(^8\)

It is shown in Christoffersen, Heston, and Jacobs (2013) that the logarithmic transformation

\(^7\)According to Christoffersen, Heston, and Jacobs (2013) even in the extreme case of \( \xi \) being equal to zero, a negative variance risk premium may arise due to negative correlation between returns and variance.

\(^8\)See Carr and Wu (2009).
of this pricing kernel takes a quadratic form with respect to returns:

\[
\ln(\tilde{M}_{t+1}) = \frac{\alpha\xi}{h_{t+1}}(R_{t+1} - r)^2 - \mu(R_{t+1} - r) + \left(\eta + (\beta - 1)\xi + \alpha\xi\left(\mu - \frac{1}{2} + \gamma^2\right)\right)h_{t+1} + \delta + \omega\xi + \phi r
\]  

(15)

This reveals that projecting the kernel on the returns can generate a U-shaped pricing kernel for \(\xi > 0\). In this case, i.e. in the case of a negative variance risk premium, the kernel assigns more weight on the tails of the distribution of shocks, which helps in better explaining the empirically observed differences between physical and risk-neutral volatility smiles (Bates (1996)). Figures in 12 demonstrate the parabolic shape of the log pricing kernel allowing for a negative variance risk premium in contrast to the previous log-linear kernel in eq. 12.

Furthermore, Christoffersen, Heston, and Jacobs (2013) have proved that considering this new variance-dependent pricing kernel in an equilibrium setting produces risk-neutral dynamics for returns and conditional variance that remain in the same affine-Garch class originally described by HN. However, the parameters and the conditional variance of the risk-neutral process will now be different than those of the simple risk-neutral HN process:

\[
\ln(S_{t+1}) = \ln(S_t) + r - \frac{1}{2} h_{t+1}^* + \sqrt{h_{t+1}^*} z_{t+1}^*
\]  

(16)

\[
h_{t+1}^* = \omega^* + \beta h_{t+1}^* + \alpha^* \left(z_{t+1}^* - \gamma^* \sqrt{h_t^*}\right)^2
\]  

(17)

In this reformulation the risk-neutral shock \(z^*\) is still linked to the physical shock \(z\) but now not only its mean is different, but also its variance:

\[
z_{t+1}^* = \sqrt{1 - 2\alpha\xi} \left(z_t + \left(\mu + \frac{\alpha\xi}{1 - 2\alpha\xi}\right) \sqrt{h_t}\right)
\]  

(18)

Under the risk-neutral measure the conditional variance is a scaled version of the physical variance:

\[
h_{t+1}^* = h_{t+1}/(1 - 2\alpha\xi) = \tilde{\xi} h_{t+1}
\]  

(19)

where \(\tilde{\xi} = (1 - 2\alpha\xi)^{-1}\) is a scaling factor that drives a wedge between the conditional variances under the two different measures. As we have seen, the case of a negative variance risk premium corresponds to \(\xi > 0\), and since \(\alpha > 0\) this is equivalent to a scaling factor \(\tilde{\xi} > 1\). This suggests that the variance process under the risk-neutral measure will exceed the physical variance, which is another stylized fact of risk-neutral dynamics. The parameters of the risk-neutral variance process are also driven by the variance risk coefficient:

\[
\omega^* = \omega/(1 - 2\alpha\xi) = \tilde{\xi}\omega
\]

\[
\alpha^* = \alpha/(1 - 2\alpha\xi)^2 = \tilde{\xi}^2\alpha
\]

\[
\gamma^* = \gamma - \phi
\]  

(20)

where the following mapping must hold for the parameters of the pricing kernel in order for this
expression to be consistent with the affine garch dynamics that we have discussed:

\[
\phi = - \left( \mu - \frac{1}{2} + \gamma \right) (1 - 2\alpha \xi) + \gamma - \frac{1}{2} \\
\delta = -(\phi + 1)r - \xi \omega + \frac{1}{2} \ln(1 - 2\alpha \xi) \\
\eta = - \left( \mu - \frac{1}{2} \right) \phi - \xi \alpha \gamma^2 + (1 - \beta)\xi - \frac{\left( \phi - 2\alpha \gamma \xi \right)^2}{2(1 - 2\alpha \xi)}
\]

(21)

With these parameter specifications the model obtains some very interesting implications. First of all, it is obvious that this framework nests that of Heston and Nandi (2000) since in the absence of variance risk premium, i.e. for \( \xi = 0 \), the kernel reduces to the log-linear case that we described in the previous section. Moreover, the risk-neutral coefficients \( \omega^* \) and \( \alpha^* \) remain the same with those of the physical process as before since the scaling factor is \( \tilde{\xi} = 1 \) and the asymmetry parameter \( \gamma^* \) becomes again equal to \( \gamma + \mu \) as in eq. 11, since the risk-aversion coefficient \( \phi \) from 21 reduces to \( -\mu \). Similarly, the risk-neutral shock distribution is also not considered to have a different variance than the distribution of physical shocks as in 10, since the conditional variance in 19 is unaffected.

However, if we consider that the variance risk enters the pricing kernel the characteristics of the risk-neutral dynamics become more realistic and may help explain empirical phenomena in option pricing. Starting from the fact that equity premium suggests a positive value for \( \mu \), this indicates that \( \phi < 0 \), which is in line with a declining marginal utility of wealth. Since the kernel is monotonic in variance, a negative variance risk premium consistent with literature would require that \( \xi > 0 \) or \( \tilde{\xi} > 1 \). A quick glimpse reveals that the risk-neutral parameters \( \omega^* \) and \( \alpha^* \) are higher from their physical counterparts, i.e. \( \omega^* > \omega \) and \( \alpha^* > \alpha \). These, taken together with the fact that returns and variance are negatively correlated, which is captured by \( \gamma > 0 \), highlight some useful properties about the dynamics under the two densities.

