Non-affine GARCH option pricing models, variance dependent kernels, and diffusion limits

Alexandru Badescu\textsuperscript{1}, Zhenyu Cui\textsuperscript{2}, and Juan-Pablo Ortega\textsuperscript{3}

Abstract

This paper investigates the pricing and weak convergence of an asymmetric non-affine, non-Gaussian GARCH model when the risk-neutralization is based on a variance dependent exponential linear pricing kernel with stochastic risk aversion parameters. The risk-neutral dynamics are obtained for a general setting and its weak limit is derived. We show how several GARCH diffusions, martingalized via well-known pricing kernels, are obtained as special cases and we derive necessary and sufficient conditions for the presence of financial bubbles. An extensive empirical analysis using both historical returns and options data illustrates the advantage of coupling this pricing kernel with non-Gaussian innovations.

Keywords: non-affine GARCH models, non-Gaussian innovations, exponential linear variance dependent pricing kernel, bivariate diffusion limit, option pricing.

\textit{JEL Classification:} C58, G12, G13.

\textsuperscript{1}Department of Mathematics and Statistics, University of Calgary, Calgary, Canada.

\textsuperscript{2}Financial Engineering Division, School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, USA.

\textsuperscript{3}Centre National de la Recherche Scientifique, Laboratoire de Mathématiques de Besançon, Université de Franche-Comté, Besançon, France.
1 Introduction

It is a well established fact that stochastic volatility (SV) option pricing models provide a significantly better alternative to the Black-Scholes (1973) model. Many empirical studies have indicated that several features, such as volatility clustering, conditional and unconditional non-normality, leverage effects, and non-monotonic stochastic discount factors (SDF) should be incorporated in the pricing framework.

The discrete-time SV option valuation literature has been generally dominated by the Generalized Autoregressive Conditionally Heteroskedastic (GARCH) type models introduced by Engle (1982) and Bollerslev (1986). Depending on the conditional mean and variance dynamics, the GARCH option pricing models can be divided into two special classes: affine and non-affine models. The family of affine GARCH processes was introduced by Heston and Nandi (2000), while the non-affine structure was proposed by Duan (1995). In both scenarios, the innovations are assumed to be conditionally Gaussian distributed and the risk-neutralization is constructed based on the so-called local risk-neutral valuation relationships (LRNVR). There are several advantages/disadvantages associated to each type of model used. On the one hand, the affine structure leads to semi-closed form expressions for the option prices, but the volatility dynamics may be viewed as too restrictive. On the other hand, the non-affine dynamics allows for flexible conditional mean and variance specification, but there are no closed-form solutions, option prices being typically computed using Monte-Carlo simulations. Extensive empirical comparisons between the two classes suggest that a simple non-affine variance dynamic with a leverage effect outperforms the affine Heston and Nandi (2000) pricing model when using options and/or VIX data (see Christoffersen et al. (2006) and Kanniainen et al. (2014) among others).

Since GARCH models based on Gaussian innovations cannot capture the skewness and kurtosis of financial data, several extensions regarding the underlying distribution have been proposed. The incompatibility between the LRNVR of Duan (1995) and the non-Gaussian GARCH setup has led to various choices for the pricing kernel which are usually justified by equilibrium and/or mathematical tractability arguments. Among the most popular SDF’s empirically used in a GARCH context are: Duan’s (1999) generalized LRNVR (see e.g., Stentoft (2008), Christoffersen et al. (2010), Simonato and Stentoft (2015)), the extended Girsanov principle (EGP) of Elliott and Madan (1998) (see Badescu and Kulperger (2008)), the conditional Esscher transform (see e.g. Siu et al. (2004), Gourieroux and Monfort (2007), Badescu and Kulperger (2008), Christoffersen et al. (2009), Chorro et al. (2012)).

The above pricing kernels have been implemented for various choices, parametric or nonparametric, of the GARCH innovations' distribution, and regardless of whether the model is affine or not. Although

---


\(^5\)Although GARCH processes are among the most popular choices for derivative pricing, there are several other important SV models in discrete-time, see for example Meddahi and Renault (2004), Darolles et al. (2006), Feunou and Tédognon (2012), Corsi et al. (2013), and Khrapov and Renault (2014) for studies on affine and realized volatility pricing models.

\(^6\)In addition to the form of the Gaussian innovations, several extensions regarding the conditional variance dynamics have been made. For example, Christoffersen et al. (2008), Majewski et al. (2015), and Bormetti et al. (2015) investigated the effect of multiple volatility components on option prices and on the term structure of variance risk premium. In this study we restrict our attention only to single component GARCH models.
the resulting option prices are theoretically different, a recent empirical study of Simonato and Stentoft (2015) showed that there are no significant differences in the pricing performance of a Johnson distributed GARCH model risk-neutralized via the generalized LRNVR and the conditional Esscher transform. One common feature of these pricing kernels is their dependence on a single (stochastic) parameter, typically quantified as the equity risk premium, which is uniquely determined by imposing that the discounted asset prices are martingales under the new measure. Therefore, the volatility risk is not directly priced in the above SDF’s.

Further insights on these pricing kernels can be gained by looking at the continuous time diffusion limits of the martingalized GARCH models. For example, in continuous time SV models the set of admissible pricing kernels from the Girsanov’s theorem is indexed by the variance risk preference parameter. Unlike in continuous-time, there is no such characterization for the aforementioned pricing kernels in discrete-time. However, one can still identify the form of this variance risk preference parameter implied by the risk-neutral GARCH weak limit. Duan (1996) showed that the LRNVR is consistent with the minimal martingale measure in continuous time which is obtained by assuming a zero market price of variance risk. Badescu et al. (2015b) proved a similar result for the conditional Esscher transform, regardless of the assumption on the GARCH innovations’ distribution. Interestingly, when the innovations are skewed, both the generalized LRNVR and EGP imply a non-zero market price of risk in continuous time which depends on the skewness and kurtosis of the driving noise.

Since the early 2000, many studies provided evidence against the hypothesis of a monotonic pricing kernel, typically referred as the pricing kernel puzzle, and proposed various alternatives (see e.g. Aït-Sahalia and Lo. (2000), Chabi-Yo et al. (2008), Bakshi et al. (2010), Chabi-Yo (2012), Song and Xiu (2016), among others). In GARCH settings, the pricing kernel puzzle has been first addressed by Christoffersen et al. (2013) who proposed an exponential affine variance dependent pricing kernel which takes into account both sources of risk, market and volatility risk. This generalizes the conditional Esscher transform, since the latter can be obtained as a special case when the variance risk preference parameter is set to zero. The variance dependent kernel has been implemented in the affine Heston and Nandi (2000) Gaussian framework and a sequential calibration exercise showed that it outperforms the conditional Esscher transform. A similar pricing kernel has been used in Majewski et al. (2015) and Bormetti et al. (2015) for pricing under a multi-component affine GARCH model, while Khrapov and Renault (2014) use a variance dependent kernel for a compound autoregressive (CAR) model. However, the only paper which investigates the option pricing within a non-Gaussian affine GARCH framework, at least to our knowledge, is provided by Babaoglu et al. (2014), who combine a U-shaped pricing kernel with a two-volatility component Inverse Gaussian model. Despite the popularity of the variance depen-

7Continuous time weak limits of various GARCH models have been extensively studied under the physical measure. As it is customary in this topic (see Nelson (1990)), assumptions need to be made as to the asymptotic behaviour of the dependence of the model parameters on the sampling frequency. One of the main issues is the non-uniqueness of such constraints, which may lead to different limit processes (see e.g. Corradi (2000)). In this paper we only follow the approach of Nelson, since this leads to a meaningful bivariate SV model which serves well our purpose of connecting the change of measure in discrete time with the Girsanov’s theorem in continuous time.
dent kernel among the affine models, there are no empirical studies on its application to option pricing for non-affine non-Gaussian GARCH models. Our aim is to address this issue from both a theoretical and an empirical point of view, and to investigate the continuous time counterpart. The contributions in this paper are articulated around three main points that we describe in the following paragraphs.

First, we risk-neutralize the price dynamics induced by asymmetric, non-affine GARCH processes driven by arbitrary non-Gaussian innovations, using a new exponential linear variance dependent pricing kernel with stochastic equity and variance risk preference parameters. This choice represents an extension of the pricing kernel formulated in Christoffersen et al. (2013) which is constructed based on constant prices of risk and on the assumption that the bivariate cumulant generating function of asset returns and variance is linear in variance, which is only meaningful in an affine GARCH setup. In fact, we argue that having constant market prices of risk at the same time is not consistent with the non-affine GARCH setting. The flexibility introduced in our approach allows us to consider richer price dynamics that, as we see in the empirical section, are capable of improving in and out-of-sample pricing performances. These results are contained mainly in Section 3, where we work out in detail the expressions needed to implement our approach for asymmetric GARCH models with innovations conditionally distributed according to the following three types of prescriptions: Gaussian, mixture of Gaussians, and mixture of exponentials.

Second, following similar parametric constraints as in Nelson (1990), we derive the weak diffusion limit of the risk-neutralized GARCH process studied in the previous section. Using a particular GARCH conditional mean specification, the resulting continuous-time stochastic volatility model nests as particular cases the affine correlated Hull-White models obtained in Badescu et al. (2015b) and Badescu et al. (2015a) and obtained by risk-neutralizing the GARCH process using the conditional Esscher transform, the Extended Girsanov Principle and the generalized LRNVR, as well as the generalized geometric mean-reverting diffusion processes with affine drift introduced in Metcalf and Hassett (1995) or in Ewald and Yang (2007). Thus, we can view the exponential affine variance dependent pricing kernel as the discrete time alternative of the Girsanov’s change of measure for continuous SV models. Next, we determine necessary and sufficient conditions for the affine correlated Hull-White diffusion limit to be a true martingale or a strict local martingale, which facilitates the classification of financial bubbles in the sense of Protter (2013).

Finally, we propose a novel method based on an exponential measure change and derive explicit exact probability density functions for the diffusion limit of the GARCH model. This is of particular interest since the probability density function can not be determined explicitly for most models and various approximation methods have been used in the literature (see Aıt Sahalia (2002; 2008).
and Choi (2015)).

Third, the practical relevance of the theoretical results that we just described is assessed by conducting an extensive empirical analysis. The exercise is based on a dataset containing more than forty thousand S&P500 option prices recorded during the period spanning January 1st, 2004 - December 31st, 2013. This study aims to prove two specific points; first, the importance of using non-trivial prices of variance risk in the stochastic discount factor and hence of going beyond the conditional Esscher transform. Second, the pertinence of non-Gaussian innovations in the GARCH model in order to better reproduce empirically observed stylized facts of the underlying returns time series and cross section of options. Using joint and sequential likelihood estimation procedures based on historical returns and option prices, our results show a significant competitive advantage on both counts. Our findings suggest that combining a Gaussian mixture distribution with a variance dependent kernel provides the best pricing framework when compared to other alternatives involving Gaussian innovations and conditional Esscher transforms.

The rest of the paper is organized as follows. In Section 2 we introduce the underlying discretized non-affine non-Gaussian GARCH setting model and illustrate its continuous time limit. The pricing kernel and risk-neutral derivations are presented in Section 3. In Section 4 we derive the weak limit of the risk-neutral GARCH model and we analyze two special diffusion classes as described above. The numerical results are contained in Section 5. Section 6 concludes the paper.

2 Notations and preliminaries

Consider a \( n \)-indexed discrete time financial market with the set trading dates \( \mathcal{T}_n = \{ l \mid l = k\tau, k = 0, 1, \ldots, nT \} \), where \( \tau := 1/n \) is the sampling period. Real-world dynamics are defined on a filtered probability space \( (\Omega_n, \mathcal{F}_n, \{ \mathcal{F}_l,n \}_l \in \mathcal{T}_n, P_n) \). We assume that the logarithm of the risky asset price process, denoted by \( \{ Y_l,n \}_l \in \mathcal{T}_n = \{ \log S_l,n \}_l \in \mathcal{T}_n \), has the following stochastic volatility structure for any \( k = 0, 1, \ldots, nT \):

\[
\Delta Y_{k\tau,n} := Y_{k\tau,n} - Y_{(k-1)\tau,n} = \mu_{k\tau,n} \tau + \sqrt{\tau} \sigma_{k\tau,n} \epsilon_{k\tau,n}, \quad \epsilon_{k\tau,n}|\mathcal{F}_{(k-1)\tau,n} \sim \mathcal{D}(0, 1), \quad (2.1)
\]

\[
\Delta \sigma_{(k+1)\tau,n}^2 := \sigma_{(k+1)\tau,n}^2 - \sigma_{k\tau,n}^2 = \alpha_0(\tau) + \alpha_1(\tau) \sigma_{k\tau,n}^2 (\epsilon_{k\tau,n} - \gamma)^2 + (\beta_1(\tau) - 1) \sigma_{k\tau,n}^2. \quad (2.2)
\]

We assume that \( \mathcal{F}_{k\tau,n} \) is the \( \sigma \)-field generated by the historical asset prices, \( \mathcal{F}_{k\tau,n} = \sigma(Y_0,n, Y_1,n, \ldots, Y_{k\tau,n}) \) for any \( k = 0, \ldots, nT \); \( \epsilon_n = \{ \epsilon_{k\tau,n} \}_{k=0,\ldots,nT} \) is a sequence of \( \mathcal{F}_{(k-1)\tau,n} \)-conditional i.i.d. random variables with zero mean and unit variance distribution \( \mathcal{D} \), \( \epsilon_{k\tau,n}|\mathcal{F}_{(k-1)\tau,n} \sim \mathcal{D}(0, 1) \). We denote by \( f_\epsilon(\cdot) \) and \( F_\epsilon(\cdot) \) the corresponding conditional p.d.f. and c.d.f., and we let \( \kappa(\cdot) \) be the conditional cumulant generating function of the driving noise under the physical measure \( P_n \) so that:

\[
\kappa(\epsilon) := \log E \left[ \exp \left( \epsilon \epsilon_{k\tau,n} \right) \mid \mathcal{F}_{(k-1)\tau,n} \right].
\]
We denote the innovations’ \( j \)th raw moments by:

\[
M_j = E \left[ e_{k\tau,n}^j | \mathcal{F}_{(k-1)\tau,n} \right].
\]

These are assumed to be finite quantities which do not depend on the sampling period \( \tau \). We do not impose yet any specific form for the conditional mean return in (2.1). The conditional variance process \( \{ \sigma^2_{k\tau,n} \}_{0 \leq k \leq nT} \) has a GARCH-type structure with sampling frequency varying parameters \( \alpha_0(\tau), \alpha_1(\tau), \beta_1(\tau) \), and \( \gamma \) (constant not dependent on \( \tau \)) satisfying standard constraints that insure non-negativity of the variance process and covariance stationarity. Note that in the special case \( \tau = 1 \), the process (2.1)-(2.2) reduces to a general asymmetric NGARCH(1,1) model.

We conclude this section by presenting the setting and the asymptotic parametric conditions required for the weak convergence of the process (2.1)-(2.2) in the physical world as the sampling frequency tends to infinity. This result will be extended in Section 4, where we will obtain the diffusion limit of the risk-neutral dynamics resulting from coupling the above model for the underlying with the kernel introduced in Section 3.

First, we introduce the right continuous with left limit (c\'adl\'ag) extension of the discretized GARCH model defined by:

\[
\{ Y_{t,n}, \sigma^2_{t,n} \}_{k\tau \leq t < (k+1)\tau} := \left\{ Y_{k\tau,n}, \sigma^2_{(k+1)\tau,n} \right\}, \quad k = 0, \ldots, nT. \tag{2.3}
\]

Similarly, we define the continuous filtration, \( \{ \mathcal{F}_{t,n} \}_{k\tau \leq t < (k+1)\tau} := \mathcal{F}_{k\tau,n}, k = 0, \ldots, nT \), and we denote by \( \mathcal{F}_{t,n}^\sigma := \mathcal{F}_{t,n} \cup \{ \sigma^2_{t,n} = \sigma^2_{t} \} \). Following the same asymptotic parametric conditions as in Nelson (1990), the following continuous time weak limit of the above model has been obtained in Badescu et al. (2015b).

**Proposition 2.1** Assume that the following asymptotic parametric conditions hold:

\[
\lim_{\tau \to 0} \frac{\alpha_0(\tau)}{\tau} = \omega_0, \quad \lim_{\tau \to 0} \frac{\alpha_1(\tau)(1 + \gamma^2) + \beta_1(\tau) - 1}{\tau} = -\omega_1, \quad \lim_{\tau \to 0} \frac{\sigma^2_t(\tau)}{\tau} = \omega_2. \tag{2.4}
\]

Then, as \( \tau \) approaches zero, the process \( \{ Y_{t,n}, \sigma^2_{t,n} \} \) defined in (2.1)-(2.2) converges weakly to a bivariate diffusion \( (Y_t, \sigma^2_t) \) which satisfies the following stochastic differential equations:

\[
\begin{align*}
\text{d}Y_t &= \mu_t dt + \sigma_t dB_{1t}, \quad \text{(2.5)} \\
\text{d}\sigma^2_t &= \left( \omega_0 - \omega_1 \sigma^2_t \right) dt + \sqrt{\omega_2} (M_4 - 2\gamma) \sigma^2_t dB_{1t} + \sqrt{\omega_2} \sqrt{M_4 - M_3^2 - 1} \sigma^2_t dB_{2t}. \quad \text{(2.6)}
\end{align*}
\]

Here, \( B_{1t} \) and \( B_{2t} \) are two independent Brownian motions on \((\Omega, \mathcal{F}, \{ \mathcal{F}_t \}_{t \in [0, \ldots, T]}, P)\) and \( \mu_t \) represents the continuous time version of \( \mu_{k\tau,n} \).