First of all, as we have already described, the conditional one-step ahead dynamics of the two measures are no longer considered to be identical, but they are now allowed to differ. Consistent with empirical findings, the risk-neutral variance always exceeds the physical variance and this wedge between them is proportional to the level of the variance risk premium. This also holds for the level of the unconditional long-run variance \( \tilde{h}^* = (\omega^* + \alpha^*)/(1 - \rho^*) \). Since \( \rho^* = \beta + \alpha^* \gamma^2 \) is the first-order autocorrelation of the risk-neutral variance, it follows that the risk-neutral variance process will be more persistent \((\rho^* > \rho)\). Similarly, the drift of the risk-neutral variance is also higher than that of the physical process taking conditional risk-neutral expectations \( E_t^*[h_{t+2}^*] = \rho^* h_{t+1}^* + (1 - \rho^*) \tilde{h}^* \). The existence of a non-negative variance risk premium also affects the variance of the risk-neutral variance \( \text{var}^*[h_{t+2}^*] = 2\alpha^* \gamma^2 + 4\alpha^* \gamma^2 h_{t+1}^* \), which makes it higher than the variance of physical variance. It also seems that the asymmetry of the risk-neutral density will be more prevalent than that of the physical density, since the returns and the risk-neutral variance are more negatively correlated, which results from the fact that \( \text{cov}^*[R_{t+1}, h_{t+2}^*] = -2\alpha^* \gamma^* h_{t+1}^* \) is also increased in magnitude.

In our empirical analysis we will try to verify the validity and the implications of the intuition discussed above. If the hypothesis of a significant negative variance risk premium holds, then the model that we have just presented is able to reproduce several patterns observed in empirical
option-pricing studies. By capturing a higher level of risk-neutral variance compared to that of physical variance we are able to explain the stylized property of the implied variance from options exceeding future realized variance, a finding which has made the short-selling of straddles a quite attractive option investment strategy. The model generates a variance process that is also more persistent and more negatively correlated with returns under the risk-neutral measure, producing a higher degree of asymmetry and inducing stronger negative skewness and fatter tails for the risk-neutral density. This is crucial in resolving the mispricing of long-term and deep out of the money options. Utilizing this parametric framework that links the GARCH dynamics with equilibrium option prices makes it feasible to reproduce the empirically observed U-shape of the log pricing kernel by projecting the kernel on the returns.

2.3 Variance Impulse Response and Variance Term-Structure

Starting from the affine GARCH dynamics of variance we may substitute out parameter \( \omega \) from the unconditional variance \( \bar{h} \) expression and rewrite the process as:

\[
 h_{t+1} = \omega + \beta h_t + \alpha (z_t - \gamma \sqrt{h_t})^2 \\
= \bar{h} + \beta (h_t - \bar{h}) + \alpha \left( (z_t - \gamma \sqrt{\bar{h}})^2 - 1 - \gamma^2 \bar{h} \right). 
\]

Rearranging and expanding the square term in the parenthesis and also substituting for the variance persistence parameter \( \rho = \beta + \alpha \gamma^2 \) gives:

\[
 h_{t+1} = \bar{h} + \rho (h_t - \bar{h}) + \alpha (z_t^2 - 1 - 2 \gamma \sqrt{h_t} z_t) \]

where we can see from the last term that the conditional expectation of the innovations to variance is zero. From this we can derive the conditional variance forecast for any \( k \)-day horizon:

\[
 E_t[h_{t+k}] = \bar{h} + \rho^k (h_t - \bar{h}) = \bar{h} + \rho^{k-1} (h_{t+1} - \bar{h}) 
\]

since \( h_{t+1} \) can also be estimated at time \( t \). As in all autoregressive processes the higher the variance persistence, the longer the time-period required for variance to revert back to its long-run unconditional level.

Since we have an estimate of the multi-step expectation of conditional variance, we can easily define the term-structure of variance for a specific maturity \( T \) in the following way:

\[
 h_{t+1:t+T} = \frac{1}{T} \sum_{k=1}^{T} E_t[h_{t+k}] \\
= \frac{1}{T} \sum_{k=1}^{T} \left( \bar{h} + \rho^{k-1} (h_{t+1} - \bar{h}) \right) \\
= \bar{h} + \frac{1 - \rho^T}{1 - \rho} \frac{(h_{t+1} - \bar{h})}{T} 
\]

with \( \sum_{k=1}^{T} \rho^{k-1} \) being a geometric series. The term-structure of volatility is obtained by taking the square root of the previous expression. Furthermore, in case we wanted to compare the
variance term structure of several models under different specifications it would be more convenient to scale the term-structure $h_{t+1:T}$ by the model’s unconditional variance $\bar{h}$, each time starting from an initial variance above or below the unconditional level.

At this point, it is also crucial to capture the impact of a return shock $z$ on variance. Hence, we can define the variance impulse response function that represents the effect of shocks on expected future variance. Starting from the conditional variance expression in 24, we can represent the GARCH dynamics as an ARCH($\infty$) process:

$$h_{t+1} = \bar{h} + \alpha \sum_{i=0}^{\infty} \rho^i \left( z_{t-i}^2 - 1 - 2\gamma \sqrt{h_{t-i}} z_{t-i} \right)$$

(29)

Based on this representation we can write the k-day ahead impulse response of variance to a specific shock $z$ relative to its unconditional level as follows:

$$vir(k) = \alpha \rho^k (z^2 - 1 - 2\gamma \sqrt{\bar{h}} z) / \bar{h}$$

(30)

Plotting the instantaneous impact of a shock on variance corresponds to the so-called “news impact curve”, with $z$ being the only stochastic parameter in a GARCH-type framework that is driven by information.

Another way to quantify the effect of a shock today on future variance at $k$-steps ahead is by taking the partial derivative of the conditional variance in 29 with respect to the shock $z^2$, which gives:

$$\frac{\partial [E_t[h_{t+k}]]}{\partial z^2} = \rho^{k-1} \alpha (1 - \gamma \sqrt{h_t} / z_t)$$

(31)

This applies similarly for the impact on the whole term-structure of variance:

$$\frac{\partial [E_t[h_{t+1:T}]]}{\partial z^2} = \frac{1 - \rho^T \alpha}{1 - \rho T} (1 - \gamma \sqrt{h_t} / z_t)$$

(32)

3 Option Pricing

One of the most important advantages of the model is that it remains in the affine garch class even under changing from physical to risk-neutral measure. This feature will be particularly useful when pricing European options as we will see. As a first step, we can proceed to find the expression for the moment generating function of the affine risk-neutral dynamics. At a next step, we will be able to value a European-style call option by discounting the expected payoff to a call-option under risk-neutral expectations:

$$C = e^{-r(T-t)} E^*_t [\max(S_T - X, 0)]$$

(33)

where $C = C(S_t, h^*_t, X, T)$ is a function of the strike price $X$, the time to maturity $T$, the conditional risk-neutral variance $h^*_t$ and the current stock price $S_t$. We can write the conditional moment generating function (MGF) of the risk-neutral dynamics as a log-linear expression of
the following form:

\[ g^*_{t,T}(\varphi) \equiv E^*[e^{\varphi \ln(S_{t+1})}] \]

\[ = e^{\varphi \ln(S_t) + A_{t,T}(\varphi) + B_{t,T}(\varphi) h^*_{t+1}} \]  

(34)

\[ = S^*_t e^{A_{t,T}(\varphi) + B_{t,T}(\varphi) h^*_{t+1}} \]  

(35)

where the coefficients \( A_{t,T} \) and \( B_{t,T} \) can be estimated recursively. Since the spot stock price at expiration \( S_T \) is only known at time \( T \), the above parameters of the MGF have to satisfy the following terminal condition:

\[ A_{T,T}(\varphi) = B_{T,T}(\varphi) = 0. \]  

(37)

Starting from this terminal condition, we can solve for the coefficients \( A_{t,T} \) and \( B_{t,T} \) at time \( t \) based on the following difference equations:

\[ A_{t,T}(\varphi) = A_{t+1,T}(\varphi) + \varphi r + \omega^* B_{t+1,T}(\varphi) - \frac{1}{2} \ln(1 - 2\alpha^* B_{t+1,T}(\varphi)) \]  

(38)

\[ B_{t,T}(\varphi) = -\frac{1}{2} \varphi^2 + 2\alpha^* \gamma^* B_{t+1,T}(\varphi) \left( \alpha^* \gamma^* B_{t+1,T}(\varphi) - \varphi \right) \]  

(39)

See Heston and Nandi (2000) for a derivation of the previous recursion formulas. Based on the MGF of the log spot price under the risk-neutral measure \( g^*(\varphi) \), we can use \( g^*(i\varphi) \) to denote the characteristic function. By inverting the characteristic function we can estimate the associated probabilities under both physical and risk-neutral expectations.\(^9\) Given the affine nature of the model, one can obtain a quasi-closed-form solution for pricing European options. Thus, under the assumption of no dividends we may rewrite the price of a European call option in \( 33 \) as:

\[ C = C(S_t, h^*_{t+1}, X, T) = e^{-r(T-t)} E_t^*[\max(S_T - X, 0)] \]

\[ = S_t P_{1,t} - e^{-r(T-t)} X P_{2,t} \]  

(40)

where the risk-neutral probabilities \( P_1 \) and \( P_2 \) can be estimated via numerical integration based on the following expressions:

\[ P_{1,t} = \left( \frac{1}{2} + \frac{e^{-r(T-t)}}{\pi} \int_0^\infty \text{Re} \left[ \frac{X^{-i\varphi} g^*_{t,T}(i\varphi + 1)}{i\varphi S_t} \right] d\varphi \right) \]  

(41)

\[ P_{2,t} = \left( \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \text{Re} \left[ \frac{X^{-i\varphi} g^*_{t,T}(i\varphi)}{i\varphi} \right] d\varphi \right) \]  

(42)

where \( \text{Re} \) represents the real part of a complex number. The intuition that lies within this pricing equation is similar to that of the standard Black-Scholes formula and thus the call price can be estimated by the current stock price times a probability \( P_1 \) minus the discounted (at the risk-free rate) strike price times a probability \( P_2 \). Probability \( P_1 \) can be interpreted as the delta of the option, i.e. the sensitivity of the call price with respect to changes in the underlying

asset price $S$, and $P_2$ as the risk-neutral probability of the option being exercised at maturity, i.e. the terminal stock price being greater than the strike price. The most important difference between this approach and the Black-Scholes formula is that the BS formula assumes a constant variance and uses only the spot asset price in the estimation. On the contrary, the GARCH option pricing formula is based on the conditional risk-neutral variance process that requires a path of observed asset prices for the underlying asset. Since a closed-form solution is available for European call options, it easily follows from put-call parity that the price of a European put option with the same strike and maturity equals:

$$P = C - S_t + X e^{-r(T-t)}$$  \hspace{1cm} (43)

We should also note here that one can further compute the sensitivity of the option price with respect to changes in the option delta (Gamma) or changes in the variance of the underlying (Vega) by differentiating the expressions 41 and 40 for the risk-neutral probability $P_1$ and the option price, respectively.

4 Data & Estimations

The estimations presented here are based on a 14-year sample of daily closing prices of the S&P500 index and the VIX implied volatility index, starting from 2000. This means that our sample comprises of a total of 3520 observations. Some statistical properties of the two time series are presented in table 1. Since VIX reflects the market-based annualized implied volatility, we adjust the VIX level to obtain a proxy for the daily implied risk-neutral variance:

$$v_t = \frac{1}{252} \left( \frac{VIX_t}{100} \right)^2$$  \hspace{1cm} (44)

As we have already described, the estimations are based on the HN–GARCH that nests the symmetric GARCH model. Motivated by the recent work of CHJ, we extend the properties of the underlying dynamics by enhancing the pricing kernel specification in such a way that it allows for a variance risk premium. We will estimate the parameters of the model in a joint likelihood framework that captures information from both returns and implied volatilities. We will show that the joint likelihood estimation further improves the results and helps the models to better describe the characteristics of both processes, at the same time producing more realistic option-price patterns.

4.1 A Joint Maximum Likelihood Estimation

The original likelihood function that we aim to maximize has the following general form:

$$L(\theta|\{z_t\}) = \prod_{t=1}^{T} f(z_t, \theta)$$  \hspace{1cm} (45)

A vector $\theta^*$ is estimated via a non-linear optimization in order to maximize the above likelihood function for a given distribution of $z$. 

14
Assuming that the innovation term of returns $\sqrt{h_t}z_{t+1}$ is conditionally normally distributed as $N \sim (0, h_t)$ this means that maximizing the likelihood function above is equivalent to maximizing its logarithm that takes the following form:

$$
\log \mathcal{L}^R = -\frac{T}{2} \log(2\pi) + \sum_{t=1}^{T} \left[ -\frac{1}{2} \log(h_t) - \frac{1}{2} z_t^2 \right] \quad (46)
$$

where $z_t = \frac{1}{\sqrt{h_t}}[r_t - r - \lambda h_t]$ is the zero-mean shock to returns. The standard approach in the literature suggests one estimates a GARCH model with return data only by maximizing this likelihood function.