The above result is a consequence of the weak convergence theorem for Markov processes which typically
holds under standard moment constraints for the GARCH model (see Nelson (1990) or Francq and Zakoian (2010) for further details). In the case of Gaussian innovations (i.e. $\kappa_\tau(\sigma_t) = \frac{1}{2} \sigma_t^2$, $M_3 = 0$, and $M_4 = 3$), this result coincides with the standard asymmetric GARCH diffusion limit of Duan (1997).

The bivariate diffusion in (2.5)-(2.6) can be viewed as a generalized Hull-White model with an affine drift for the variance process and driven by two correlated Brownian motions. This can be easily seen if we write the asset price dynamics as:

\[
\begin{align*}
    dS_t &= \left( \mu_t + \frac{1}{2} \sigma_t^2 \right) dt + \sigma_t dB_{1t}, \\
    d\sigma_t^2 &= (\omega_0 - \omega_1 \sigma_t^2) dt + \zeta \sigma_t^2 dB_{3t}.
\end{align*}
\]  

(2.7) \hspace{1cm} (2.8)

Here $B_{3t}$ is another standard Brownian motion which has a constant correlation $\rho$ with $B_{1t}$ given by:

\[
\rho := \frac{\sqrt{\omega_2} (M_4 - 2\gamma)}{\zeta} = \frac{\sqrt{\omega_2} (M_3 - 2\gamma)}{\sqrt{\omega_2 (M_3 - 2\gamma)^2 + \omega_2 (M_4 - M_3^2 - 1)}}.
\]  

(2.9)

When dealing with stochastic differential equations for diffusions, one needs to verify whether the underlying coefficients are smooth enough so that there exists a unique weak solution. Here we recall the Engelbert-Schmidt conditions (see Theorem 5.15, p341, Karatzas and Shreve (1988)) for a stochastic differential equation with drift and diffusion functions $a(\cdot)$ and $b(\cdot)$, respectively:

\[
\forall x \in J, \quad b(x) \neq 0, \quad \frac{1}{b^2(\cdot)}, \quad \frac{a(\cdot)}{b^2(\cdot)} \in L^1_{loc}(J),
\]  

(2.10)

where $L^1_{loc}(J)$ denotes the class of locally integrable functions (i.e. the functions mapping $J$ to $\mathbb{R}$ that are integrable on compact subsets of $J$). It is straightforward to check that (2.7)-(2.8) satisfy these conditions for the state space $J = (0, \infty)$.

### 3 Exponential linear variance dependent pricing kernels

In this section, we derive the risk-neutral dynamics of the asset price under a family of exponential linear pricing kernels and we discuss the relationships between this pricing kernel and other potential candidates. We start by introducing, for any $k = 0, 1, \ldots, nT$, the following stochastic discount factor (SDF):

\[
N_{k, T} = \exp \left( -\theta_{k, n}^{(1) Y_{k, T}} - \theta_{k, n}^{(2) \Delta \sigma_{(k+1)T}} - \kappa(\Delta Y, \Delta \sigma^2) \left( -\theta_{k, n}^{(1)}, -\theta_{k, n}^{(2)} \right) \right).
\]  

(3.1)

We assume that $\theta_{n}^{(1) Y_{k, T}} = \theta_{k, n}^{(1)}$ and $\theta_{n}^{(2) \Delta \sigma_{(k+1)T}} = \theta_{k, n}^{(2)}$ are $\mathcal{F}_n$-predictable processes quantifying the prices of equity and variance risk. Here, $\kappa(\Delta Y, \Delta \sigma^2)(\cdot; \cdot)$ represents the joint conditional cumulant generating function of changes in log-prices and conditional variances. Since $\Delta Y_t$ has a linear
dependence on \( \epsilon_{k,\tau,n} \), and \( \Delta \sigma^2_{k,\tau,n} \) has a quadratic dependence on \( \epsilon_{k,\tau,n} \), we have

\[
\kappa(\Delta Y, \Delta \sigma^2)(z_1, z_2) = \log E \left[ \exp \left( z_1 \Delta Y_{k,\tau,n} + z_2 \Delta \sigma^2_{(k+1)\tau,n} \right) \mathcal{F}_{(k-1)\tau,n} \right]
= z_1 \mu_{k,\tau,n} + z_2 \left( \alpha_0(\tau) + \sigma^2_{k,\tau,n} (\alpha_1(\tau) \gamma^2 + \beta_1(\tau) - 1) \right)
+ \kappa(\epsilon, \epsilon) \left( z_1 \sqrt{\tau} \sigma_{k,\tau,n} - 2z_2 \gamma \alpha_1(\tau) \sigma^2_{k,\tau,n}, z_2 \alpha_1(\tau) \sigma^2_{k,\tau,n} \right). \tag{3.2}
\]

Substituting (3.2) into (3.1) leads to the following representation of the SDF (3.1):

\[
N_{k,\tau,n} = \exp \left( A_{k,\tau,n} \epsilon_{k,\tau,n} + B_{k,\tau,n} \epsilon^2_{k,\tau,n} - \kappa(\epsilon, \epsilon) (A_{k,\tau,n}, B_{k,\tau,n}) \right), \tag{3.3}
\]

where the coefficients \( A_n = \{A_{k,\tau,n}\}_{k=0,...,nT} \) and \( B_n = \{B_{k,\tau,n}\}_{k=0,...,nT} \) are \( \mathcal{F}_n \)-predictable processes given by:

\[
A_{k,\tau,n} = -\sqrt{\tau} \theta^{(1)}_{k,\tau,n} \sigma_{k,\tau,n} + 2\gamma \alpha_1(\tau) \theta^{(2)}_{k,\tau,n} \sigma^2_{k,\tau,n}, \tag{3.4}
\]
\[
B_{k,\tau,n} = -\alpha_1(\tau) \theta^{(2)}_{k,\tau,n} \sigma^2_{k,\tau,n}. \tag{3.5}
\]

We notice that the exponential linear pricing kernel reduces to an exponential quadratic pricing kernel in \( \epsilon_{k,\tau,n} \) (or a second-order conditional Esscher type pricing kernel with respect to \( \epsilon_{k,\tau,n} \), see Monfort and Pegoraro (2012) for details on this SDF) when applied to the asymmetric GARCH structure in (2.1)-(2.2). The specific form of the coefficients \( A_{k,\tau,n} \) and \( B_{k,\tau,n} \) and, in particular, their dependence on the sampling period \( \tau \) and the market prices of risk, will be crucial later on at the time of deriving the continuous time diffusion limit.\(^9\) Moreover, as in Christoffersen et al. (2013), we can show that the logarithm of the pricing kernel is a quadratic function of the asset returns. Indeed, regardless of the GARCH innovation distribution used, the pricing kernel is U-shaped when the variance premium is negative (i.e. \( \theta^{(2)}_{k,\tau,n} < 0 \)). The main result of this section is contained in the following proposition.

**Proposition 3.1** Suppose that the predictable processes \( \theta^{(1)}_n \) and \( \theta^{(2)}_n \) satisfy the following relation for any \( k = 0, \ldots, nT \):

\[
\mu_{k,\tau,n} = \tau - \frac{1}{\tau} \left( \kappa(\epsilon, \epsilon) (A_{k,\tau,n} + \sqrt{\tau} \sigma_{k,\tau,n}, B_{k,\tau,n}) - \kappa(\epsilon, \epsilon) (A_{k,\tau,n}, B_{k,\tau,n}) \right). \tag{3.6}
\]

Then, the following statements hold:

(i) The process \( Z_n := \{Z_{k,\tau,n}\}_{k=0,...,nT} \) defined by \( Z_{k,\tau,n} := \prod_{l=1}^{k} N_{l,\tau,n} \) with \( Z_{0,n} = 1 \) is a \( P \)-martingale and \( Z_{T,n} \) defines an equivalent measure \( Q_n \) via \( \frac{dQ_n}{dP} := \frac{dZ_{T,n}}{dZ_{0,n}} \).

---

\(^9\)If we would start directly with a second-order Esscher transform as our pricing kernel, we would have to make a priori assumptions on the dependence of both Esscher coefficients on the time scale \( \tau \). Depending on such assumptions, we would obtain different GARCH diffusion limits. For example, if the coefficient of \( \epsilon_{k,\tau,n} \) in (3.3) is of order \( \sqrt{\tau} \) and that of \( \epsilon_{k,\tau,n}^2 \) is of order \( \tau \), we would obtain the same GARCH limit as in the standard Esscher case, so the market price of variance risk \( \theta^{(2)}_{k,\tau,n} \) has no impact on the risk-neutral diffusion case. This is no longer the case if we work with the SDF in (3.1) for which the continuous limit is uniquely determined.
(ii) The innovation process \( \epsilon^*_n := \{ \epsilon^*_{kT,n} \}_{k=0,\ldots,nT} \) defined by:

\[
\epsilon^*_{kT,n} = \frac{\epsilon_{kT,n} - P_{kT,n}}{\sqrt{R_{kT,n}}},
\]

is a sequence of \( F(k-1)_\tau \)-conditionally uncorrelated zero mean and unit variance random variables with a different conditional distribution under \( Q_n \) denoted by, \( \epsilon^*_{kT,n} \sim D^*(0,1) \). Here \( P_{kT,n} \) and \( R_{kT,n} \) are the first and second partial derivatives of \( \kappa_{(\epsilon,\epsilon^2)}(\cdot,\cdot) \) with respect to the first argument under \( P_n \) and evaluated at \( A_{kT,n} \) and \( B_{kT,n} \):

\[
P_{kT,n} = \frac{\partial \kappa_{(\epsilon,\epsilon^2)}(A_{kT,n},B_{kT,n})}{\partial z_1} \quad \text{and} \quad R_{kT,n} = \frac{\partial^2 \kappa_{(\epsilon,\epsilon^2)}(A_{kT,n},B_{kT,n})}{\partial z_1^2}.
\]

(iii) The risk-neutralized asset return dynamics are given by:

\[
\begin{align*}
\Delta Y_{kT,n} &= \kappa_{(\epsilon,\epsilon^2)}(A_{kT,n},B_{kT,n}) - \kappa_{(\epsilon,\epsilon^2)}(A_{kT,n} + \sqrt{\tau} \sigma_{kT,n},B_{kT,n}) \\
&\quad + \left( r + \frac{P_{kT,n} \sigma_{kT,n}}{\sqrt{\tau}} \right) \tau + \sqrt{\tau} R_{kT,n} \sigma_{kT,n} \epsilon^*_{kT,n} + \epsilon^*_{kT,n} \sim D^*(0,1),
\end{align*}
\]

\[
\begin{align*}
\Delta \sigma^2_{(k+1)\tau,n} &= \alpha_0(\tau) + \alpha_1(\tau) \sigma^2_{kT,n} \left( \sqrt{R_{kT,n}} \epsilon^*_{kT,n} + P_{kT,n} - \gamma \right)^2 + (\beta_1(\tau) - 1) \sigma^2_{kT,n}.
\end{align*}
\]

The relation (3.6) ensures that the discounted asset prices are martingales under the new probability measure \( Q_n \). This is the reason why we call it the martingale condition/constraint. Note that this identity provides a relation between the two market prices of risk, but it does not uniquely determine them. The risk-neutral innovation distribution, denoted by \( D^*(0,1) \), is a priori not known and has to be determined based on the assumptions on the underlying GARCH innovation distribution \( D(0,1) \). A useful relation in this sense is the connection between \( \kappa^*_n(\cdot) \) and \( \kappa_{(\epsilon,\epsilon^2)}(\cdot,\cdot) \). Indeed, using (3.3) and (3.7), we can write:

\[
\kappa^*_n(z) = -z \frac{P_{kT,n}}{\sqrt{R_{kT,n}}} + \kappa_{(\epsilon,\epsilon^2)} \left( \frac{z}{\sqrt{R_{kT,n}}} + A_{kT,n},B_{kT,n} \right) - \kappa_{(\epsilon,\epsilon^2)}(A_{kT,n},B_{kT,n}).
\]

We notice that the conditional variance of the asset returns per time interval of length \( \tau \) under \( Q_n \), denoted by \( \sigma^2_{kT,n} \), is proportional to the physical one and the constant of proportionality is given by \( R_{kT,n} \):

\[
\sigma^2_{kT,n} = R_{kT,n} \sigma^2_{kT,n}.
\]

The risk-neutral conditional variance obviously exceeds the physical one whenever \( R_{kT,n} > 1 \). Depending on the assumption made regarding the GARCH driving noise distribution, this generally leads to a specific constraint for the market prices of risk \( \theta^{(1)}_{kT,n} \) and \( \theta^{(2)}_{kT,n} \).

If we let \( \theta^{(2)}_{kT,n} = 0 \), the exponential linear SDF reduces to the well-known conditional Esscher transform. Moreover, the risk-neutral dynamics of the asset price under the conditional Esscher transform can be
obtained as a special case of Proposition 3.1. Indeed, if we take \( \theta_{kT,n}^{(2)} = 0 \) in equation (3.1) for all \( k = 0, \ldots, nT \), we have, \( A_{kT,n} = -\sqrt{T} \theta_{kT,n}^{(1)} \sigma_{kT,n} \) and \( B_{kT,n} = 0 \), so the risk-neutral dynamics reduce to those found in Badescu et al. (2015b).

An important issue when computing derivative prices based on the exponential linear SDF (3.1) is the numerical tractability of this equation, since closed-form expressions for the bivariate cumulant generating function of \((\varepsilon, \varepsilon^2)\) are only available for a few special cases. In general, \( \kappa(\varepsilon, \varepsilon^2) \) is evaluated directly using the p.d.f. of \( \varepsilon \) but it can also be computed using a randomization approach which uses the moment generating function of \( \varepsilon \) as we now explain. First, we let:

\[
\kappa(\varepsilon, \varepsilon^2)(z_1, z_2) := \log \mathbb{E} \left[ \exp(z_1 \varepsilon_{kT,n} + z_2 \varepsilon_{kT,n}^2) \mid \mathcal{F}_{(k-1)T,n} \right].
\]

Using a conditional version of the randomization idea of Keller-Ressel and Muhle-Karbe (2012), we have:

\[
\kappa(\varepsilon, \varepsilon^2)(z_1, z_2) = \log \mathbb{E} \left[ \exp \left( z_2 \left( \varepsilon_{kT,n} + \frac{z_1}{2z_2} \right)^2 - \frac{z_1^2}{4z_2} \right) \mid \mathcal{F}_{(k-1)T,n} \right]
\]

\[
= \log \mathbb{E} \left[ \exp \left( \sqrt{2z_2} \left( \varepsilon_{kT,n} + \frac{z_1}{2z_2} \right) U \right) \mid \mathcal{F}_{(k-1)T,n} \right] - \frac{z_1^2}{4z_2},
\]

where \( U \) is a standard normal random variable conditional on \( \mathcal{F}_{(k-1)T,n} \) and it is independent of \( \varepsilon_{kT,n} \). The second expectation is taken with respect to the product law of \( \varepsilon_{kT,n} \) and \( U \). Recalling that the conditional moment generating function of \( \varepsilon_{kT,n} \) is \( M_{\varepsilon}(z) := \mathbb{E} \left[ \exp(z \varepsilon_{kT,n}) \mid \mathcal{F}_{(k-1)T,n} \right] \), we write the last expectation in (3.13) as:

\[
\mathbb{E} \left[ \exp \left( \sqrt{2z_2} \left( \varepsilon_{kT,n} + \frac{z_1}{2z_2} \right) U \right) \mid \mathcal{F}_{(k-1)T,n} \right] = \int_{-\infty}^{\infty} e^{\frac{z_1}{2z_2} u} M_{\varepsilon}(\sqrt{2z_2} u) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.
\]

Thus, we can write the joint cumulant generating function as a one dimensional integral of the conditional moment generating function of \( \varepsilon_{kT,n} \mid \mathcal{F}_{(k-1)T,n} \):

\[
\kappa(\varepsilon, \varepsilon^2)(z_1, z_2) = \log \int_{-\infty}^{\infty} e^{\frac{z_1}{2z_2} u} M_{\varepsilon}(\sqrt{2z_2} u) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du - \frac{z_1^2}{4z_2}.
\]

We notice that (3.14) can be easily evaluated when \( M_{\varepsilon}(u) \) is a quadratic function in \( u \), which is clearly the case for conditional Gaussian innovations. However, it is important to identify other non-Gaussian cases for which closed-form expressions can also be obtained. If these are not available, one has to rely on Monte-Carlo simulations or numerical integration.
3.1 Conditional Gaussian economies

Conditional Gaussian innovations have been the main distributional assumptions in the context of GARCH option pricing models based on a variance dependent pricing kernel. As we already mentioned, when $\epsilon_{k,n} \sim \mathcal{N}(0,1)$, we can integrate and simplify (3.14), which yields the following closed-form expression for the bivariate cumulant generating function of $(\epsilon, \epsilon^2)$:

$$\kappa_{(\epsilon, \epsilon^2)}(z_1, z_2) = \frac{1}{2} \left( \frac{z_1^2}{1 - 2z_2} - \log(1 - 2z_2) \right).$$

(3.15)

Using this result we characterize the dynamical evolution of the asset returns in the following corollary.