We can construct a joint maximum likelihood estimation by utilizing VIX data simultaneously. We can assume that the difference between two volatility processes is normally distributed with sample variance $\hat{s}^2$. Here we consider that the model-free implied variance proxy $v_t$ is derived from the VIX directly and that the conditional risk-neutral variance $h^*_t$ is implied by the GARCH process. The log-likelihood for variance will be:

$$
\log \mathcal{L}^V = -\frac{T}{2} \log(2\pi \hat{s}^2) - \frac{1}{2\hat{s}^2} \sum_{t=1}^{T} (v_t - h^*_t)^2 \quad (47)
$$

We should note here that we have decided to use directly the VIX as the market-implied risk-neutral variance in the above likelihood function instead of indirectly working the cross-section of index options as in CHJ. However, there can be several other approaches to the objective function. For instance Christoffersen, Heston, and Jacobs (2013) as well as Kanniainen et al. (2014) use Black-Scholes Vega-weighted option valuation errors. Another approach would be to simply use dollar mean square error (MSE) of option prices as a loss function in the MLE estimation. See Christoffersen, Jacobs, and Ornthanalai (2013) for a more detailed discussion about the objective loss functions that can be used in GARCH option-pricing. Generally, it has been shown that using implied risk-neutral variance errors is almost equivalent to using vega-weighted option pricing errors and hence we expect comparable results. Hao and Zhang (2013) also consider a joint-likelihood estimation with returns and the VIX, but they only focus on fitting and comparing the risk-neutral dynamics for several GARCH specifications under a log-linear pricing kernel. Our approach is quite different since we simultaneously fit the physical dynamics of returns and the risk-neutral dynamics of variance at the same time taking into consideration a variance-dependent kernel in order to capture variance risk premium. Finally, the joint likelihood function that we want to maximize takes the following simple form:

$$
\log \mathcal{L} = \log \mathcal{L}^R + \log \mathcal{L}^V \quad (48)
$$

We first estimate the model under the standard pricing-kernel specification that only allows for an equity risk premium. We consider both a symmetric (G) and an asymmetric (HN) case for the GARCH specification. Each model is estimated in three different ways; (i) by maximizing the log-likelihood of the returns process, (ii) by maximizing the log-likelihood that arises from the dynamics of risk-neutral variance, or (iii) by maximizing their joint log-likelihood function. Finally, we extend the estimation framework to allow for a non-negative variance risk-premium
via the more flexible pricing-kernel approach as in CHJ. In the next section we discuss on the obtained results, summarized in Table 2.

5 Results & Comments

A quick glimpse on the maximized likelihood results reveals that the model fit improves dramatically as we move from the symmetric (G) case ($\gamma = 0$) to the asymmetric (HN) case, which underlines the existence of a sizable asymmetry in the behavior of volatility.\footnote{Further Likelihood-Ratio (LR) Tests have been carried out to examine the validity of the restriction on the asymmetry parameter $\gamma$. Results strongly indicate that the extra parameter for the asymmetry is required at all conventional levels of significance.} Even if the model parameters are estimated by maximizing only the return-based likelihood component as in almost all GARCH applications, we can see that the HN model is able to improve not only the fit for returns, but also the fit for volatility. Since it is very important to accurately capture the variance dynamics, and in order to examine which model can provide an adequate fit for volatility irrespective of that for returns, we perform the same estimations this time maximizing the variance-driven likelihood component. We observe that for the symmetric case the variance likelihood slightly improves as expected, but the total fit of the model is damaged due to the fact that the parameters (especially $\mu$) are unreasonably distorted in order to match the volatility process, and can no longer explain the properties of returns (this is also in line with the findings of Hao and Zhang (2013)). However, when estimating the asymmetric model solely on variance, we find that the model again improves the fit for variance, while at the same time it maintains a decent fit for the returns process, resulting to an even higher total likelihood which is also higher than the likelihood obtained by a joint maximization on the symmetric model. This implies that a joint likelihood maximization under the asymmetric model can further improve the overall fit of the system and drive the estimated parameter values to plausible levels, which is indeed justified by the output of the joint MLE on HN. Since joint likelihood maximization yields improved results for the total fit of the system, while at the same time requiring the same number of parameters to be estimated, it is always preferable to the other two single MLE approaches (see also Kanniainen et al. (2014)).

If we further relax the restriction that requires the variance risk factor $\xi$ to be zero (or $\tilde{\xi} = 1$) and essentially allow for richer kernel dynamics as in CHJ (see the last column of table 2), we observe that the total likelihood obtained from the joint MLE is again significantly improved. It is crucial to underline that now the model does not have to sacrifice the fit for any of the two processes since the results are approximately similar to those obtained from maximizing the two distinct likelihood components separately in the HN case. The improvement of the CHJ model is due to the risk-adjustment of the variance process when moving from the physical to the risk-neutral measure, which is achieved via properly pricing variance risk in the kernel specification. Ignoring the premium for variance risk is equivalent to trying to match the conditional physical variance with the conditional (model-free) risk-neutral variance implied from options, which is too restrictive and not in line with the heavily-documented observation in the literature that the two variance processes differ. By estimating a scaling factor $\tilde{\xi}$, the CHJ model is able to capture the stylized property of the risk-neutral variance exceeding the physical variance
and hence can better explain the joint dynamics of physical returns and implied volatilities simultaneously. Generally, our results are directly comparable with those of Christoffersen, Heston, and Jacobs (2013) who use option-valuation errors instead of implied variance errors, but we find the variance scaling parameter to be slightly higher. This could be primarily due to the fact that we are using a more recent data sample and it could indicate that the variance risk premium is becoming more pronounced lately, also leading to increased straddle returns.

The gradual improvement of the model fit as we relax the restrictions regarding asymmetry and variance risk is also revealed graphically from figures 2, 3 and 4. For the symmetric GARCH case the conditional variance path performs quite bad in tracking the daily-adjusted implied volatility index and cannot reproduce the significant asymmetric properties or capture the pronounced jumps in volatility. The fit of the HN model that allows for a leverage effect seems to be somewhat improved (figure 2), yet not adequate. As we have described above, considering a joint maximum likelihood estimation can help significantly improve the fit for the variance process along with that of returns, which can be seen from figure 3. The HN model appears to do a fairly good job in capturing the big volatility jumps during the recent financial crisis when a joint estimation is employed. However, enhancing the pricing kernel process to include a variance risk factor (CHJ model) produces a risk-neutral variance path that explains even more accurately the model-free VIX implied variance, since it relaxes the restrictive requirement of physical and risk-neutral variance being identical conditionally and hence it is able to capture a higher variance level under the risk-neutral measure (black line) compared to the level of variance under the physical measure (grey line).

A closer look on the obtained physical and risk-neutral parameters validates the hypotheses that we discussed in the methodology section. We see that even if the physical variance process is considered to be symmetric (G), there can still arise a small degree of asymmetry under the risk-neutral measure, which stems from the fact that we have considered a non-zero premium for equity risk $\mu$ that drives $\gamma^*$. Similarly, for the HN model the asymmetry increases even further under risk-neutral expectations when there exists a positive premium for equity risk (or equivalently the marginal utility is decreasing, $\phi < 0$). Taken together with the fact that for the G and the HN cases the remaining parameters of the GARCH process are the same under both measures, this indicates that the variance process will be more persistent under risk-neutral expectations. Furthermore, the unconditional mean and variance of the risk-neutral variance process will also be higher than that of the physical variance process. Similarly, the unconditional correlation between returns and risk-neutral variance is more negative than that between returns and physical variance, which again is driven by the stronger asymmetry under risk-neutral expectations.

For the CHJ model the interpretation is slightly more complicated because the risk-neutral parameters are now driven by the variance risk premium parameter as well. The degree of asymmetry is no longer required to be higher under the risk-neutral dynamics, since the risk-preference parameter $\phi$ now depends on both the equity and variance risk premium factors $\mu$ and $\xi$. The variance risk factor is found to be greater than zero, suggesting a scaling parameter $\tilde{\xi}$ that is higher than unity and statistically significant. As we have already analyzed, this scalar drives the conditional risk-neutral variance at a higher level compared to its conditional physical

17
counterpart. Furthermore, it increases the value of the risk-neutral parameters $\alpha^*$ and $\omega^*$. With these parameters being higher than the corresponding physical parameters, the resulting risk-neutral variance process under the CHJ model is more persistent than the physical variance process and with a higher unconditional mean and variance. The CHJ unconditional correlation between returns and variance also seems to be stronger under risk-neutral expectations.

In order to get a better understanding of the different behavior among the models, it is crucial to graphically demonstrate the impact of a shock on variance under all specifications. We first plot the news impact curves for the 3 models in 5 using the obtained parameters from the joint maximum likelihood estimations. The symmetric nature of the G model is salient since both positive and negative shocks always increase variance by the same percentage. For the HN and the CHJ models the news impact curve is shifted not only to the right but also downwards in order to capture the asymmetric effect of shocks on physical variance. Even though the degree of asymmetry is higher under the CHJ parameterization, the resulting impact on variance is smoother than that suggested by HN. We find that, despite the stronger asymmetry in the CHJ case, positive(negative) shocks tend to decrease(increase) volatility by less compared to the effect of a similar shock under the HN.

Plotting the instantaneous impact of a return shock on risk-neutral variance is even more important, especially for the CHJ specification where we expect shocks to exhibit a different behavior under the two measures. In the presence of a variance risk premium, the risk-neutral density differs from the physical density not only in terms of mean but also in terms of volatility. Figure 6 shows that the news-impact curve for risk-neutral variance is right-shifted as expected, suggesting a higher risk-neutral drift, and also reveals that the instantaneous effects of shocks on the risk-neutral variance process are less pronounced compared to the effects on physical variance. This observation can be quite important when it comes to implementing the model for option-pricing purposes.

In Figure 7 we show the 200-step ahead variance forecasts for the models starting from an initial variance level that is four times above and below the level of the long-run unconditional variance. Plots are normalized by the unconditional variance of each model in order to achieve comparability. We can see that the CHJ model is able to produce a more persistent physical variance process. Since the models demonstrate a somewhat different attitude towards shocks, it can be the case that a similar shock may have a stronger impact on variance for a particular model, but this shock may also die out relatively faster, and vice versa. The variance impulse responses in figure 8 show how a shock of prespecified magnitude gradually fades away along a k-day horizon. We consider positive and negative shocks of 1,3 and 5 standard deviations and we plot the variance impulse responses as a percentage change relative to each model’s unconditional variance. Note again that under volatility asymmetry the models can reduce volatility after a positive shock, which is not the case for the G model. We also see that for HN shocks have a more pronounced effect on variance than for CHJ, but they die out faster due to lower persistence.

The different level of persistence and news impact curve of the CHJ risk-neutral variance, yields an interesting comparison between the k-day impulse response functions of both physical and risk-neutral variances (Figure 9). We find that for small positive/negative shocks the effect
on physical variance is higher. However, as we underlined before, the risk-neutral variance process is more persistent than the physical variance, resulting to a slower decay of return shock effects on risk-neutral variance. For shocks of higher magnitude the impact is quite mixed since we find large positive shocks to reduce risk-neutral volatility by more than they reduce physical volatility, which can also be seen from the news impact curves in 6.

Plotting the physical and risk-neutral variance term-structures uncovers the superiority of the CHJ approach in terms of relaxing the strain that requires the conditional variance under the two measures to be identical. Under the assumption of a log-linear pricing kernel (in G and HN cases) the conditional error distribution has the same variance after changing from the physical to the risk-neutral measure. Practically, this results in considering that at time $t+1$ the two conditional variances are the same, and they can only diverge for longer-term forecasts (see Figure 10). Again, for illustration purposes we start from an initial volatility level four times above/below the long-run (annualized) volatility level. Introducing a variance risk premium in the pricing kernel specification (in the CHJ model) essentially drives a wedge between the two conditional variances (Figure 11). The conditional risk-neutral variance is now allowed to differ and is always higher than the physical variance, which is more realistic and also in line with empirical observations.