**Corollary 3.2** If $\epsilon_{k,n} | F_{(k-1)n} \sim \mathcal{N}(0,1)$, then the risk-neutral dynamics of the asset return under $Q_n$ is given by:

$$\Delta Y_{k,n} = \left( r - \frac{\sigma_{k,n}^2}{2 (1 - 2B_{k,n})} \right) \tau + \sqrt{\tau} \sigma_{k,n} \sqrt{\frac{1}{1 - 2B_{k,n}}} \epsilon_{k,n}^*, \quad \epsilon_{k,n}^* \sim \mathcal{N}(0,1),$$

(3.16)

$$\Delta \sigma_{(k+1)n}^2 = \alpha_0(\tau) + \alpha_1(\tau) \sigma_{k,n}^2 \left( \sqrt{\frac{1}{1 - 2B_{k,n}}} \epsilon_{k,n}^* \right. + \left. \frac{A_{k,n}}{1 - 2B_{k,n}} - \gamma \right)^2 + (\beta_1(\tau) - 1) \sigma_{k,n}^2,$$

(3.17)

where $A_{k,n}$ and $B_{k,n}$ are provided in (3.4)-(3.5).

We notice that the exponential linear pricing kernel preserves the distribution of the asset returns after the measure change since the innovations are conditionally Gaussian distributed under $Q_n$. Additionally, by examining (3.16), it is easy to check that the discounted asset prices are martingales. The conditional risk-neutral variance $\sigma_{k,n}^2$ per time interval of length $\tau$ no longer has a GARCH structure since both $A_{k,n}$ and $B_{k,n}$ depend on $\sigma_{k,n}$. In this case $\sigma_{k,n}^2$ is related to the conditional variance under the physical measure in the following way:

$$\sigma_{k,n}^* = \frac{\sigma_{k,n}^2}{1 + 2\alpha_1(\tau) \theta_{k,n}^{(2)} \sigma_{k,n}^2},$$

(3.18)

Since the GARCH coefficient $\alpha_1(\tau) \geq 0$, the risk-neutral conditional variance exceeds the physical counterpart if and only if the variance premium is negative, $\theta_{k,n}^{(2)} < 0$. Notice that such a result does not necessarily hold for other innovation distributions.

In this case the martingale equation (3.6) can be re-written as:

$$\mu_{k,n} = r - \frac{\sigma_{k,n}^2}{2 (1 - 2B_{k,n})} - \frac{\sigma_{k,n}^2}{\sqrt{\tau}} \frac{A_{k,n}}{1 - 2B_{k,n}}.$$

(3.19)
Substituting the expressions for the market prices of risk into (3.19), we obtain an explicit dependence between $\theta_{n}^{(1)}$ and $\theta_{n}^{(2)}$:

$$
\theta_{k,T,n}^{(1)} = \frac{1}{2} + \frac{\mu_{k,T,n} - r}{\sigma_{k,T,n}^{2}} - 2^{\alpha_{1}(\tau)} \sigma_{k,T,n} \left( \sqrt{r - \mu_{k,T,n}} \sigma_{k,T,n} - \gamma \right) \theta_{k,T,n}^{(2)}. \quad (3.20)
$$

The well-known conditional Esscher transform under Gaussian innovations is obtained by setting $\theta_{k,T,n}^{(2)} = 0$, for all $k = 0, \ldots, nT$. In that case we recover the option pricing model in Duan (1995) derived via his local risk-neutral valuation relationship (LRNVR). In that case the physical and risk-neutral conditional variances are equal, and the Esscher parameter for the market price of equity risk is given by:

$$
\theta_{k,T,n}^{(1)} = \frac{1}{2} + \frac{\mu_{k,T,n} - r}{\sigma_{k,T,n}^{2}}. \quad (3.21)
$$

### 3.2 Relation with the variance dependent kernel in Christoffersen et al (2013)

The risk-neutral dynamics presented above are in agreement with those obtained in Christoffersen et al. (2013) using a variance dependent pricing kernel for the Heston and Nandi (2000) model. However, there are some major differences regarding both the underlying model and the pricing kernel used, and we now discuss them below.

First, using the same parametric notation, we recall that the variance dependent SDF of Christoffersen et al. (2013) adapted to the sampling frequency implied by $\tau$ is:

$$
N_{k,T,n} = \exp \left( (r - \delta) \tau + \eta \sigma_{k,T,n}^{2} + \phi \Delta Y_{k,T,n} + \xi \Delta \sigma_{(k+1)\tau,n}^{2} \right). \quad (3.22)
$$

This expression is the discrete-time version of the continuous time SDF used in the Heston (1993) option pricing model. Unlike our pricing kernel, their construction is based on four (constant) parameters: $\delta$ and $\eta$ are the time preference parameters, usually identified from the bond martingale constraint, while $\phi$ and $\xi$ govern the aversion to equity and variance risks, respectively, and satisfy the standard martingale constraint for the discounted asset price. When applied to affine GARCH type models driven by Gaussian or Inverse Gaussian innovations (see Babaoglu et al. (2014) for the latter case), this change of measure leads to semi closed form solutions for European option prices. Despite this important advantage, the specification in (3.22) is somewhat restrictive since it requires that the bivariate cumulant generating function of $\Delta Y_{k,T,n}$ and $\Delta \sigma_{(k+1)\tau,n}^{2}$ evaluated at $\eta$ and $\xi$ is linear in $\sigma_{t}^{2}$. Such a property is consistent with affine option pricing models (see Khrapov and Renault (2014)), but unfortunately cannot be implemented in non-affine frameworks such as our setup.

Indeed, if we use the affine conditional mean specification $\mu_{k,T,n} = r + (\lambda - 1/2) \sigma_{k,T,n}^{2}$ and we identify the kernel parameters by imposing the existence condition $E \left[ N_{k,T,n} | F_{(k-1)\tau,n} \right] = 1$, we obtain the following
solutions:
\[
\begin{align*}
\xi & = 0, \\
\delta & = r(1 + \phi), \\
\eta & = -\phi \left( \lambda - \frac{1}{2} \right) \tau - \frac{1}{2} \phi^2 \tau.
\end{align*}
\]

The remaining parameter $\phi$ is determined by $\mathbb{E} \left[ \exp \left( \Delta Y_{k\tau,n} \right) N_{k\tau,n} \mid \mathcal{F}_{(k-1)\tau,n} \right] = \exp \left( r\tau \right)$.\(^{10}\) Thus, since $\xi = 0$, the variance dependent pricing kernel reduces to the standard conditional Esscher transform.

On the other hand, one can show that using the exponential linear pricing kernel for the Heston and Nandi model leads to the same dynamics as that coming from the variance dependent kernel in (3.22), but with stochastic market prices of risk. This provides us with an added flexibility to consider richer dynamics on the variance risk preferences and investigate their impact on option pricing. Moreover, unlike in affine settings, we observe that in the context of a Gaussian non-affine GARCH model, the variance dependent exponential linear pricing kernel with constant risk preferences (i.e. $\theta^{(1)}_{k\tau,n} = \theta^{(1)}$ and $\theta^{(2)}_{k\tau,n} = \theta^{(2)}$, for all $k = 0, \ldots, nT$) is not consistent with the absence of arbitrage opportunities. Indeed, if we take the same conditional mean specification as above and substitute it into (3.19), we obtain $\theta^{(1)} = \lambda$ and $\theta^{(2)} = 0$, which corresponds to a pricing kernel with no variance risk premium. Thus, we require at least one risk preference to be stochastic. The choice of $\theta^{(2)}_{k\tau,n}$ will be of particular interest in Section 4 when we compute different GARCH diffusion limits.

### 3.3 Conditional non-Gaussian economies

Since financial asset returns empirically exhibit negative skewness and leptokurtosis, it is natural to investigate alternative non-Gaussian pricing frameworks. In this section we present two ways for handling the bivariate cumulant generating function of $(\epsilon, \epsilon^2)$ in a non-Gaussian setting. First, we propose a second-order Taylor approximation for $\kappa_{(\epsilon, \epsilon^2)}(\cdot, \cdot)$ without specifying a parametric distribution for the GARCH innovations. Second, we identify two specific non-Gaussian cases for which we can explicitly compute this quantity.

#### 3.3.1 A second-order Taylor expansion

The next result derives the risk-neutral GARCH dynamics when we approximate the bivariate cumulant generating function of $(\epsilon, \epsilon^2)$ by its second-order Taylor expansion.

**Corollary 3.3** The risk-neutral dynamics of the asset return when $\kappa_{(\epsilon, \epsilon^2)}(\cdot, \cdot)$ is replaced by its second-
order Taylor expansion is given by:

\[ \Delta Y_{k\tau,n} = \left( r - \frac{1}{2} \sigma_{k\tau,n}^2 \right) \tau + \sqrt{\tau} \sigma_{k\tau,n} \epsilon_{k\tau,n}, \quad \epsilon_{k\tau,n} \sim \mathcal{N}(0,1), \]
\[ \Delta \sigma_{(k+1)\tau,n}^2 = \alpha_0(\tau) + \alpha_1(\tau) \sigma_{k\tau,n}^2 (\epsilon_{k\tau,n} + A_{k\tau,n} + M_3 B_{k\tau,n} - \gamma)^2 + (\beta_1(\tau) - 1) \sigma_{k\tau,n}^2, \]

where \( A_{k\tau,n} \) and \( B_{k\tau,n} \) are provided in (3.4)-(3.5).

Notice that under this approximation, the risk-neutral variance is the same as the physical one (i.e. \( \sigma_{k\tau,n}^2 = \sigma_{k\tau,n}^2 \)). Moreover, the GARCH innovations are conditional Gaussian distributed under \( Q_n \).

The martingale equation (3.6) leads to another explicit dependence between \( \theta_{(1)}^{(1)} \) and \( \theta_{(2)}^{(2)} \), slightly different from the one obtained in the Gaussian case in (3.20):

\[ \theta_{k\tau,n}^{(1)} = \frac{1}{2} + \frac{\mu_{k\tau,n} - r}{\sigma_{k\tau,n}^2} - 2 \frac{\alpha_1(\tau)}{\sqrt{\tau}} \sigma_{k\tau,n} \left( \frac{M_3}{2} - \gamma \right) \theta_{k\tau,n}^{(2)}. \]

Equation (3.25) can be used as an alternative relationship between the risk preference parameters in the numerical implementation of non-Gaussian GARCH option pricing models, especially for cases when solving (3.6) is computationally demanding.

### 3.3.2 Conditional Gaussian mixture innovations

The Gaussian mixture distribution (GM) has been previously used for option pricing in a GARCH setup (see e.g. Badescu et al. (2008) or Rombouts and Stentoft (2015)). However, the empirical performance of such model is, to our knowledge, yet to be investigated under a variance dependent pricing kernel. As in the Gaussian case, we shall use the results in Proposition 3.1 to derive the risk-neutral dynamics.

First, we assume that the driving noise is distributed according to a conditional Gaussian mixture density, \( \epsilon_{k\tau,n} \sim \text{GM}(p_i, m_i, h_i) \) with \( I \) components, where the \( p_i \)'s are the mixing probabilities with \( \sum_{i=1}^{I} p_i = 1 \), and the \( m_i \)'s and the \( h_i \)'s are the conditional means and variances of the Gaussian random variables which satisfy:

\[ \sum_{i=1}^{I} p_i m_i = 0, \quad \text{and} \quad \sum_{i=1}^{I} p_i (m_i^2 + h_i) = 1, \]

so that the total distribution has mean zero and variance one. The conditional cumulant generating function of a Gaussian mixture is given by:

\[ \kappa_{\epsilon}(z) = \log \left( \sum_{i=1}^{I} p_i \exp \left( z m_i + \frac{z^2}{2} h_i \right) \right). \]

\(^{11}\)We notice that, despite starting with non-Gaussian GARCH innovations under \( P \), we obtain a Gaussian driving noise under \( Q_n \). This property is strongly connected to the approximation used and will not hold under higher order Taylor expansions for \( \kappa_{\epsilon}(\sigma_{k\tau,n}^2) \). We note that deriving the risk-neutral dynamics under a third-order Taylor approximation is not only more tedious but also requires information on the fifth moment of \( \epsilon_{k\tau,n} \).
Replacing this identity into (3.14) we obtain the following expression for the bivariate conditional cumulant generating function:

\[
\kappa_{(\epsilon, \epsilon')}(z_1, z_2) = \log \left( \sum_{i=1}^{l} p_i \exp \left( -\frac{1}{2} \log (1 - 2z_2 h_i) - \frac{m_i^2}{2h_i} + \frac{1}{2} \left( z_1 + \frac{m_i}{h_i} \right)^2 \frac{h_i}{1 - 2z_2 h_i} \right) \right). \tag{3.27}
\]

We can now use Proposition 3.1 once again to characterize the risk-neutral return evolution.

**Corollary 3.4** If \( \epsilon_{k\tau,n} F_{(k-1)\tau,n} \sim \text{GM}(p_i, m_i, h_i) \), then the risk-neutral dynamics for the asset returns under \( Q_n \) are given by:

\[
\Delta Y_{k\tau,n} = \tau \log \left( \sum_{i=1}^{l} \exp \left( \left( \sqrt{\tau} \sigma_{k\tau,n} (A_{k\tau,n} + m_i h_i) + \frac{\tau \sigma_{k\tau,n}^2}{2} \frac{h_i}{1 - 2B_{k\tau,n} h_i} \right) \right) \right)
+ \sqrt{\tau} \sigma_{k\tau,n}^2 P_{k\tau,n} + \tau R_{k\tau,n} \sigma_{k\tau,n} \epsilon_{k\tau,n}, \quad \epsilon_{k\tau,n}^* \sim \text{GM}(p_i, m_i, h_i), \tag{3.28}
\]

\[
\Delta \sigma^2_{(k+1)\tau,n} = a_0(\tau) + a_1(\tau) \sigma_{k\tau,n}^2 \left( \sqrt{R_{k\tau,n}} \epsilon_{k\tau,n}^* + P_{k\tau,n} - \gamma \right)^2 + (\beta_1(\tau) - 1) \sigma_{k\tau,n}^2. \tag{3.29}
\]

Here, \( p_i, m_i, h_i \) are given by:

\[
p_i = \frac{\exp \left( -\frac{1}{2} \log (1 - 2B_{k\tau,n} h_i) - \frac{m_i^2}{2h_i} + \frac{1}{2} \left( A_{k\tau,n} + \frac{m_i}{h_i} \right)^2 \frac{h_i}{1 - 2B_{k\tau,n} h_i} \right)}{\sum_{i=1}^{l} \exp \left( -\frac{1}{2} \log (1 - 2B_{k\tau,n} h_i) - \frac{m_i^2}{2h_i} + \frac{1}{2} \left( A_{k\tau,n} + \frac{m_i}{h_i} \right)^2 \frac{h_i}{1 - 2B_{k\tau,n} h_i} \right)}, \tag{3.30}
\]

\[
m_i = \left( A_{k\tau,n} + \frac{m_i}{h_i} \right) \frac{h_i}{1 - 2B_{k\tau,n} h_i} \frac{1}{\sqrt{R_{k\tau,n}}} - \frac{P_{k\tau,n}}{\sqrt{R_{k\tau,n}}}, \tag{3.31}
\]

\[
h_i = \frac{1}{1 - 2B_{k\tau,n} h_i} \frac{1}{R_{k\tau,n}}. \tag{3.32}
\]

where \( P_{k\tau,n} \) and \( R_{k\tau,n} \) are given by:

\[
P_{k\tau,n} = \sum_{i=1}^{l} p_i h_i \left( A_{k\tau,n} + \frac{m_i}{h_i} \right) \frac{1}{1 - 2B_{k\tau,n} h_i}, \tag{3.33}
\]

\[
R_{k\tau,n} = \sum_{i=1}^{l} p_i h_i^2 \left( A_{k\tau,n} + \frac{m_i}{h_i} \right)^2 \frac{1}{(1 - 2B_{k\tau,n} h_i)^2} + \sum_{i=1}^{l} p_i h_i \frac{1}{1 - 2B_{k\tau,n} h_i} - P_{k\tau,n}^2. \tag{3.34}
\]

We notice that under the exponential linear pricing kernel the asset returns also have a conditional Gaussian mixture distribution with time varying parameters for the Gaussian mixing components. Thus the underlying distributional assumption is also stable under the risk-neutral measure \( Q \).

### 3.3.3 Conditional mixed-Exponential innovations

The mixed-exponential (ME) distribution has been used by Cai and Kou (2011) in the modeling of the jump sizes in a new class of jump diffusions option pricing models. Furthermore, this model can also
be used to approximate Lévy processes. The ME distribution is dense with respect to the class of all distributions in the sense of weak convergence (see Botta and Harris (1986)). In this subsection we show that the use of such a density in modeling the GARCH innovations also leads to a closed-form expression for the bivariate cumulant generating function of \((\epsilon, \epsilon^2)\).