5.1 Implied Kernel Dynamics and Option-Price Calibration

Before proceeding to an analysis of how the models calibrate to option prices, it is crucial to visualize and explain the implied pricing kernel patterns that translate the physical dynamics to risk-neutral dynamics. Therefore, in Figure 12 we plot the logarithmic pricing kernel described by equation 15. We consider 3 different maturity horizons (1 day, 30 days and 90 days) and we illustrate the resulting pricing kernels suggested by the model parameters obtained by the MLEs. As we have shown, in the absence of a variance risk-premium the pricing kernel collapses to the standard power kernel of equation 12, as in Brennan (1979) and Rubinstein (1976).

Here, we show the log-linear kernel produced by the G and HN models. Note that the kernel becomes steeper if we consider the parameters obtained by the joint maximum likelihood estimations that utilize both returns and risk-neutral variance. This reveals that there is valuable information embedded in option prices that should not be ignored when trying to specify the implied kernel dynamics. Thus, incorporating implied-volatility information within a joint likelihood framework should be superior in properly capturing the shape of the pricing-kernel and in accurately pricing options.

However, the power kernel is very restrictive and cannot address most of the empirical irregularities in option prices. We observe that allowing for a non-zero variance risk premium in the pricing kernel specification generates a U-shaped pattern. This shape is in also line with the semi-parametric evidence recently presented by Christoffersen, Heston, and Jacobs (2013). By plotting the ratio of the logarithmic risk-neutral to physical density they reject the assumption of a monotonic kernel and they uncover a U-shaped relation between the two densities that is stable across different subsamples. Our estimation procedure seems to be able to reproduce this empirical observation of a non-monotonic logarithmic kernel. Even though the kernel appears to be monotonic in both returns and variance, its logarithmic representation in 15 is
a quadratic function of returns after projecting the impact of variance on returns. Generally, the unique properties of this specification can contribute significantly towards resolving several robust option-pricing anomalies, such as the returns to option straddles and the mispricing of long-term and deep-out-of-the-money options.

For the calibration exercise we consider put and call options with maturity of 30-days and a strike price of 100$. We fix the annual risk-free rate to 3% and we let the moneyness to range between 0.9 and 1.1. We first compare the different option-price patterns for the symmetric (G) and the asymmetric (HN) cases using the parameters obtained by a standard return-based MLE (Figure 13). We verify that under the asymmetric model the out-of-the-money put options (and also the in-the-money calls) are more expensive. This is more intuitive in terms of explaining the empirically observed profits to short strategies of OTM put options and also consistent with the pronounced implied volatility smirks. Thus, incorporating asymmetric volatility effects in a model is very important for producing theoretical option prices that are closer to those of the market. Following Heston (1993), researchers have underlined the importance of allowing for a negative correlation structure between innovations to returns and volatility when pricing options. Bakshi et al. (1997) and Pan (2002) among others also emphasize the role of jumps for option-pricing, not only in explaining the observed volatility smiles and but also in reconciling the implied dynamics of both returns and variance.

As a next step we show how the option prices change if we employ a joint maximum likelihood estimation. From figure 14 we can see that the option premia are shifted upwards at all moneyness levels. This suggests that the implied volatility data that we have introduced in the joint MLEs contain additional information regarding the risk factors embedded in option prices, which makes option prices more expensive. The option premia increase even further if we consider a separate factor for variance risk in the CHJ model (Figure 15). Since the variance-depended kernel can generate a more flexible risk-adjustment this approach produces even more accurate option prices. Yet, this hypothesis still needs to be tested with actual option prices. Finally, and in order to compare the model with a benchmark option-pricing model, we demonstrate a calibration of the standard Black-Sholes model. It is easy to observe that the BS approach significantly undervalues OTM puts (and ITM calls), which directly results from the flat shape of the implied volatility function across maturities and its inability to capture asymmetric volatility effects and non-linearities in the pricing kernel.

6 Summary & Conclusions

There is a sizable stream in empirical finance literature that presents evidence of a significant divergence of risk-neutral densities from their physical counterparts, especially in the post-crash period of ’87. Consistently with this finding, numerous robust stylized facts have also been reported repeatedly. Option-implied volatilities have been found to exceed subsequent realized volatilities, providing biased predictions of future volatilities. The shape of the implied volatility smile has also become much steeper compared to that in the pre-crash period, and the fact that higher risk-neutral moments appear to differ from those of the physical densities has made pricing out-of-the-money and long-term options a quite non-trivial task. While attempting to
explain these empirical option-pricing anomalies, academics and practitioners have uncovered a significant variance risk factor that is embedded in option prices. This premium for bearing variance risk implies a change in investors’ preferences and has been primarily attributed to the stochastic nature of volatility as well as its pronounced asymmetric behavior. Option-pricing models that also try to incorporate variance risk are normally constructed within a stochastic volatility framework and require complex estimation procedures. Other GARCH-type approaches that are much easier to implement and can also be used for option-valuation purposes have been only assuming simple pricing kernel specifications, until very recently, that could not take into account variance risk.

Here, we utilize the asymmetric affine-GARCH process of Heston and Nandi (2000) that nests the symmetric GARCH model as a special case, and also allows for European options to be priced conveniently in closed-form. However, as we have already mentioned, the development of this approach has been based on the implicit assumption of a standard log-linear pricing kernel that only accounts for an equity risk premium, which has been proved to be quite restrictive and inadequate to explain empirical irregularities. In order to overcome this drawback, Christoffersen, Heston, and Jacobs (2013) have very recently introduced a variance-dependent pricing kernel that has several important implications and which may dramatically contribute towards resolving some of the commonly observed option-pricing anomalies. Combining the affine-GARCH framework with the richer properties of this enhanced pricing kernel that allows us to explicitly capture variance risk premium, significantly improves our ability to reconcile the underlying physical and risk-neutral dynamics of returns and variance, respectively.

Under this parametric framework and if there is indeed a negative price for variance risk, then the model is able to successfully reproduce several empirical option-pricing patterns. First of all, by capturing a higher level of risk-neutral variance compared to that of physical variance, it may explain the stylized property of option-implied variance exceeding future realized variance, which has made the short-selling of straddles a quite attractive option investment strategy in the post-crash period. Furthermore, the model generates a variance process that is also more persistent and more negatively correlated with returns under the risk-neutral measure, producing a higher degree of asymmetry and inducing stronger negative skewness and fatter tails for the risk-neutral density. This is crucial in resolving the mispricing of long-term and deep out of the money options. Generally, the superiority of this framework lies within the fact that in the presence of a non-zero variance risk premium the model can generate more flexible non-linear dynamics for the logarithmic pricing kernel, which can be more realistic and lead to more accurate valuation for European options.