We assume that \(\epsilon_{k,T,n}\) follows a mixed-exponential distribution, \(\epsilon_{k,T,n} \sim \text{ME}(p_u, p_d, p_i, q_j, a_i, b_i)\), where the parameters satisfy the standard constraints: \(p_u \geq 0, \quad q_d = 1 - p_u \geq 0, \quad p_i \in (-\infty, \infty), \quad i = 1, \ldots, I; \quad \sum_{i=1}^I p_i = 1, \quad q_j \in (-\infty, \infty), \quad j = 1, \ldots, J; \quad \sum_{j=1}^J q_j = 1 \text{ and } a_i > 1, \quad i = 1, \ldots, I, \quad b_j > 0, \quad j = 1, \ldots, J.\) In addition, the parameters \(p_i\) and \(q_j\) need to satisfy some conditions to guarantee that the p.d.f. is always non-negative and is a true probability density function. From Cai and Kou (2011), a simple sufficient condition is \(\sum_{i=1}^I p_i a_i \geq 0, \) for all \(i = 1, \ldots, I,\) and \(\sum_{j=1}^J q_j b_j \geq 0, \) for all \(j = 1, \ldots, J.\)

The conditional p.d.f. \(f_{\epsilon}\) of \(\epsilon_{k,T,n}\) is given by:

\[
f_{\epsilon}(y) = p_u \sum_{i=1}^I p_i a_i \exp(-a_i y) \mathbb{I}_{y \geq 0} + q_d \sum_{j=1}^J q_j b_j \exp(b_j y) \mathbb{I}_{y < 0}.
\]  

(3.35)

The joint cumulant generating function is calculated in the following proposition.

**Proposition 3.5** When \(\epsilon_{k,T,n} \sim \text{ME}(p_u, p_d, p_i, q_j, a_i, b_i)\) conditional on \(\mathcal{F}_{(k-1)T,n}\), the joint cumulant generating function of \((\epsilon_{k,T,n}, \epsilon^2_{k,T,n})\) is given by:

\[
\kappa_{(\epsilon, \epsilon^2)}(z_1, z_2) = \log \left( p_u \sum_{i=1}^I p_i a_i e^{-\frac{(z_1 - a_i z_2)^2}{4 z_2}} \sqrt{\frac{\pi}{z_2}} \Phi \left( \frac{a_i - z_1}{\sqrt{2 z_2}} \right) + q_d \sum_{j=1}^J q_j b_j e^{-\frac{(z_1 + b_j)^2}{4 z_2}} \sqrt{\frac{\pi}{z_2}} \Phi \left( \frac{z_1 + b_j}{\sqrt{2 z_2}} \right) \right)
\]  

(3.36)

Unfortunately, even though we obtain a closed-form expression for the bivariate cumulant generating function of \((\epsilon_{k,T,n}, \epsilon^2_{k,T,n})\), we are not able to fully characterize the risk-neutral dynamics of the asset returns since we cannot identify the law of \(\epsilon^2_{k,T,n}\) under \(Q_n\). In that case, option prices can be computed by simulating asset paths under \(P\) and making use of the closed-form Radon-Nikodym derivative.

### 4 Diffusion limits under the exponential linear SDF

In this section we derive the diffusion limit of the risk-neutralized process in Proposition 3.1 based on the same parametric assumptions used to establish the weak convergence of the asset prices under the physical measure. The main result is contained in the next proposition.

**Proposition 4.1** Assume that the parametric conditions in (2.4) hold. Then, as \(\tau\) approaches zero, the risk-neutral continuous-time extended processes \(\{Y_{t,n}, \sigma_{t,n}^2\}\) of (3.8)-(3.9) converge weakly to the
following bivariate diffusions \((Y_t, \sigma_t^2)\) that satisfy:

\[
dY_t = \left( r - \frac{1}{2} \sigma_t^2 \right) dt + \sigma_t dB_{1t},
\]

\[
d\sigma_t^2 = \left( \omega_0 \left( \omega_1 + \sqrt{\omega_2 (M_3 - 2\omega_3)} \frac{\mu_t - r + \sigma_t^2/2}{\sigma_t} + \omega_2 (M_4 - M_3^2 - 1) \theta_t^{(2)} \sigma_t^2 \right) \sigma_t^2 \right) dt
+ \sqrt{\omega_2 (M_3 - 2\gamma)} \sigma_t^2 dB_{1t} + \sqrt{\omega_2} \sqrt{M_4 - M_3^2 - 1} \sigma_t^2 dB_{2t}.
\]

(4.2)

Here \(B_{1t}^*\) and \(B_{2t}^*\) are two independent Brownian motions on \(\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, Q\}\), and \(\mu_t\) and \(\theta_t^{(2)}\) are the continuous-time limits of \(\mu_{k\tau,n}\) and \(\theta_{k\tau,n}^{(2)}\), respectively.

In order for the above limit to exist, we require that \(M_4 M_2 M_3 + 1\). We notice that the drift of \(\sigma_t^2\) only depends on the continuous version of the market price of volatility risk \(\theta_t^{(2)}\). In fact, unlike the discrete-time GARCH case, the continuous limit of the martingale condition (3.6) leads to a closed-form expression which allows us to express the market price of equity risk \(\theta_t^{(1)}\) as a linear function of \(\theta_t^{(2)}\).

This observation is explained in more detail in the Appendix.

The risk-neutral dynamics in (4.1)-(4.2) can also be obtained by applying the Girsanov theorem (see e.g. Karatzas and Shreve (1988)) to the continuous time limit of the GARCH diffusion model from (2.5)-(2.6). Indeed, if we define the following Radon-Nikodym process with respect to the Brownian motion \(\{B_{1t}, B_{2t}\}\) on \(\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, P\}\):

\[
\frac{dQ}{dP} \bigg|_{\mathcal{F}_T} = \exp \left( - \sum_{i=1}^{2} \left( \int_0^T \nu_t^{(i)} dB_{it} + \frac{1}{2} \int_0^T (\nu_t^{(i)})^2 dt \right) \right),
\]

the results in Proposition 4.1 are recovered by taking:

\[
\nu_t^{(1)} = \frac{\mu_t - r + \sigma_t^2}{\sigma_t^2},
\]

\[
\nu_t^{(2)} = \sqrt{\omega_2} \sqrt{M_4 - M_3^2 - 1} \theta_t^{(2)} \sigma_t^2.
\]

(4.3)

(4.4)

This can be easily verified by taking the two independent Brownian motions in (4.1)-(4.2) according to the prescription \(dB_{1t}^* = dB_{1t} + \nu_t^{(1)} dt\) and \(dB_{2t}^* = dB_{2t} + \nu_t^{(2)} dt\). Various choices of \(\mu_{k\tau,n}\) and \(\theta_{k\tau,n}^{(2)}\) show that our GARCH diffusion model nests some important special cases. A discussion in this sense is provided below.

### 4.1 Conditional mean specifications and uniqueness of weak solutions

The choice of a specific conditional mean return is in general justified by numerical tractability arguments even though it is often the case that \(\mu_{k\tau,n}\) contains extra model parameters which have an economic...
interpretation. Thus, we propose the following conditional mean specification to address both issues:

$$
\mu_{k\tau,n} = r + \lambda \sigma_{k\tau,n} - \frac{\kappa_e (\sqrt{\tau} \sigma_{k\tau,n})}{\tau}.
$$

(4.5)

First, as it is done in other studies on GARCH option pricing models, we can interpret $\lambda$ as the unit risk premium parameter. Indeed, if we compute the expected simple gross return under (4.5), we get:

$$
E[\exp(\Delta Y_{k\tau,n})|\mathcal{F}_{(k-1)\tau,n}] = E[\exp(\mu_{k\tau,n}\tau + \sqrt{\tau} \sigma_{k\tau,n}\epsilon_{k\tau,n})|\mathcal{F}_{(k-1)\tau,n}] = \exp((r + \lambda \sigma_{k\tau,n})\tau).
$$

When $\tau = 1$, equation (4.5) reduces to the conditional mean return used in Christoffersen et al. (2010). Although this specification does not improve the numerical tractability when computing option prices in discrete time, it simplifies the drift term in the conditional variance of the GARCH diffusion model from (4.2). Under this assumption, the market price of equity risk in Girsanov’s theorem is constant which can be shown by calculating the continuous limit of $\mu_{k\tau,n}$. Indeed, using a second-order Taylor expansion for $\kappa_e (\sqrt{\tau} \sigma_{k\tau,n})$ around the origin, we have:

$$
\kappa_e (\sqrt{\tau} \sigma_{k\tau,n}) = \frac{1}{2} \tau \sigma_{k\tau,n}^2 + o(\tau).
$$

Letting $\tau$ approach zero, we obtain $\mu_t = r + \lambda \sigma_t - \sigma_t^2/2$, which leads to $\mu_t^{(1)} = \lambda$.

In order to analyze the risk-neutral GARCH diffusion model under the above mean specification in more detail, we introduce some simplifying notations: we let $a = -\omega_1 - \lambda \sqrt{\omega_2} (M_3 - 2 \omega_3)$, $b = -\omega_2 (M_4 - M_3^2 - 1)$, and $c = \omega_0$.

Since the diffusion parameters $\omega_0$, $\omega_1$, $\omega_2$ and $\omega_3$ are obtained based on the implied GARCH parameters, they obey certain constraints. This leads to $c \geq 0$ and $b \leq 0$.

Using the notations from (2.9) we can re-write (4.1)-(4.2) as follows:

$$
dS_t = r S_t dt + \sigma_t S_t dB_{1t}^*,
$$

(4.6)

$$
d\sigma_t^2 = \left(c + a \sigma_t^2 + b \theta_t^{(2)}(2) \sigma_t^4\right) dt + \zeta \sigma_t^2 dB_{3t}^*.
$$

(4.7)

Here, $B_{3t}^*$ is a standard Brownian motion under $Q$ such that $dB_{1t}^* \cdot dB_{3t}^* = \rho dt$. If we assume that the variance risk preference process is a time-homogeneous function of $\sigma_t$ (i.e. $\theta_t^{(2)} = \theta^{(2)}(\sigma_t)$) with $\theta^{(2)}(\cdot) \in L_{loc}(J)$, we can show that the two-dimensional diffusion defined in (4.6)-(4.7) has a unique weak solution. This follows immediately by verifying the Engelbert-Schmidt conditions introduced in (2.10) in Section 2 and from the fact that $S_t$ can be uniquely represented as the stochastic exponential related to a stochastic integral of $\sigma_t$ under $Q$ (i.e. $S_t = \mathcal{E}(L_t)$ with $L_t := \int_0^t \sigma_s dB_{1s}^*$).

\footnote{These constraints follow from the stationarity conditions imposed to the underlying GARCH model. Moreover, when financial returns are estimated using GARCH models, it is generally found that $c \approx 0$ (c is of the order $10^{-6}$) and $M_3 < 0$.}
4.2 Nested GARCH diffusion models and financial bubbles

In this subsection we analyze different choices of discrete-time versions of $\theta_i^{(2)}$ as a function of $\sigma_t$ and show how some of the existing GARCH diffusion option pricing models can be obtained as particular cases of our model. In particular, we shall focus on two main classes of diffusion processes embedded in (4.7).

The first class is represented by the generalized Hull-White models with affine volatility drift and it can be obtained by taking either $\theta_i^{(2)} = 0$ or $\theta_i^{(2)} = C/\sigma_i^2$ for some real constant $C$ and for all $k = 0, \ldots, nT$. The first case corresponds to the conditional Esscher transform for which we recover the continuous time limit obtained in Badescu et al. (2015b) and where we obtain that $\nu_i^{(2)} = 0$ in (4.4). The corresponding limiting diffusion can hence be obtained by applying the minimal martingale measure in continuous time to the model in (2.5)-(2.6). The second case leads to the same type of diffusion, but with a different volatility drift coefficient. Assuming a particular value for the constant $C$, we recover another important GARCH diffusion limit. Indeed, if we let $C = -\lambda M_3/\sqrt{\omega_2} (M_4 - M_3^2 - 1)$, we obtain the continuous time limit derived in Badescu et al. (2015b) for the underlying GARCH model (2.1)-(2.2) risk-neutralized via the extended Girsanov principle. The limiting diffusion can also be obtained in that case by applying the Girsanov change of measure in continuous time with $\nu_i^{(2)} = -\lambda M_3/\sqrt{M_4 - M_3^2 - 1}$. Finally, taking $C = \lambda M_3 (E_{\Psi}/F_{\Psi} - M_3)/\sqrt{\omega_2} (M_4 - M_3^2 - 1)$, we obtain the GARCH diffusion option pricing model corresponding to Duan’s (1999) generalized LRNVR derived in Badescu et al. (2015a), where $E_{\Psi} = E \left[ \epsilon_{k\tau,n}^2 \Psi (\epsilon_{k\tau,n}) | F_{(k-1)\tau,n} \right]$, $F_{\Psi} = E \left[ \epsilon_{k\tau,n} \Psi (\epsilon_{k\tau,n}) | F_{(k-1)\tau,n} \right]$ and $\Psi(\cdot) = \Phi (F_{\cdot}(\cdot))$. This can also be obtained via Girsanov’s theorem with $\nu_i^{(2)} = \lambda M_3 (E_{\Psi}/F_{\Psi} - M_3)/\sqrt{M_4 - M_3^2 - 1}$. We can write the resulting generalized Hull-White variance limit equation under $Q$ as:

$$d\sigma_t^2 = \left(c + a' \sigma_t^2 \right) dt + \zeta \sigma_t^2 dB_t^Q,$$  \hspace{1cm} (4.8)

where $a' = a$ if $\theta_i^{(2)} = 0$ and $a' = a + bC$ if $\theta_i^{(2)} = C/\sigma_i^2$.

The second class of processes is represented by the generalized geometric mean-reverting diffusion process with affine drift (see e.g. Metcalf and Hassett (1995) or Ewald and Yang (2007)). These models are obtained by taking the variance risk preference parameter to be a non-zero constant (i.e. $\theta_i^{(2)} = \theta^{(2)} \neq 0, \quad k = 0, \ldots, nT$). This further implies that $\theta_i^{(2)} = \theta^{(2)}$, and substituting this value into the variance equation (4.7), we have the following dynamics:

$$d\sigma_t^2 = \left(c + a\sigma_t^2 + b \theta^{(2)} \sigma_t^4 \right) dt + \zeta \sigma_t^2 dB_t^Q.$$ \hspace{1cm} (4.9)

Under the above specification, the drift of the variance process is a quadratic function of $\sigma_t^2$. We notice that when the variance risk premium is negative (i.e. $\theta^{(2)} < 0$) the coefficient of $\sigma_t^4$ is non-negative since $b \leq 0$.

In the remaining part of this subsection, we investigate the presence of financial bubbles in markets...
driven by the GARCH diffusion limit from (4.6)-(4.7). Since a financial bubble exists when the discounted underlying price process is a strict local martingale (see Protter (2013)), we are interested in identifying under what circumstances the underlying stock price in our model is a true martingale under $Q$. More specifically, following Bernard et al. (2015), we derive a necessary and sufficient condition for the martingale property, which involves checking local integrability of various deterministic test functions. Here we restrict our attention to the special case of a generalized Hull-White variance specification from (4.8), since the proof for the generalized geometric mean-reverting case is similar but rather tedious. The main result is given below.

**Proposition 4.2** The asset price $\{S_t\}_{0 \leq t < T}$ in (4.6) with the variance specification from (4.8) is a true martingale if and only if $\rho \leq 0$, or equivalently, if and only if $M_3 \leq 2\gamma$.

Note that the above condition is automatically satisfied for asset returns which exhibit negative skewness if the leverage parameter $\gamma$ is positive.

### 4.3 Explicit density function of GARCH volatility diffusion limit

In this section, we derive closed-form expressions for the probability density functions of the risk-neutral variance processes for both classes of diffusion limits given in (4.8) and (4.9). The following results characterize the joint density functions related to $\sigma_t^2$ under $Q$.

**Proposition 4.3** If the risk-neutral variance $\sigma_t^2$ satisfies the dynamics in (4.8), the joint probability density function of $\left(\sigma_t^2, \int_0^t \sigma_s^{-2} ds, \int_0^t \sigma_s^{-4} ds\right)$ for any $t \in (0, \infty)$ is given by:

$$Q \left( \int_0^t \sigma_s^{-4} ds \in dg, \int_0^t \sigma_s^{-2} ds \in dy, \sigma_t^2 \in dz \right) = \exp \left( -c \frac{z^{-1} + \frac{c}{2} - \sigma_0^{-2}}{\zeta^2} - c \left( \frac{a'}{\zeta^2} - 1 \right) y - \frac{\nu^2 \zeta^2 t}{2} \right)$$

$$\times \frac{\zeta \nu^{-1}}{8\sigma_0^{2\nu}} \cdot e_{ig} \left( \frac{\zeta^2 t}{8}, \frac{y \zeta}{2}, \frac{\sigma_0^{-2} + z^{-1}}{\zeta}, \frac{2(\sigma_0^2 z)^{-\frac{1}{2}}}{\zeta} \right) dg dy dz.$$

Here $\nu = \frac{a'}{\zeta^2} - \frac{1}{2}$ and $e_{ig}(v, t, z, x)$ is the function defined on page 645 of Borodin and Salminen (2002).

Note that the density function of $\sigma_t^2$ can be obtained by integrating out $g$ and $y$.

For the generalized mean-reverting case, it is not possible to obtain the probability density explicitly when $c \neq 0$ in (4.9). However, since $c$ is usually a very small non-negative number, we shall only consider the derivation for the case $c = 0$.

**Proposition 4.4** If the risk-neutral variance $\sigma_t^2$ satisfies the dynamics in (4.9) with $c = 0$, the joint

\footnote{Note that equations (4.6)-(4.7) ensure that the asset price process is a strict local martingale under $Q$.}
probability density function of \( \left( \sigma_t^2, \int_0^t \sigma_s^2 \, ds, \int_0^t \sigma_s^4 \, ds \right) \) for any \( t \in (0, \infty) \) is given by:

\[
Q \left( \int_0^t \sigma_s^2 \, ds \in dg, \int_0^t \sigma_s^2 \, ds \in dy, \sigma_t^2 \in dz \right) = \exp \left( \frac{\beta_2^{(2)}}{\xi^2} \left( z - \frac{\beta_0^{(2)} g}{2} \right) - \frac{\beta_2^{(2)} g y - \nu^2 \xi^2 t}{2} \right) \times \frac{\zeta^{\nu-1}}{8 \sigma_0^{2\nu}} \cdot \text{ei}_g \left( \frac{\xi^2 t}{8}, \frac{y \xi}{2}, \frac{\sigma_0^2 + z}{\xi}, \frac{2(\sigma_0^2 + z)^{1/2}}{\xi} \right) \, dg \, dy \, dz.
\]

Here \( \nu = \frac{\alpha}{\xi^2} - \frac{1}{2} \) and \( \text{ei}_g(v,t,z,x) \) is the function defined on page 645 of Borodin and Salminen (2002).

As in the previous case, the density function of \( \sigma_t^2 \) is obtained by integrating out \( g \) and \( y \).

5 Empirical analysis

In this section we investigate the in-sample and out-of-sample pricing performances of the asymmetric NGARCH model in (2.1)-(2.2) based on Gaussian and Gaussian mixture innovations and that has been risk-neutralized via the exponential pricing kernel introduced in Section 3. We conduct this study using an extensive dataset of European calls on the S&P500 index. The empirical assessment will be carried out for the above models with parameters obtained using historical information about option prices and returns of the underlying in different combinations. This will result in two different model estimation approaches that are used separately in the in-sample and out-of-sample empirical exercises, namely, the joint likelihood and the sequential estimation methods, respectively.

5.1 Data description

The empirical pricing performance is tested using two datasets of S&P500 call options obtained from OptionMetrics, whose prices were quoted during the period spanning January 1st, 2004–December 31st, 2013. Both datasets comprise contracts with maturities between 20 and 250 days and moneyness between 0.9 and 1.1. In order to only use significant contracts, we applied various filters similar to those introduced in Bakshi et al. (1997). The first dataset, called Sample A, consists of 20,912 call prices quoted every Wednesday for the reference period and is used for both the joint and sequential estimation exercises. The second dataset, called Sample B contains 21,228 call prices recorded every Thursday for the same period and is only used for the out-of-sample performance assessment. The basic features of these two datasets, including the number of contracts, average prices, and implied volatilities are reported in tables 1 and 2 for an array of different maturities and moneyness intervals. The moneyness is defined as the ratio between the future price and the strike price \( (Mo = F/K) \), so options with \( Mo < 1 \) are out-of-money (OTM) and those with \( Mo > 1 \) are in-the-money (ITM). The average price and implied volatility for Sample A are $57.767 and 18.6%, respectively, while the corresponding values for Sample B are $58.132 and 18.9%, respectively.
The return sample dataset using in both empirical analyses consists of daily returns which cover the period January 1st, 1995–December 22nd, 2013. The mean and variance of the asset returns are $2.8799 \cdot 10^{-4}$ and $1.5408 \cdot 10^{-4}$, respectively, while the skewness and kurtosis are $-0.2405$ and $10.8928$, respectively. In each of the empirical exercises, we shall use different subsets of this series as it is explained in the following subsections.

5.2 Estimation methodology

The in-sample and out-of-sample performance assessments are conducted using different estimation strategies based on three different likelihoods: the returns likelihood, the option likelihood, and the joint likelihood based on returns and option data.

The returns likelihood is constructed for the NGARCH model in (2.1)-(2.2) sampled at a daily frequency (i.e. $\tau = 1$), together with a conditional mean specification like in (4.5), that is:

$$\mu_t = r + \lambda \sigma_t - \kappa_t (\sigma_t).$$

(5.1)

We note that the parameter $r$ in the mean specification (5.1) has not been estimated at the time of maximizing the likelihood but has been set in advance equal to the average one year T-Bill rate for the corresponding period. We assume that the GARCH innovations follow either a Gaussian distribution or a Gaussian mixture distribution with two components. In the former case, we have $\epsilon_t | F_{t-1} \sim N(0, 1)$ and the likelihood depends only on the model parameters $\theta := (\alpha_0, \alpha_1, \beta_1, \gamma, \lambda)$ that are subjected to standard constraints that ensure the positivity of the conditional variance process and the second order stationarity of the returns process. In the latter case, we let $\epsilon_t | F_{t-1} \sim GM((p_1, p_2), (m_1, m_2), (h_1, h_2))$, where the mixing probabilities satisfy $p_1 + p_2 = 1$ and the means and variances of the mixture components satisfy (3.26). These constraints reduce the number of parameters to be estimated and hence the Gaussian mixture likelihood depends exclusively on $\theta := (\alpha_0, \alpha_1, \beta_1, \gamma, \lambda, p_1, m_1, h_1)$.

Given a log-return value $r_t$ at time $t$, we denote the corresponding conditional returns log-likelihood by:

$$\log L^R_t (r_t, \theta) := \log \left( \frac{1}{\sigma_t} f_r \left( \frac{r_t - \mu_t}{\sigma_t} \right) \right).$$

Here, $f_r$ stands for the pdf of the model innovations, that is, $f_r(z) = 1/\sqrt{2\pi} \exp(-z^2/2)$ in the Gaussian case and

$$f_r(z) = p_1 \left( \frac{1}{\sqrt{2\pi h_1}} \exp \left( \frac{(z - m_1)^2}{2h_1} \right) \right) + p_2 \left( \frac{1}{\sqrt{2\pi h_2}} \exp \left( \frac{(z - m_2)^2}{2h_2} \right) \right),$$

for a mixture of two Gaussians. The log-likelihood corresponding to the sample $r := \{r_1, \ldots, r_T\}$ is denoted by $\log L^R_t (r_t, \theta) := \sum_{t=1}^{T} \log L^R_t (r_t, \theta)$.

We follow the approach of Trolle and Schwartz (2009) in deriving the options likelihood. For each model, we use a set of option market prices $C^{Mkt} := \{C^{Mkt}_1, \ldots, C^{Mkt}_N\}$, and assume that the vega
weighted option valuation errors $\xi_i$ defined by

$$\xi_i := \frac{C_{i \text{Mkt}} - C_{i \text{Mod}}}{BSV_{i \text{Mkt}}}, \quad i = 1, \ldots, N,$$

are independent and normally distributed with mean zero and variance $\bar{\xi}^2 := \frac{1}{N} \sum_{i=1}^{N} \xi_i^2$. In this expression $C_{i \text{Mod}}$ is the model price and $BSV_{i \text{Mkt}}$ is the corresponding Black-Scholes vega of the option. Given a market option price $C_{i \text{Mkt}}$, we let the corresponding conditional options log-likelihood given by:

$$\log L_0(C_{i \text{Mkt}}, \theta, \theta^O) := \log \left( \frac{1}{BSV_{i \text{Mkt}}} f_\xi \left( \frac{C_{i \text{Mkt}} - C_{i \text{Mod}}}{BSV_{i \text{Mkt}}} \right) \right),$$

Here, $f_\xi(z) = 1/(\sqrt{2\pi\bar{\xi}}) \exp(-z^2/(2\bar{\xi}^2))$, $\theta$ are the model parameters, and $\theta^O$ are the parameters of the pricing kernel that has been used to compute the model-based prices $C_{i \text{Mod}}$. The log-likelihood corresponding to the option prices $C_{i \text{Mkt}}$ is denoted by $\log L_0(C_{i \text{Mkt}}, \theta, \theta^O) := \sum_{i=1}^{N} \log L_0(C_{i \text{Mkt}}, \theta, \theta^O)$. Since the non-affine GARCH structure does not lead to semi-closed expressions, we evaluate the option prices using Monte-Carlo simulation. In general this can be performed in two equivalent ways, either by generating the asset paths directly under the risk-neutral measure, or by simulating under the physical measure and weighting the option payoff by the corresponding Radon-Nykodim derivative path. We follow the second approach due to the complicated form of the risk-neutral GARCH dynamics in the Gaussian mixture case (see Corollary 3.4). For each GARCH innovation distribution, we consider two special cases for the exponential linear pricing kernel from (3.1), depending on the form of the market price of variance risk $\theta^{(2)}_t$. First, we take $\theta^{(2)}_t = 0$, which corresponds to the conditional Esscher transform, while for the second specification we assume take $\theta^{(2)}_t := \theta^{(2)} / \bar{\xi}^2$, where $\theta^{(2)}$ is a real constant representing the variance risk aversion parameter. Throughout this section, we refer to this specification as the variance dependent pricing kernel. Note that in the conditional Esscher transform case, there are no parameters from the pricing kernel to be estimated since we impose $\theta^{(2)}_t = 0$ and $\theta^{(1)}_t$ is uniquely determined from the corresponding martingale constraint. For the second pricing kernel specification, we only estimate the variance risk aversion parameter, so that $\theta^O = \theta^{(2)}$. The market price of equity risk $\theta^{(1)}$ is then evaluated using the martingale equation (3.20) for Gaussian innovations, and the second-order Taylor approximation based martingale equation (3.25) in the Gaussian mixture case; several non-reported experiments that compare the Gaussian mixture GARCH option prices obtained using the solution of the exact martingale identity (3.6) with those of (3.25), show that the differences are not significant. The prices $C_{i \text{Mod}}$ are computed using 20,000 paths. These paths are initialized for the pricing of each contract by using the return of the underlying asset in the corresponding date and the spot volatility obtained from the historical return estimation. An important point that needs to be emphasized is that, following Eichler et al. (2011), we use the same random numbers in the generation of the Monte Carlo paths at each step of the maximization of the likelihood in order to reduce the stochastic noise and to make possible the optimization algorithm convergence. For discounting the expected option
payoff we use the corresponding period T-Bill rates interpolated to match the option maturity.

The joint likelihood estimation has become very popular in calibrating the model parameters using both historical returns and cross-section of option data (see e.g. Santa-Clara and Yan (2010), Christoffersen et al. (2013), and Ornthanalai (2014) among others). Given a set of market option prices $C_{\text{Mkt}}$ and historical log-return values $r$ at time $t$, we define the corresponding joint log-likelihood $\log L^J(r, C_{\text{Mkt}}, \theta, \theta^O)$ as a weighted function of the return and option likelihoods:

$$\log L^J(r, C_{\text{Mkt}}, \theta, \theta^O) := \frac{T + N}{2} \log L^R(r, \theta) + \frac{T + N}{2} \log L^O(C_{\text{Mkt}}, \theta, \theta^O)$$

(5.2)

Note that the pricing kernel parameters $\theta^O$ is present only in the option likelihood, while the GARCH model parameters $\theta$ appear in both likelihood functions.

5.3 Empirical results

The findings obtained in the in-sample and out-of-sample empirical exercises are described in the following subsections.

5.3.1 Joint likelihood estimation using returns and options

We provide an initial evidence of the modelling properties coming from the combination of using a non-zero variance risk aversion in the pricing kernel and the use of a GARCH model with Gaussian mixture innovations. The joint likelihood estimation is carried out in two steps.

First, we maximize the returns likelihoods corresponding to the NGARCH models with Gaussian and Gaussian mixture innovations constructed using 2,520 daily log-returns of the S&P500 index for the period December 22nd, 1999–December 30th, 2009. This step is not only useful for comparison purposes, but also to determine daily spot volatilities that will be used to initialize the Monte Carlo paths generated to carry out the option pricing. The results are illustrated in Table 3. We notice that the likelihood corresponding to the process with Gaussian mixture innovations exhibits a higher optimal value, which hints a better adequacy of this model to the data; this observation is confirmed by the ordering of the (not reported) AIC and BIC statistics. The values of the estimated Gaussian NGARCH model parameters are in the same range as those obtained in many other previous empirical studies. The implied persistency is 0.9917. The parameters do not change much for the Gaussian Mixture NGARCH model, the implied persistency being 0.9948. We notice that in the mixture case, the first Gaussian component has an associated probability of 92.22% and its mean and variance are 0.0804 and 0.8486, respectively. The second Gaussian component has a probability of 7.98% and exhibits a negative mean of -0.9534 and a higher variance of 1.8099. Thus, we argue that the first mixture component captures the
“business as usual” state of the economy, while the second models the “crash” component. The implied skewness of the NGARCH innovation is negative and the kurtosis is greater than that of a standard Gaussian distribution.

In the second step, we proceed by estimating these models using their joint likelihoods with the same historical returns series set and a subset of the Sample A option dataset that comprises prices quoted during the period January 1st, 2009 - December 31st, 2009 (the dataset contains 1,829 contracts). The estimation is implemented by maximizing the joint likelihood constructed as we described in Section 5.2 with respect to the model parameter values and, in the case of mixed Gaussian innovations, with respect to the distribution parameters also. Additionally, when the variance dependent pricing kernel is used, optimization is also carried out with respect to the variance premium parameter $\theta^{(2)}$. The results of this estimation exercise are presented in Table 4. The figures in this table evidence the importance of using non-Gaussian innovations combined with an exponential linear pricing kernel which contains a non-zero price of variance risk in this context. Indeed, when using Gaussian innovations, a maximum normalized joint log-likelihood of 2.0982 (respectively, 2.2802) is attained when using the conditional Esscher pricing kernel (respectively, the variance dependent pricing kernel). When comparing the performance of the conditional Esscher transform with its variance dependent analog, the main gain in the latter is obtained out of the options part of the likelihood that goes from 0.65 to 0.80 for the optimal parameter values. In the Gaussian mixture case, the maximum normalized log-likelihood values attained with both the conditional Esscher (2.2846) and the variance dependent (2.3453) kernels are superior to those obtained in the Gaussian case, and the gain when going from Esscher to its variance dependent counterpart comes again from the options part of the likelihood that goes from 0.8092 to 0.8713. The returns likelihood does not vary much across the models.

We notice that the values of the NGARCH parameters in the joint estimation exercise differ significantly from those obtained using only on the returns likelihood. For example, in the Gaussian case the model persistence implied from both returns and options is 0.7514 when using the conditional Esscher transform and 0.7312 for the variance dependent kernel. There are also significant changes in the estimated parameters for the Gaussian mixture distribution. The first component used to describe “business as usual” market conditions has an associated probability of 98.76%, while the “crash” component happens in 1.24% of the cases. This latter component is very noisy with high negative means (-1.8975 for the conditional Esscher transform and -3.3821 for the variance dependent pricing kernel) and very high values for the corresponding variances (36.5387 for the conditional Esscher transform and 40.5314 for the variance dependent pricing kernel).

Finally, Figure 1 depicts the dependence of the joint log-likelihood on the variance premium parameter $\theta^{(2)}$. Note that all the model and distribution parameters have been kept fixed and set to their optimal values and it is only the variance premium parameter $\theta^{(2)}$ that varies. In both cases this function is non-convex, which hints the difficulties that can sometimes be faced at the time of finding the optimal variance premium value, as well as the interest of having preliminary estimates for it. In the Gaussian case, we
obtain $\theta^{(2)} = -0.328$, which indicates that the log-ratio of the risk-neutral and physical conditional return densities has a parabolic form. The negative value of the variance risk premium parameter is consistent with the findings of Christoffersen et al. (2013) and Bormetti et al. (2015) when variance dependent pricing kernels with constant market prices of risk are applied to affine Gaussian GARCH settings. For the Gaussian mixture NGARCH model, the above log-ratio will no longer have a parabolic form, so a negative value of $\theta^{(2)}$ is no longer expected. Our estimation results indicate that $\theta^{(2)} = 0.009$ in this case.

5.3.2 Sequential estimation and out-of-sample pricing performance

We carry out an extensive out-of-sample pricing performance assessment using the entire options dataset described in Tables 1 and 2. This section compares the performance of the NGARCH models with Gaussian and Gaussian mixture innovations when the pricing is carried out using exclusively the variance dependent pricing kernel.

The parameters of the models used in this study will be determined using a sequential estimation process (see Broadie et al. (2007), Christoffersen et al. (2013) among others) that is updated according to the scheme that we now describe: for the first Wednesday in Sample A of the study, a maximum likelihood estimation of the two models is performed using the historical daily returns of the underlying corresponding to the ten preceding years (2,520 daily observations). The obtained model parameters are then kept fixed and a value of the variance premium parameter $\theta^{(2)}$ that maximizes the options likelihood corresponding to the options quoted that Wednesday is computed. The model/pricing kernel parameters obtained using this sequential procedure are then used to price the options quoted the next day (the corresponding Thursday from Sample B), as well as those quoted the next Wednesday from Sample A. This procedure is repeated iteratively for every Wednesday from Sample A. Note that although the $\theta^{(2)}$ is calibrated weekly, the NGARCH and innovation parameters are re-estimated on a monthly basis using a rolling window of size 2,520 observations.