In our empirical analysis we verify the validity of these implications by employing a somewhat different estimation approach. We build a joint likelihood function that incorporates both spot and forward-looking information in order to be able to capture the properties of the pricing kernel process as well. We implement the model using returns of the S&P500 index, but instead of simultaneously working with the wide cross-section of index options as in Christoffersen, Heston, and Jacobs (2013), we propose to use the daily variance dynamics inferred from the VIX index to approximate the implied risk-neutral variance process. Since the VIX is estimated on a daily basis from a proper weighting of index options, then minimizing the implied variance errors
in the objective function should be technically equivalent to minimizing a cross-section of vega-weighted option-valuation errors. Furthermore, this technique is much more straightforward and computationally more efficient and, as we show, it can produce similar results. We find strong evidence that support the hypothesis of priced volatility risk, since the model fits the data much better after allowing for this more flexible pricing kernel process. Our findings indicate a U-shaped pattern for the logarithmic kernel, which can be very helpful in explaining option-pricing puzzles after projecting the impact of volatility risk on the return process. This is also in line with the semi-parametric evidence documented in Christoffersen, Heston, and Jacobs (2013) regarding the implied shape of the kernel. We also demonstrate how these results can be used to calibrate more realistic option price characteristics across moneyness and maturity. Our final objective will be to evaluate the out-of-sample option-valuation performance of our estimation methodology.
References


Duan, J., & Simonato, J. (2001). American option pricing under garch by a markov chain


25


Appendix

Table 1: Statistical Properties of Sample

<table>
<thead>
<tr>
<th></th>
<th>mean (ann)</th>
<th>volat. (ann)</th>
<th>skewness</th>
<th>kurtosis</th>
<th>1st autocorr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500 (r)</td>
<td>0.0171</td>
<td>0.2088</td>
<td>-0.1753</td>
<td>10.70</td>
<td></td>
</tr>
<tr>
<td>VIX (v)</td>
<td>0.2328</td>
<td>0.0580</td>
<td>4.2461</td>
<td>28.99</td>
<td>0.9695</td>
</tr>
</tbody>
</table>

Figure 1: Daily log-returns on the S&P500 index and daily VIX risk-neutral variance.
Figure 2: Conditional variance path of a symmetric GARCH (left) and an asymmetric HN-GARCH model (right) both estimated by maximizing the likelihood of returns versus the daily VIX variance. The red line corresponds to the VIX and the black line to the model-based conditional variance estimates.

Figure 3: Conditional variance paths of a symmetric GARCH (left) and an asymmetric HN-GARCH (right) model obtained by the joined maximum likelihood approach versus the daily VIX variance and volatility.

Figure 4: Conditional variance and volatility from an asymmetric HN-GARCH model estimated with a joint maximum likelihood at the same time allowing for variance risk premium $\xi$ as in CHJ. The black line corresponds to the risk-neutral variance process and the grey line corresponds to the physical variance process.
Figure 5: Here we plot the News Impact Curves of all 3 models in order to show the impact of a return shock on variance. The x-axis represents the return innovations in standard deviations \((z)\) and the y-axis shows the percentage change in variance compared to the model-implied unconditional variance. The red line is for the symmetric Garch model, the blue line for the asymmetric HN model and the black line for the CHJ model. We use the parameters obtained by the joint MLEs.

Figure 6: This plot demonstrates the instantaneous impact of a return shock on both the physical and the risk-neutral variance in the CHJ approach. The blue-marked line is for the effect of a shock on physical variance and the red-marked line for the effect on risk-neutral variance. We can see that in the presence of variance risk \((\xi > 0)\), a shock \(z\) in returns does not only shift the distribution of the risk-neutral shock \(z^*\), but also changes its variance.
Figure 7: This figure shows the k-day ahead variance forecasts for the 3 different models, normalized by the unconditional variance level of each model. Graphs display how conditional variance slowly reverts back to its unconditional level. In the first graph the initial conditional variance level is 4 times smaller than its unconditional level \((h_t = 0.25 \times E_h)\) and in the second the conditional variance is 4 times larger than the unconditional variance level \((h_t = 4 \times E_h)\). We use the parameters of the joint MLEs for all models. We notice that the persistence implied by the CHJ approach is higher compared to that implied by the other two models.
Figure 8: Here we plot the variance impulse responses for different levels of return shocks \( z \). The dotted line is for the symmetric Garch, the dashed line for the HN and the solid line for the CHJ model. We use the parameters of the joint MLEs for all models. The black lines show the impact of a negative return shock on variance and the grey lines the impact of a positive shock. We see once more that for the symmetric Garch positive and negative shocks have the same positive impact on variance. For the two asymmetric models, we can see that a positive shock slightly decreases the level of variance, while a similar in magnitude negative shock tends to increase variance much more than a positive shock tends to decrease it. Again, we show the percentage of the change in variance compared to the level of the unconditional variance.
Figure 9: Here we plot the variance impulse responses for different levels of return shocks $z$. We consider 3 different levels of positive/negative return shocks as before and we demonstrate the impact of a shock on the physical and the risk-neutral variance of the CHJ model. The solid lines are for the physical variance and the dashed lines for the risk-neutral variance. The black color corresponds to the impact of a negative return shock and the grey color to the impact of a positive shock. Again, we plot the percentage change in variance compared to the physical/risk-neutral unconditional variance. We see that especially negative shocks affect physical variance by more than they affect risk-neutral variance.
Figure 10: This graph shows the term structure of volatility for a HN model. We use the parameters of the joint MLE and we plot the term structure of both the physical (solid line) and the risk-neutral variance (dashed line). The first graph starts from an initial variance 4 times smaller than the unconditional variance and the second starts from an initial variance level 4 times larger. Notice that the HN model suggests that the conditional physical variance is identical to the conditional risk-neutral variance at time $t + 1$, which places a strain when modeling volatility dynamics.