The pricing performance is measured using the Implied Volatility Root Mean Squared Error (IVRMSE) indicator, defined below:

$$IVRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (IV^{\text{Mod}} - IV^{\text{Mkt}})^2} \times 100.$$ 

The results of this exercise are presented in Table 5 for the Gaussian and Gaussian mixture NGARCH models in the pricing of contracts grouped in various moneyness and time-to-maturity intervals. The table contains three modules. In the top one, marked “In-Sample Error”, we report the average pricing errors committed each Wednesday of Sample A at the time of pricing the options that have been used to optimize the options likelihood with respect to $\theta^{(2)}$ the very same Wednesday. In the module marked “Next Day Pricing Error” we report the average pricing errors committed each Thursday from Sample
B, using the models whose parameters have been estimated the preceding day. We emphasize that even though in this case we use models that have been calibrated the day before, the spot volatility necessary in the generation of Monte Carlo paths is updated. Finally, in the module marked “Next Week Pricing Error” we report the average pricing errors committed each Wednesday from Sample A using the models whose parameters have been estimated the Wednesday of the preceding week.

The figures in Table 5 show that the mixture distribution consistently outperforms the Gaussian counterpart both in and out-of-sample. The overall IVRMSE of the Gaussian mixture NGARCH model for the three exercises reported are: 3.334 (in sample), 3.648 (next day) and 3.802 (next week), which are all smaller than the corresponding values for the Gaussian NGARCH model: 3.547 (in sample), 3.963 (next day) and 4.041 (next week). This overall improvement could be in principal attributed to the performance of the mixture model for longer maturity options. For example, for the in-sample scenario, the improvement is of 10.9% for deep OTM options and 22.2% for deep ITM options for maturities between 180 and 250 days, while for the next week out-of-sample exercise, the improvement is 12.3% for OTM options to 20.3% for ITM options for the same maturity category. This fact can be potentially explained by the effect of the “crash” component of the mixture density for the long maturity options.

We notice that although the Gaussian mixture NGARCH outperforms the Gaussian NGARCH for almost all ATM and ITM option classes, there are a few cases where its performance is worse. For instance, this is the case for deep OTM shorter maturity options.

6 Conclusions

This paper investigates the pricing and weak convergence of an asymmetric non-affine, non-Gaussian GARCH model when the risk-neutralization is based on a variance dependent exponential linear pricing kernel with stochastic market prices of risk. The risk-neutral dynamics is derived for a general setting and specific distributional choices for the driving noise are further discussed. We emphasize the importance of using stochastic market prices of risk, since a standard variance dependent kernel with constant prices of risk cannot be applied in our setting.

The weak limit of our general risk-neutral GARCH model is derived under standard parametric constraints. The resulting stochastic volatility process generalizes the class of GARCH diffusion option pricing models from the financial literature. We show that for several specific choices of the variance risk aversion parameter, the GARCH diffusion limits constructed via the conditional Esscher transform, extended Girsanov principle, and Duan’s generalized LRNVR are obtained as special cases of our model. Moreover, there is a one-to-one relationship between the market prices of risk from the variance dependent pricing kernel in the GARCH case and their counterparts from the Girsanov’s theorem applied to the GARCH diffusion limit. For two particular diffusions, we derive explicit expressions for the density of the process variance. Finally, we investigate the presence of financial bubbles by deriving necessary and sufficient conditions for the discounted continuous risk-neutral GARCH diffusion limit to be a true
We provide a detailed empirical analysis to illustrate the importance of combining a non-Gaussian distribution for the GARCH innovations with the proposed pricing kernel. In a first numerical experiment, using both historical returns and a cross-section of options, we perform a joint likelihood estimation procedure to calibrate the model and pricing kernel parameters for a Gaussian NGARCH and a Gaussian mixture NGARCH model with two components. Our results indicate that the Gaussian mixture NGARCH based on a non-zero market price of variance risk provides the best fit. Next, using ten years of options data we test the out-of-sample performance for the above option pricing models constructed solely with the variance dependent kernel. As in the previous case, the Gaussian mixture model consistently outperforms its Gaussian counterpart, especially for longer maturity ITM options for which the improvement is around 22%. The Gaussian model provides a slightly better alternative only for shorter maturity OTM options.

The proposed framework can be further extended to a more sophisticated parametric mixture structure where the volatility of each component follows its own GARCH dynamic, or to a non-parametric setting based on a kernel density estimator for the GARCH innovations. The pricing methodology can also be used for volatility derivatives and joint estimation based on asset returns and VIX data.
7 Appendix

7.1 Proof of Proposition 3.1

(i) Using the fact that \( \theta_{k,n}^{(1)} \) and \( \theta_{k,n}^{(2)} \) are \( \mathcal{F}_{k,n} \)-predictable processes, we proceed by computing the one-step conditional mean of \( N_{k,n} \) from relation (3.1). For any \( k = 0, \ldots, nT \) we have:

\[
E\left[N_{k,n}|\mathcal{F}_{(k-1)\tau,n}\right] = E\left[\exp\left(-\theta_{k,n}^{(1)}\Delta Y_{k,n} - \theta_{k,n}^{(2)}\Delta \sigma_{n(k+1)\tau,n} - \kappa(\Delta Y, \Delta \sigma) \left(-\theta_{k,n}^{(1)}, -\theta_{k,n}^{(2)}\right)\right)|\mathcal{F}_{(k-1)\tau,n}\right] = 1.
\]

Thus, the martingale property of \( Z_n \) follows immediately:

\[
E\left[Z_{k,n}|\mathcal{F}_{(k-1)\tau,n}\right] = Z_{(k-1)\tau,n} E\left[N_{k,n}|\mathcal{F}_{(k-1)\tau,n}\right] = Z_{(k-1)\tau,n}.
\]

Since \( Z_{T,n} := dQ_n/dP_n > 0 \) is non-negative by construction, the fact that \( Q_n \) is an equivalent probability measure with respect to \( P_n \) follows from:

\[
E^*[1] = E[Z_{T,n}] = E[Z_{0,n}] = 1.
\]

Here we denote by \( E^*[\cdot] \) the expectation under the risk-neutral measure \( Q_n \). The proof that \( Q_n \) is a risk-neutral measure is at the end of (ii).

(ii) Since \( \epsilon_{k,n} := \sqrt{R_{k,n}}\epsilon_{k,n}^* + P_{k,n} \) and both \( P_{k,n} \) and \( R_{k,n} \) are \( \mathcal{F}_{k,n} \)-predictable, we have that \( \mathcal{F}_{k,n} = \sigma(\epsilon_0, \ldots, \epsilon_{k,n}) = \sigma(\epsilon_0^*, \ldots, \epsilon_{k,n}^*) \). Then we compute the conditional cumulant generating function of \( \epsilon_{k,n} \) under \( Q_n \). Next, we evaluate the conditional moment generating function of \( \epsilon_{k,n} \) under \( Q_n \), denoted by \( \kappa^*_\epsilon(z) \), as a function of the bivariate cumulant generating function of \( (\epsilon_{k,n}, \epsilon_{2,k,n}^2) \) under \( P_n \). Let \( u > 0 \) and for any \( z \in (-u, u) \). Using (3.3) we have:

\[
\kappa^*_\epsilon(z) = \log E^*\left[\exp\left(\epsilon_{k,n}^*\right)|\mathcal{F}_{(k-1)\tau,n}\right] = \log E\left[\exp\left(\epsilon_{k,n}^*\right)|\mathcal{F}_{(k-1)\tau,n}\right] = \log E\left[\exp\left(\epsilon_{k,n}^*\right)|\mathcal{F}_{(k-1)\tau,n}\right] = \log E\left[\exp\left(\epsilon_{k,n}^* + A_{k,n}\epsilon_{k,n}^* + B_{k,n}\epsilon_{k,n}^2 - \kappa(\epsilon, \epsilon^2) (A_{k,n}, B_{k,n})\right)|\mathcal{F}_{(k-1)\tau,n}\right] = \kappa(\epsilon, \epsilon^2) (z + A_{k,n}, B_{k,n}) - \kappa(\epsilon, \epsilon^2) (A_{k,n}, B_{k,n}).
\]

It follows that the first and second raw moments of \( \epsilon_{k,n} \) under \( Q_n \) are given by:

\[
E^*\left[\epsilon_{k,n}|\mathcal{F}_{(k-1)\tau,n}\right] = \frac{d\kappa^*_\epsilon(z)}{dz}|_{z=0} = \frac{\partial\kappa(\epsilon, \epsilon^2)(A_{k,n}, B_{k,n})}{\partial z_1} := P_{k,n},
\]

\[
\text{Var}^*\left[\epsilon_{k,n}|\mathcal{F}_{(k-1)\tau,n}\right] = \frac{d^2\kappa^*_\epsilon(z)}{dz^2}|_{z=0} = \frac{\partial^2\kappa(\epsilon, \epsilon^2)(A_{k,n}, B_{k,n})}{\partial z_1^2} := R_{k,n}.
\]

We define \( \epsilon_{k,n}^* = (\epsilon_n - P_n)/\sqrt{R_n} \) the new innovation process under \( Q_n \) with mean zero and unit variance, and denote its distribution by \( D^*(0, 1) \). In order to show that \( \epsilon_{k,n}^* \) are conditionally uncorrelated,
we let $l > s > 1$ and have:

$$
\mathbb{E}^* \left[ e^{*_{k,t,n} e^{*_{(k-s)\tau,n}} \mathcal{F}_{(k-l)\tau,n}} \right] = \mathbb{E}^* \left[ e^{*_{k,t,n} e^{*_{(k-s)\tau,n}} \mathcal{F}_{(k-l)\tau,n}} \big| \mathcal{F}_{(k-l)\tau,n} \right] = \mathbb{E}^* \left[ e^{*_{(k-s)\tau,n} e^{*_{k,t,n}} \mathcal{F}_{(k-l)\tau,n}} \big| \mathcal{F}_{(k-l)\tau,n} \right] = 0.
$$

Now we show that if the market prices of risk $\delta_{k,t,n}^{(1)}$ and $\delta_{k,t,n}^{(2)}$ satisfy (3.6) for any $k = 0, \ldots, nT$, then discounted asset prices are martingales under $Q_n$ (i.e. $Q_n$ is a risk-neutral measure). This is equivalent to showing that $\kappa^*_\Delta Y(1) = r \tau$ holds for all $k = 0, \ldots, nT$. Indeed we have:

$$
\kappa^*_\Delta Y(1) = \log \mathbb{E}^* \left[ \exp (\Delta Y_{k,t,n}) \big| \mathcal{F}_{(k-l)\tau,n} \right] = r \mu_{k,t,n} + \kappa^*_e \left( \sqrt{\tau} \sigma_{k,t,n} \right).
$$

Using (7.1), the above martingale constraint becomes:

$$
\mu_{k,t,n} = r - \frac{1}{\tau} \left( \kappa_{(e,\epsilon^2)} (A_{k,t,n} + \sqrt{\tau} \sigma_{k,t,n}, B_{k,t,n}) - \kappa_{(e,\epsilon^2)} (A_{k,t,n}, B_{k,t,n}) \right),
$$

which coincides with (3.6).

(iii) The risk-neutral dynamics (3.8)-(3.9) follow by substituting (3.6) and (3.7) into (2.1)-(2.2).

### 7.2 Proof of Corollary 3.2

Using the expression in (3.15) for the conditional bivariate cumulant generating function in the Gaussian case we evaluate the first and second partial derivatives with respect to $z_1$ and we find:

$$
P_{k,t,n} = \frac{A_{k,t,n}}{1 - 2B_{k,t,n}} \quad \text{and} \quad R_{k,t,n} = \frac{1}{1 - 2B_{k,t,n}}.
$$

Moreover, for any $x \in \mathbb{R}$ we have:

$$
\kappa_{(e,\epsilon^2)} (z_1 + x, z_2) - \kappa_{(e,\epsilon^2)} (z_1, z_2) = \frac{1}{1 - 2x_2} \left( z_1 x + \frac{x^2}{2} \right).
$$

Evaluating (7.4) at $z_1 = P_{k,t,n}$, $z_2 = R_{k,t,n}$ and $x = \sqrt{\tau} \sigma_{k,t,n}$, and substituting this, together with (7.3), into the general risk-neutral dynamics from (3.8)-(3.9), we obtain the desired equations (3.16)-(3.17) for the Gaussian case. The martingale constraint is obtained by substituting the same quantities into equation (3.6). The only remaining thing to be shown is that $e^{*_{k,t,n}} \sim \mathcal{N}(0, 1)$ given $\mathcal{F}_{k,t,n}$. Using (3.10), we compute the conditional cumulant generating function of $e^{*_{k,t,n}}$ under $Q_n$. For any $z \in \mathbb{R}$ we have:

$$
\kappa^*_{e^*} (z) = -z \frac{P_{k,t,n}}{\sqrt{R_{k,t,n}}} + \kappa_{(e,\epsilon^2)} \left( \frac{z}{\sqrt{R_{k,t,n}}} + A_{k,t,n}, B_{k,t,n} \right) - \kappa_{(e,\epsilon^2)} (A_{k,t,n}, B_{k,t,n})
$$

$$
= -z \frac{P_{k,t,n}}{\sqrt{R_{k,t,n}}} + \frac{1}{1 - 2B_{k,t,n}} \left( \frac{z^2}{2R_{k,t,n}} + z \frac{A_{k,t,n}}{\sqrt{R_{k,t,n}}} \right).
$$
Replacing (7.3) into the above relationship we obtain that $\kappa^*_k(z) = z^2/2$, and therefore $\epsilon^*_{k\tau,n}$ follow standard Gaussian distributions under $Q_n$. ■

### 7.3 Proof of Corollary 3.3

Using a second-order Taylor expansion around the origin, we can approximate $\kappa_{(e,\varepsilon^2)}(z_1, z_2)$ by:

$$
\kappa_{(e,\varepsilon^2)}(z_1, z_2) \approx \kappa_{(e,\varepsilon^2)}(0, 0) + z_1 \frac{\partial \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_1} + z_2 \frac{\partial \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_2} + \frac{1}{2} \left( z_1^2 \frac{\partial^2 \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_1^2} + 2z_1z_2 \frac{\partial^2 \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_1 \partial z_2} + z_2^2 \frac{\partial^2 \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_2^2} \right). 
$$

(7.5)

Using the relationships between joint moments and cumulants we find:

$$
\kappa_{(e,\varepsilon^2)}(0, 0) = 0, \quad \frac{\partial \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_1} = 0, \quad \frac{\partial \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_2} = 1, \\
\frac{\partial^2 \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_1^2} = 1, \quad \frac{\partial^2 \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_1 \partial z_2} = M_3, \quad \frac{\partial^2 \kappa_{(e,\varepsilon^2)}(0, 0)}{\partial z_2^2} = M_4 - 1.
$$

Substituting the above in (7.5) we obtain the following approximation:

$$
\kappa_{(e,\varepsilon^2)}(z_1, z_2) \approx z_2 + \frac{1}{2} \left( z_1^2 + 2z_1z_2 M_3 + z_2^2 (M_4 - 1) \right).
$$

(7.6)

Using a similar approach to the previous proof, we first find:

$$
P_{k\tau,n} = A_{k\tau,n} + M_3 B_{k\tau,n} \quad \text{and} \quad R_{k\tau,n} = 1.
$$

It follows that:

$$
\kappa_{(e,\varepsilon^2)} \left( A_{k\tau,n} + \sqrt{\tau} \sigma_{k\tau,n}, B_{k\tau,n} \right) - \kappa_{(e,\varepsilon^2)} \left( A_{k\tau,n}, B_{k\tau,n} \right) = \frac{\tau \sigma^2_{k\tau,n}}{2} + \sqrt{\tau} \sigma_{k\tau,n} \left( A_{k\tau,n} + M_3 B_{k\tau,n} \right).
$$

Substituting the above relations into (3.8)-(3.9) leads to the dynamics from (3.23)-(3.24). The martingale constraint (3.25) follows immediately. Using (3.10), the conditional cumulant generating function of $\epsilon^*_{k\tau,n}$ under $Q_n$ is given by:

$$
\kappa^*_k(z) = -z \frac{P_{k\tau,n}}{\sqrt{R_{k\tau,n}}} + \kappa_{(e,\varepsilon^2)} \left( \frac{z}{\sqrt{R_{k\tau,n}}} + A_{k\tau,n}, B_{k\tau,n} \right) - \kappa_{(e,\varepsilon^2)} \left( A_{k\tau,n}, B_{k\tau,n} \right) \\
= -z \left( A_{k\tau,n} + M_3 B_{k\tau,n} \right) + \kappa_{(e,\varepsilon^2)} \left( z + A_{k\tau,n}, B_{k\tau,n} \right) - \kappa_{(e,\varepsilon^2)} \left( A_{k\tau,n}, B_{k\tau,n} \right) = \frac{z^2}{2}.
$$

Therefore, $\epsilon^*_{k\tau,n} \sim N(0, 1)$ under $Q_n$. ■
7.4 Proof of Proposition 3.4

Taking the first and second partial derivatives of $\kappa_{(\epsilon, \epsilon^2)}(z_1, z_2)$ from (3.27) with respect to $z_1$ at $z_1 = A_k, n$ and $z_2 = B_k, n$, we obtain the desired relationships for $P_k, n$ and $Q_k, n$ from (3.33) and (3.34), where the risk-neutral weights, denoted by $p_i^{*, k}, n$, are given by (3.30). As in the Gaussian case, we next evaluate for any real-valued $x$ the following quantity:

$$\kappa_{(\epsilon, \epsilon^2)}(z_1 + x, z_2) - \kappa_{(\epsilon, \epsilon^2)}(z_1, z_2) = \log \left( \sum_{i=1}^{l} p_i^*(z_1, z_2) \exp \left( \left( \left( z_1 + \frac{m_i}{h_i} \right) x + \frac{x^2}{2} \right) \frac{h_i}{1 - 2z_2 h_i} \right) \right), \quad (7.7)$$

where $p_i^*(z_1, z_2)$ are given by:

$$p_i^*(z_1, z_2) = \frac{p_i \exp \left( -\frac{1}{2} \log (1 - 2z_2 h_i) - \frac{m_i^2}{2h_i} + \frac{1}{2} \left( z_1 + \frac{m_i}{h_i} \right) \frac{h_i}{1 - 2z_2 h_i} \right)}{\sum_{i=1}^{l} \frac{p_i \exp \left( -\frac{1}{2} \log (1 - 2z_2 h_i) - \frac{m_i^2}{2h_i} + \frac{1}{2} \left( z_1 + \frac{m_i}{h_i} \right) \frac{h_i}{1 - 2z_2 h_i} \right)}{}}.$$

Taking $z_1 = A_k, n$, $z_2 = B_k, n$ and $x = \sqrt{7} \sigma_k, n$, and using the fact that $p_i^*(A_k, n, B_k, n) = p_i^*, k, n$, we obtain the risk-neutral dynamics from (3.28)-(3.29) together with the corresponding martingale constraint. Finally, we verify that $\epsilon_k^*, n$ follows a conditional Gaussian mixture distributions with mean zero and unit variance and have its parameters given in (3.28). We have

$$\kappa_{\epsilon^*}^* (z) = -z \frac{P_{k, n}}{\sqrt{R_{k, n}}} + \kappa_{(\epsilon, \epsilon^2)} \left( \frac{z}{\sqrt{R_{k, n}}} + A_{k, n}, B_{k, n} \right) - \kappa_{(\epsilon, \epsilon^2)} (A_{k, n}, B_{k, n})$$

$$= -z \frac{P_{k, n}}{\sqrt{R_{k, n}}} + \log \left( \sum_{i=1}^{l} p_i^{*, k, n} \exp \left( \left( A_{k, n} + \frac{m_i}{h_i} \right) \frac{z}{\sqrt{R_{k, n}}} + \frac{z^2}{2R_{k, n}} \frac{h_i}{1 - 2B_{k, n} h_i} \right) \right)$$

$$= \log \left( \sum_{i=1}^{l} p_i^{*, k, n} \exp \left( m_i^{*, k, n} z + \frac{h_i^{*, k, n} z^2}{2} \right) \right),$$

where $m_i^{*, k, n}$ and $h_i^{*, k, n}$ are those given in (3.31)-(3.32).
7.5 Proof of Proposition 3.5

We proceed as follows:

\[
\kappa_{(\epsilon, \epsilon^2)}(z_1, z_2) = E \left[ \exp \left( z_1 \epsilon_{k\tau, n} + z_2 \epsilon_{k\tau, n}^2 \right) \mid \mathcal{F}_{(k-1)\tau, n} \right] = \int_{-\infty}^{\infty} \exp \left( z_1 y + z_2 y^2 \right) f_{\epsilon}(y) \, dy
\]

\[
= \int_{-\infty}^{\infty} \exp \left( z_1 y + z_2 y^2 \right) \left( p_u \sum_{i=1}^{l} p_i a_i \exp (-a_i y) 1_{y \geq 0} + q_d \sum_{j=1}^{f} q_j b_j \exp (b_j y) 1_{y < 0} \right) \, dy
\]

\[
= p_u \sum_{i=1}^{l} p_i a_i \int_{0}^{\infty} \exp \left( (z_1 - a_i) y + z_2 y^2 \right) \, dy + q_d \sum_{j=1}^{f} q_j b_j \int_{-\infty}^{0} \exp \left( (z_1 - b_j) y + z_2 y^2 \right) \, dy
\]

\[
= p_u \sum_{i=1}^{l} p_i a_i e^{-\frac{(z_1 - a_i)^2}{2z_2}} \sqrt{\frac{\pi}{z_2}} \Phi \left( \frac{a_i - z_1}{\sqrt{z_2}} \right) + q_d \sum_{j=1}^{f} q_j b_j e^{-\frac{(z_1 - b_j)^2}{2z_2}} \sqrt{\frac{\pi}{z_2}} \Phi \left( \frac{z_1 + b_j}{\sqrt{z_2}} \right).
\]

Here we used the following identity

\[
\int_{0}^{\infty} \exp \left( c(y + d)^2 \right) \, dy = \sqrt{\frac{\pi}{cd}} \Phi(-\sqrt{cd}), \quad c > 0.
\]

7.6 Proof of Proposition 4.1

First, we notice that although the GARCH innovations under \( Q_n \) have mean zero and unit variance, we do not have any information concerning the higher moments. In particular, unlike in the physical world, we do not know whether the higher moments are independent of the time step \( \tau \) or not. Thus, we make the following notations for the limiting risk-neutral raw moments, provided that they exist:

\[
M_i^* = \lim_{\tau \to 0} E^* \left[ (\epsilon^*)^i_{k\tau, n} \mid \mathcal{F}_{(k-1)\tau, n} \right]. \tag{7.8}
\]

We follow the standard approach on weak convergence analysis to compute the drift \( \Psi(X_t) \) and the diffusion \( \Sigma(X_t) \) of the risk-neutral limiting diffusion:

\[
dx_t = \Psi(X_t) \, dt + \Sigma(X_t) \, dB_t^*.
\tag{7.9}
\]

Here, \( X_t = (Y_t, \sigma_t^2)^T \) and \( B_t^* = (B_{1t}^*, B_{2t}^*)^T \), with \( B_{1t}^* \) and \( B_{2t}^* \) being two independent standard Brownian motions under \( Q \). To proceed, we evaluate the conditional first and second limiting risk-neutral moments of \( \Delta Y_{k\tau, n} \) and \( \sigma_{(k+1)\tau, n}^2 \). All limits considered are conditional on the filtration \( \mathcal{F}_{(k-1)\tau, n} \). First, using (3.4)-(3.5) and the parametric constraints, we notice that both \( A_{k\tau, n} \) and \( B_{k\tau, n} \) are \( \mathcal{O}(\sqrt{\tau}) \) as \( \tau \) approaches zero such that:

\[
\lim_{\tau \to 0} \frac{A_{k\tau, n}}{\sqrt{\tau}} = -\theta_t^{(1)} \sigma_t + 2\gamma \sqrt{\omega_t^2} \sigma_t^{(2)} \quad \text{and} \quad \lim_{\tau \to 0} \frac{B_{k\tau, n}}{\sqrt{\tau}} = -\sqrt{2} \theta_t^{(2)} \sigma_t^2. \tag{7.10}
\]
Examining the quantities to compute, we notice that there are a few types of special limits which show up throughout the proof, and we deal with them below:

- First limit type: \( \lim_{\tau \to 0} \frac{\kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n}, B_{k\tau,n})}{\tau} \).
- Second limit type: \( \lim_{\tau \to 0} \frac{P_{k\tau,n}}{\sqrt{\tau}} \).
- Third limit type: \( \lim_{\tau \to 0} R_{k\tau,n} \) and \( \lim_{\tau \to 0} \frac{1-R_{k\tau,n}}{\sqrt{\tau}} \).

For the first type of limits it is sufficient to use a second-order Taylor expansion for \( \kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n}, B_{k\tau,n}) \) around the origin. Using results from the previous proof, we can write:

\[
\kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n}, B_{k\tau,n}) = B_{k\tau,n} + \frac{1}{2}(A_{k\tau,n}^2 + 2A_{k\tau,n}B_{k\tau,n}M_3 + B_{k\tau,n}^2(M_4 - 1)) + o(\tau). \tag{7.11}
\]

For the second type of limits, we only need a first-order Taylor expansion of the first derivative of \( \kappa_{(\epsilon,\epsilon^2)}(z_1, z_2) \) with respect to \( z_1 \) about the origin. Using the above results, we can write:

\[
P_{k\tau,n} := \frac{\partial \kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n}, B_{k\tau,n})}{\partial z_1} = A_{k\tau,n} + M_3B_{k\tau,n} + o(\sqrt{\tau}). \tag{7.12}
\]

For the third type of limits we shall also perform a first-order Taylor expansion of the second derivative of \( \kappa_{(\epsilon,\epsilon^2)}(z_1, z_2) \) with respect to \( z_1 \) about the origin. Using the following joint cumulant relationships:

\[
\frac{\partial^3 \kappa_{(\epsilon,\epsilon^2)}(0,0)}{\partial z_1^3} = M_3 \quad \text{and} \quad \frac{\partial^3 \kappa_{(\epsilon,\epsilon^2)}(0,0)}{\partial z_1^2 \partial z_2} = M_4 - 1,
\]

we have:

\[
R_{k\tau,n} := \frac{\partial^2 \kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n}, B_{k\tau,n})}{\partial z_1^2} = 1 + M_3A_{k\tau,n} + (M_4 - 1)B_{k\tau,n} + o(\sqrt{\tau}). \tag{7.13}
\]

Using (7.10) and (7.11), we have:

\[
\lim_{\tau \to 0} \frac{1}{\tau} E^* \left[ \Delta Y_{k\tau,n} \bigg| F^*_n \right] = \lim_{\tau \to 0} \left( r + \frac{\sigma_{k\tau,n} P_{k\tau,n}}{\sqrt{\tau}} + \kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n}, B_{k\tau,n}) - \frac{\kappa_{(\epsilon,\epsilon^2)}(A_{k\tau,n} + \sqrt{\tau} \sigma_{k\tau,n}, B_{k\tau,n})}{\tau} \right)
\]

\[
= r + \sigma_t \lim_{\tau \to 0} \frac{A_{k\tau,n} + M_3B_{k\tau,n}}{\sqrt{\tau}} - \sigma_t^2 \lim_{\tau \to 0} \frac{A_{k\tau,n} + M_3B_{k\tau,n}}{\sqrt{\tau}}
\]

\[
= r - \frac{\sigma_t^2}{2}.
\]
Using (7.10), (7.12) and (7.13), we can calculate the limiting expected conditional variance:

\[
\lim_{\tau \to 0} \frac{1}{\tau} E^* \left[ \Delta \sigma^2_{(k+1)_{\tau}, n} \mid F_{(k-1)\tau,n} \right] = \lim_{\tau \to 0} \frac{\alpha_0(\tau)}{\tau} + \sigma_i^2 \lim_{\tau \to 0} \frac{\alpha_1(\tau)(1 + \gamma^2) + \beta_1(\tau) - 1}{\tau} - \sigma_i^2 \lim_{\tau \to 0} \frac{\alpha_1(\tau)}{\tau}
\]

\[
+ \sigma_i^2 \lim_{\tau \to 0} \frac{\alpha_1(\tau)}{\tau} E^* \left[ R_{\tau,n}(\epsilon^*_{\tau,n})^2 + P_{\tau,n} - 2\gamma P_{\tau,n} \left| F_{(k-1)\tau,n} \right| \right]
\]

\[
= \omega_0 - \frac{\omega_1 \sigma_i^2}{\sqrt{\omega_2 \sigma_i^2}} \lim_{\tau \to 0} \frac{R_{\tau,n} - 1}{\sqrt{\tau}} - 2\gamma \sqrt{\omega_2 \sigma_i^2} \lim_{\tau \to 0} \frac{P_{\tau,n}}{\sqrt{\tau}}
\]

\[
= \omega_0 - \frac{\omega_1 \sigma_i^2}{\sqrt{\omega_2 \sigma_i^2}} \left( M_3 \lim_{\tau \to 0} \frac{A_{\tau,n}}{\sqrt{\tau}} + (M_4 - 1) \lim_{\tau \to 0} \frac{B_{\tau,n}}{\sqrt{\tau}} \right)
\]

\[
= 2\gamma \sqrt{\omega_2 \sigma_i^2} \left( \lim_{\tau \to 0} \frac{A_{\tau,n}}{\sqrt{\tau}} + M_3 \lim_{\tau \to 0} \frac{B_{\tau,n}}{\sqrt{\tau}} \right)
\]

\[
= \omega_0 - \omega_1 \sigma_i^2 \sqrt{\omega_2 (M_3 - 2\gamma) \theta_i^{(1)} \theta_i^{(2)} \sigma_i^4} - \omega_2 \left( M_4 - 1 + 4\gamma^2 - 4\gamma M_3 \right) \theta_i^{(2)} \sigma_i^4.
\]

Furthermore, we can express \( \theta_i^{(1)} \) as a function of \( \theta_i^{(2)} \) by taking the limit of the martingale equation (3.6):

\[
\theta_i^{(1)} = \frac{1}{\sigma_i^2} \left( \mu_t - r + \frac{\sigma_i^2}{2} \right) - \sqrt{\omega_2} (M_3 - 2\gamma) \theta_i^{(2)} \sigma_i.
\]

Replacing this into the conditional first moment of the conditional variance process, the drift term of the limiting diffusion is given by:

\[
\Psi(Y_t, \sigma_i^2) = \left( \omega_0 - \left( \omega_1 + \sqrt{\omega_2} (M_3 - 2\omega_3) \frac{\mu_t - r + \sigma_i^2/2}{\sigma_i} + \omega_2 \left( M_4 - M_3^2 - 1 \right) \theta_i^{(2)} \sigma_i^2 \right) \sigma_i^2 \right) \tag{7.14}
\]

The second-order moments are computed in a similar fashion. We have:

\[
\lim_{\tau \to 0} \frac{1}{\tau} \text{Var}^* \left[ \Delta Y_{\tau,n} \mid F_{(k-1)\tau,n} \right] = \lim_{\tau \to 0} \frac{1}{\tau} \text{Var}^* \left[ R_{\tau,n} \text{Var}^* \left[ \epsilon_{\tau,n} \mid F_{(k-1)\tau,n} \right] \right] = \sigma_i^2.
\]

\[
\lim_{\tau \to 0} \frac{1}{\tau} \text{Var}^* \left[ \Delta \sigma^2_{(k+1)_{\tau}, n} \mid F_{(k-1)\tau,n} \right] = \lim_{\tau \to 0} \frac{1}{\tau} \text{Var}^* \left[ \frac{\alpha_1(\tau)}{\tau} \text{Var}^* \left[ \left( \sqrt{R_{\tau,n} \epsilon^*_{\tau,n}} + P_{\tau,n} - \gamma \right)^2 \mid F_{(k-1)\tau,n} \right] \right]
\]

\[
= \omega_2 \sigma_i^2 \text{Var}^* \left[ \left( \sqrt{R_{\tau,n} \epsilon^*_{\tau,n}} + P_{\tau,n} - \gamma \right)^2 \mid F_{(k-1)\tau,n} \right]
\]

\[
= \omega_2 \left( M_3^2 - 1 + 4\gamma^2 - 4\gamma M_3 \right) \sigma_i^4.
\]

\[
\lim_{\tau \to 0} \frac{1}{\tau} \text{Cov}^* \left[ \Delta Y_{\tau,n}, \Delta \sigma^2_{(k+1)_{\tau}, n} \mid F_{(k-1)\tau,n} \right] = \sqrt{\omega_2} \sigma_i^2 \lim_{\tau \to 0} \text{Cov}^* \left[ \sqrt{R_{\tau,n} \epsilon^*_{\tau,n}} \left( \sqrt{R_{\tau,n} \epsilon^*_{\tau,n}} + P_{\tau,n} - \gamma \right)^2 \mid F_{(k-1)\tau,n} \right]
\]

\[
= \sqrt{\omega_2} \left( M_3^2 - 2\gamma \right) \sigma_i^3.
\]

Therefore, the risk-neutral second moment matrix is given by:

\[
\Sigma(Y_t, \sigma_i^2) \Sigma^T(Y_t, \sigma_i^2) = \begin{pmatrix} \sigma_i^2 & \sqrt{\omega_2} (M_3^2 - 2\gamma) \sigma_i^3 \\ \sqrt{\omega_2} (M_3^2 - 2\gamma) \sigma_i^3 & \omega_2 \left( M_4 - 1 + 4\gamma^2 - 4\gamma M_3 \right) \sigma_i^4 \end{pmatrix}.
\]
Finally, using a Cholesky decomposition of $\Sigma(Y_t, \sigma_t^2)$, we obtain the diffusion coefficient of the GARCH diffusion limit:

$$
\Sigma(Y_t, \sigma_t^2) = \begin{pmatrix}
\sigma_t & 0 \\
\sqrt{\omega_2 (M_3^c - 2\gamma)} \sigma_t^2 & \sqrt{\omega_2 (M_4^c - (M_3^c)^2 - 1)} \sigma_t^2
\end{pmatrix}.
$$

(7.15)