Figure 11: This graph shows the volatility term structure for the CHJ model for both the physical (solid line) and the risk-neutral variance (dashed line). Again, the first graph starts from an initial variance 4 times smaller than the unconditional variance and the second starts from an initial variance 4 times larger than the unconditional variance level. Since the CHJ approach allows for a variance risk premium, the two variance processes are allowed to differ by a scaling parameter and thus we can observe a wedge between the physical and the risk-neutral variance even at time $t + 1$, which is more realistic.
Figure 12: These plots show that the logarithm of the pricing kernel obtains a U-shaped pattern for values of $\xi > 0$, i.e. in the presence of a variance risk premium. We use the values obtained by the MLEs (in order to find $\phi, \eta$ and $\delta$) and we set the variance to its unconditional level. We plot the kernels for 3 different maturities: 1 day, 1 month and 3 months. The red-marked line is the U-shaped variance-dependent pricing kernel of the CHJ model. The blue and black lines correspond to the log-linear kernel implied by the HN and the Garch models, respectively. For the marked lines (left) we have used the parameters of the joint MLEs and for the simple dashed lines (right) those of the return-based MLEs.
Figure 13: This figure shows the theoretical closed-form option prices for the case of a symmetric GARCH model and an asymmetric HN model. Here we use the parameters obtained by the return-based MLE. We allow the spot price to range from 90 to 110\$, with a strike price of 100\$. The annual risk-free rate is set to 3\% and the time to maturity is 30 days.

Figure 14: In this figure we show how the theoretical option prices for the asymmetric HN model change when considering a joint MLE. Again, the spot price ranges from 90 to 110\$, with a strike price of 100\$. The annual risk-free rate is 3\% and the time to maturity is 30 days.
Figure 15: Here we compare the option prices of the HN model those of the CHJ approach that allows for a variance risk premium. Again, the spot price ranges from 90 to 110$, with a strike price of 100$. The annual risk-free rate is set to 3% and the time to maturity is 30 days.

Figure 16: HN and CHJ option prices (of the previous figure) versus those implied by a standard Black-Sholes model.
Table 2: Parameters and Statistics

### Physical Parameters

<table>
<thead>
<tr>
<th></th>
<th>GARCH</th>
<th>HN–GARCH</th>
<th>CHJ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\max \mathcal{L}^R$</td>
<td>$\max \mathcal{L}^V$</td>
<td>$\max \mathcal{L}$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>4.933</td>
<td>128.861</td>
<td>17.146</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$\omega$</td>
<td>3.59e-13</td>
<td>1.46e-15</td>
<td>1.49e-13</td>
</tr>
<tr>
<td></td>
<td>(0.8581)</td>
<td>(0.8488)</td>
<td>(0.6356)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>8.96e-06</td>
<td>7.86e-06</td>
<td>1.87e-05</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.934</td>
<td>0.852</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\tilde{\xi} = (1 - 2\alpha \xi)^{-1}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

### Risk-Neutral Parameters

<table>
<thead>
<tr>
<th></th>
<th>$\omega^*$</th>
<th>$\alpha^*$</th>
<th>$\beta^*$</th>
<th>$\gamma^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega^*$</td>
<td>3.59e-13</td>
<td>1.46e-15</td>
<td>1.49e-13</td>
<td>1.33e-12</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$\alpha^*$</td>
<td>8.96e-06</td>
<td>7.86e-06</td>
<td>1.87e-05</td>
<td>3.41e-06</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>0.934</td>
<td>0.852</td>
<td>0.952</td>
<td>0.760</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>$\gamma^*$</td>
<td>4.93</td>
<td>128.86</td>
<td>17.14</td>
<td>252.4</td>
</tr>
<tr>
<td>$\tilde{\xi} = (1 - 2\alpha \xi)^{-1}$</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
</tbody>
</table>

### Log–Likelihood

<table>
<thead>
<tr>
<th></th>
<th>$\log \mathcal{L}$</th>
<th>$\log \mathcal{L}^R$</th>
<th>$\log \mathcal{L}^V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log \mathcal{L}$</td>
<td>36269.22</td>
<td>31484.40</td>
<td>37543.13</td>
</tr>
<tr>
<td>$\log \mathcal{L}^R$</td>
<td>37006.06</td>
<td>38119.05</td>
<td>38365.20</td>
</tr>
<tr>
<td>$\log \mathcal{L}^V$</td>
<td>38546.29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Physical Properties

<table>
<thead>
<tr>
<th></th>
<th>Persistence, $\rho$</th>
<th>Long-run Vol., $\sqrt{E h}$</th>
<th>Volatility of $h_t$, $\sigma_h$</th>
<th>Corr($r_{t+1}, h^*_{t+1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistence, $\rho$</td>
<td>0.9340</td>
<td>0.8520</td>
<td>0.9520</td>
<td>0.9772</td>
</tr>
<tr>
<td>Long-run Vol., $\sqrt{E h}$</td>
<td>0.1850</td>
<td>0.1157</td>
<td>0.3133</td>
<td>0.1943</td>
</tr>
<tr>
<td>Volatility of $h_t$, $\sigma_h$</td>
<td>0.0507</td>
<td>0.0445</td>
<td>0.1058</td>
<td>0.0865</td>
</tr>
<tr>
<td>Corr($r_{t+1}, h^*_{t+1}$)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.9748</td>
</tr>
</tbody>
</table>

### Risk-Neutral Properties

<table>
<thead>
<tr>
<th></th>
<th>Persistence, $\rho^*$</th>
<th>Long-run Vol., $\sqrt{E h^*}$</th>
<th>Volatility of $h^<em><em>t$, $\sigma</em>{h^</em>}$</th>
<th>Corr($r_{t+1}, h^*_{t+1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistence, $\rho^*$</td>
<td>0.9342</td>
<td>0.9825</td>
<td>0.9575</td>
<td>0.9777</td>
</tr>
<tr>
<td>Long-run Vol., $\sqrt{E h^*}$</td>
<td>0.1853</td>
<td>0.3366</td>
<td>0.3330</td>
<td>0.1963</td>
</tr>
<tr>
<td>Volatility of $h^<em><em>t$, $\sigma</em>{h^</em>}$</td>
<td>0.0509</td>
<td>0.1775</td>
<td>0.1187</td>
<td>0.0874</td>
</tr>
<tr>
<td>Corr($r_{t+1}, h^*_{t+1}$)</td>
<td>-0.0812</td>
<td>-0.9681</td>
<td>-0.4532</td>
<td>-0.9753</td>
</tr>
</tbody>
</table>