Finally, we show that the limiting risk-neutral third and fourth moments defined in (7.8) coincide with the corresponding physical ones (i.e. $M_3^c = M_3$ and $M_4^c = M_4$). To do this, we compute the limiting third and fourth risk-neutral cumulants. The result follows from using (3.10) and first-order Taylor expansions around the origin for the third and fourth partial derivatives of $\kappa(x, \rho)(\cdot, \cdot)$ with respect to the first argument:

$$
K_3^c := \lim_{\tau \to 0} \frac{d^3 \kappa^c_s(z)}{dz^3} |_{z=0} = \lim_{\tau \to 0} \frac{1}{R_{kT,n}^{k/2}} \partial^3 \kappa(x, \rho) (A_{kT,n}, B_{kT,n}) = M_3,
$$

$$
K_4^c := \lim_{\tau \to 0} \frac{d^4 \kappa^c_s(z)}{dz^4} |_{z=0} = \lim_{\tau \to 0} \frac{1}{R_{kT,n}^{k/2}} \partial^4 \kappa(x, \rho) (A_{kT,n}, B_{kT,n}) = M_4 - 3.
$$

This concludes the proof. ■

7.7 Proof of Proposition 4.2

The main idea of the proof is to utilize the deterministic necessary and sufficient conditions for the martingale property in Proposition 4.1 of Bernard et al. (2015), and this involves checking the finiteness of some test functions evaluated at the boundaries. Using similar notations with Bernard et al. (2015), we let $b(x) = \sqrt{x}, \mu(x) = c + a'x, \sigma(x) = \zeta x$. Following Proposition 2.8 of Bernard et al. (2015), we introduce the measure $\tilde{Q}$ under which the risk-neutral variance process satisfies the following SDE:

$$
\sigma_t^2 = (c + a' \sigma_t^2 + \rho \zeta \sigma_t^2) dt + \zeta \sigma_t^2 dB_t.
$$

We denote $\zeta = \frac{4a'}{\lambda} - 1$, $\omega = \frac{4a'}{\lambda}$ and $\eta = \frac{2a'}{\lambda} \geq 0$, and compute the scale function under $\tilde{Q}$:

$$
\bar{s}(x) = \int e^{-\frac{y}{\lambda} \sqrt{\lambda + x} \cdot \sqrt{2a' + 2\rho \zeta n^3/2}} dy = C_1 \int x^{-\frac{z+1}{2}} e^{-\zeta \sqrt{\eta y} / y} dy, \quad x \in J,
$$

where $C_1$ is a positive constant. The other test functions are given below:

$$
\bar{v}(x) = C_2 \int x^{-\frac{z+3}{2}} e^{-\zeta \sqrt{\eta y} / y} \left( \int y^{-\frac{z+1}{2}} e^{-\zeta \sqrt{\eta z} / z} dz \right) dy,
$$

$$
\bar{v}_b(x) = C_2 \int x^{-\frac{z+1}{2}} e^{-\zeta \sqrt{\eta y} / y} \left( \int y^{-\frac{z+1}{2}} e^{-\zeta \sqrt{\eta z} / z} dz \right) dy.
$$
Note that there is an extra term $e^{-\eta/z}$ and/or $e^{\eta/y}$ in the above expressions compared to the original test functions for the correlated Hull-White stochastic volatility model (the special case corresponding to $c = 0$) given in Section 5.4 of Bernard et al. (2015), which we recall here

$$\bar{s}(x) = C_1 \int_x^y \frac{x+1}{y} e^{-\eta\sqrt{y}} dy, \quad x \in J,$$

$$\bar{v}(x) = C_2 \int_x^y \frac{x+1}{y} e^{\eta\sqrt{y}} \left( \int_y^x e^{-\eta\sqrt{z}} dz \right) dy,$$

$$\bar{v}_b(x) = C_2 \int_x^y \frac{x+1}{y} e^{\eta\sqrt{y}} \left( \int_y^x e^{-\eta\sqrt{z}} dz \right) dy.$$

We first establish the following lemma about the asymptotic relations of our test functions $\bar{v}, \bar{v}_b$ and the corresponding test functions $\bar{v}, \bar{v}_b$ for the Hull-White stochastic volatility model.

**Lemma 7.1** For $\eta \geq 0$, we have the following statements:

(i) $\bar{v}(x) \geq \bar{v}(x)$ for all $x \in J$, and similarly for $\bar{v}_b(x)$ and $\bar{v}_b(x)$.

(ii) $\bar{v}(x) \sim \bar{v}(x)$ as $x \to \infty$, and similarly for $\bar{v}_b(x)$ and $\bar{v}_b(x)$.

**Proof:** For (i), we have:

$$\bar{v}(x) = C_2 \int_x^y \frac{x+1}{y} e^{\eta\sqrt{y}} \left( \int_y^x e^{-\eta\sqrt{z}} dz \right) dy = C_2 \int_x^y \frac{x+1}{y} e^{\eta\sqrt{y}} \left( \int_y^x e^{-\eta\sqrt{z}} dz \right) dy \geq C_2 \int_x^y \frac{x+1}{y} e^{\eta\sqrt{y}} \left( \int_y^x e^{-\eta\sqrt{z}} dz \right) dy = : \bar{v}(x).$$

Note that in the last inequality we consider two cases: (1) if $x \geq y$, then $\eta/y \geq \eta/z$ for $z \in (y, x)$, and (2) if $x < y$, then $\eta/y < \eta/z$ for $z \in (x, y)$, but $\int_x^y e^{-\eta\sqrt{z}} e^{\eta/y} dy = -\int_y^x e^{-\eta\sqrt{z}} e^{\eta/y} dz$, and the inequality still follows. The proof for $\bar{v}_b(x)$ and $\bar{v}_b(x)$ can be done in a similar way.

For (ii), note that we have $\lim_{x \to \infty} e^{-\eta/z} = 1$ and $\lim_{y \to \infty} e^{\eta/y} = 1$, and the result follows. This completes the proof of the lemma.

In Table 6 we recall the classification table Table 5.7 from Proposition 5.11 of Bernard et al. (2015) which summarizes the essential information for applying their Proposition 4.1. Combining Table 6 with Lemma 7.1 (i) and (ii), we construct a similar classification for our model in (4.6), which is illustrated in Table 7.

The two “undetermined” entries in Table 7 are due to the fact that knowing $\bar{v}(\ell) < \infty$, combined with $\bar{v}(\ell) \geq \bar{v}(\ell)$ (from Lemma 7.1 (i)), is not sufficient to determine whether $\bar{v}(\ell)$ is infinite or not. However,
we observe that it does not matter in determining the martingale property of \( S \). For all parameter settings, from Table 7, we always have \( \bar{v}(\ell) = \infty \) and \( \bar{v}(r) = \infty \). This, together with Proposition 4.1 of Bernard et al. (2015), allows us to conclude that the stock price \( \{ S_t \}_{t \in [0, \ldots, T]} \) in (4.6) with the variance dynamics from (4.8) is a true martingale if and only if \( \bar{\varpi}(\gamma) = 1 \), which is equivalent to \( \varpi \leq 0 \) from Table 7. By definition, \( \varpi \leq 0 \) is equivalent to \( \rho \leq 0 \). The remaining result follows from the expression in (2.9). This completes the proof. ■

### 7.8 Proof of Proposition 4.3

The idea of the proof is to first introduce a measure \( \tilde{Q} \) with a simpler process under it (e.g. geometric Brownian motion), and then apply a measure change, so that the process will have the same SDE as (4.8) under the new measure. Thus, we consider under \( \tilde{Q} \) the following geometric Brownian motion

\[
\frac{d\sigma^2_t}{\sigma^2_t} = a' \sigma^2_t dt + \zeta \sigma^2_t dB^*_t,
\]

with \( \sigma^2_0 > 0 \) and state space \( J = (0, \infty) \). From Feller’s test of explosions, \( \sigma^2_t \) does not explode from \( J \) under \( \tilde{Q} \). Define \( f(x) = \exp(-c x^{-1}/\zeta^2) \), and it is easy to check that \( f(x) > 0, x \in J \) and \( f(\cdot) \in C^2(J) \).

The Radon-Nicodym derivative associated with the exponential measure change in Palmowski and Rolski (2002) is given by:

\[
\frac{dQ}{d\tilde{Q}} = \frac{f(\sigma^2_T)}{f(\sigma^2_0)} \exp \left( -\int_0^T \frac{\mathcal{L}^\tilde{Q} f(\sigma^2_t)}{f(\sigma^2_t)} dt \right).
\]  \hspace{1cm} (7.16)

Note that \( \mathcal{L}^\tilde{Q} \) is the infinitesimal generator under \( \tilde{Q} \) and is given by:

\[
\frac{\mathcal{L}^\tilde{Q} f(x)}{f(x)} = c \left( \frac{a'}{\zeta^2} - 1 \right) x^{-1} + \frac{c^2}{2\zeta^2} x^{-2}.
\]  \hspace{1cm} (7.17)

Using Girsanov’s theorem, the risk-neutral variance \( \sigma^2_t \) satisfy the following SDE under \( Q \):

\[
\frac{d\sigma^2_t}{\sigma^2_t} = (c + a' \sigma^2_t) dt + \zeta \sigma^2_t dB^*_t,
\]

which has the same unique-in-law weak solution as that of (4.8). Since \( c \geq 0 \), from Feller’s test of explosions, \( \sigma^2_t \) does not explode from \( J \) under \( Q \). Then following Palmowski and Rolski (2002), we conclude that \( Q \) is equivalent to \( \tilde{Q} \).
Now we have:

\[
Q \left( \int_0^t \sigma_s^{-4} ds \in dg, \int_0^t \sigma_s^{-2} ds \in dy, \sigma_t^2 \in dz \right) = E^* \left[ 1 \left\{ \int_0^t \sigma_s^{-4} ds \in dg, \int_0^t \sigma_s^{-2} ds \in dy, \sigma_t^2 \in dz \right\} \right]
\]

\[
= \tilde{E} \left[ f(\sigma_t^2) \exp \left( -\frac{c^2}{2\xi} \int_0^t \sigma_s^{-4} ds - c \left( \frac{\alpha'}{\xi^2} - 1 \right) \int_0^t \sigma_s^{-2} ds \right) \right]
\]

\[
= \frac{f(z)}{f(\sigma_0^2)} \exp \left( -\frac{c^2}{2\xi^2} g - c \left( \frac{\alpha'}{\xi^2} - 1 \right) y \right) \tilde{Q} \left( \int_0^t \sigma_s^{-4} ds \in dg, \int_0^t \sigma_s^{-2} ds \in dy, \sigma_t^2 \in dz \right)
\]

\[
= \exp \left( -\frac{c}{\xi^2} \left( z^{-1} + \frac{c_8}{2} - \sigma_0^{-2} \right) - c \left( \frac{\alpha'}{\xi^2} - 1 \right) y - \nu^2 \frac{c^2 t}{2} \right) \frac{\zeta \nu^{-1}}{8}\sigma_0^{2} \cdot e^{i\gamma \left( \frac{\zeta^2 t}{8}, \frac{y}{2}, \frac{\sigma_0^{-2} + z^{-1}}{\zeta}, \frac{2(\sigma_0^2 z)^{-\frac{3}{2}}}{\zeta} \right)} dg dy dz.
\]

Here we denoted \( \nu = \frac{\alpha'}{\xi^2} - \frac{1}{2} \). In the last equality we have used the formula (9.121.8) on page 620 of Borodin and Salminen (2002) with substitution \( \beta \mapsto -1 \). The special function \( e^{i\gamma(x,t,z)} \) is defined on page 645 of Borodin and Salminen (2002). This completes the proof. \( \blacksquare \)

### 7.9 Proof of Proposition 4.4

Following a similar idea as the proof of Proposition 4.3, we introduce a new measure \( \tilde{Q} \), under which we have the following geometric Brownian motion (with different coefficients than in the previous proposition) \( d\tilde{\sigma}_t^2 = a\sigma_t^2 dt + \zeta \sigma_t^2 dB_t^\tilde{Q} \) with \( \sigma_0^2 > 0 \) and state space \( J = (0, \infty) \). If we let \( k = \theta^{(2)} \) with \( k \geq 0 \) (i.e. this corresponds to a negative variance risk premium, \( \theta^{(2)} \leq 0 \)) and define \( f(x) = \exp(kx/\zeta^2) \), it is easy to check that \( f(x) > 0, x \in J \) and \( f(\cdot) \in C^2(J) \). Define the new measure \( Q \) similar in form as (7.16) with:

\[
\mathcal{L}^Q f(x) = \frac{ak}{\sigma^2 x} + \frac{k^2}{2\zeta^2 x^2}.
\]

It follows that under \( Q \) the risk-neutral variance satisfies the SDE below:

\[
d\tilde{\sigma}_t^2 = (a\sigma_t^2 + k\sigma_t^2)dt + \zeta \sigma_t^2 dB_t^\tilde{Q},
\]

which has the same unique-in-law weak solution as that of (4.9). From Feller’s test of explosions, we can check that \( \sigma_t^2 \) does not explode from \( J \) under \( Q \). Thus, \( Q \) is equivalent to \( \tilde{Q} \). The result follows in a similar way as the proof of Proposition 4.3 using the formula (9.121.8) on page 620 of Borodin and Salminen (2002) with substitution \( \beta \mapsto \alpha - 1 \) and \( \nu = \frac{\alpha}{\xi^2} - \frac{1}{2} \). This completes the proof. \( \blacksquare \)
References


<table>
<thead>
<tr>
<th>Maturities</th>
<th>Number of Contracts</th>
<th>Moneyness ($F/K$)</th>
<th>Across Maturities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>[0.900, 0.950]</td>
<td>[0.950, 0.975]</td>
</tr>
<tr>
<td>$T &lt; 30$</td>
<td></td>
<td>1862</td>
<td>1413</td>
</tr>
<tr>
<td>$30 \leq T &lt; 80$</td>
<td>3019</td>
<td>2224</td>
<td>2506</td>
</tr>
<tr>
<td>$80 \leq T &lt; 180$</td>
<td>1274</td>
<td>693</td>
<td>803</td>
</tr>
<tr>
<td>$180 \leq T \leq 250$</td>
<td>502</td>
<td>217</td>
<td>346</td>
</tr>
<tr>
<td>Across Maturities</td>
<td></td>
<td>6657</td>
<td>4547</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>Average Prices</th>
<th>Moneyness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; 30$</td>
<td></td>
<td>2.667</td>
</tr>
<tr>
<td>$30 \leq T &lt; 80$</td>
<td>6.874</td>
<td>14.302</td>
</tr>
<tr>
<td>$80 \leq T &lt; 180$</td>
<td>21.699</td>
<td>38.042</td>
</tr>
<tr>
<td>$180 \leq T \leq 250$</td>
<td>39.526</td>
<td>63.879</td>
</tr>
<tr>
<td>Across Maturities</td>
<td></td>
<td>17.691</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>Average Implied Volatilities</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; 30$</td>
<td></td>
<td>0.167</td>
</tr>
<tr>
<td>$30 \leq T &lt; 80$</td>
<td></td>
<td>0.159</td>
</tr>
<tr>
<td>$80 \leq T &lt; 180$</td>
<td></td>
<td>0.161</td>
</tr>
<tr>
<td>$180 \leq T \leq 250$</td>
<td></td>
<td>0.166</td>
</tr>
<tr>
<td>Across Maturities</td>
<td></td>
<td>0.163</td>
</tr>
</tbody>
</table>

Table 1: Basic features of the Sample A option dataset (Wednesdays). Prices in this dataset correspond to the period January 1st, 2004–December 31st, 2013.
### Table 2: Basic features of the Sample B option dataset (Thursdays). Prices in this dataset correspond to the period January 1st, 2004–December 31st, 2013.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>$[0.900, 0.950]$</th>
<th>$[0.950, 0.975]$</th>
<th>$[0.975, 1.000]$</th>
<th>$[1.000, 1.025]$</th>
<th>$[1.025, 1.050]$</th>
<th>$[1.050, 1.100]$</th>
<th>Moneyness</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; 30$</td>
<td>1533</td>
<td>1394</td>
<td>1412</td>
<td>879</td>
<td>286</td>
<td>165</td>
<td>5669</td>
</tr>
<tr>
<td>$30 \leq T &lt; 80$</td>
<td>3176</td>
<td>2368</td>
<td>2551</td>
<td>1772</td>
<td>377</td>
<td>256</td>
<td>10500</td>
</tr>
<tr>
<td>$80 \leq T &lt; 180$</td>
<td>1288</td>
<td>690</td>
<td>764</td>
<td>599</td>
<td>175</td>
<td>127</td>
<td>3643</td>
</tr>
<tr>
<td>$180 \leq T \leq 250$</td>
<td>471</td>
<td>213</td>
<td>320</td>
<td>291</td>
<td>78</td>
<td>43</td>
<td>1416</td>
</tr>
<tr>
<td>Across Maturities</td>
<td>6468</td>
<td>4665</td>
<td>5047</td>
<td>3541</td>
<td>916</td>
<td>591</td>
<td>21228</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>Average Prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; 30$</td>
<td>3.011</td>
</tr>
<tr>
<td>$30 \leq T &lt; 80$</td>
<td>6.961</td>
</tr>
<tr>
<td>$80 \leq T &lt; 180$</td>
<td>21.800</td>
</tr>
<tr>
<td>$180 \leq T \leq 250$</td>
<td>39.623</td>
</tr>
<tr>
<td>Across Maturities</td>
<td>17.848</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>Average Implied Volatilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; 30$</td>
<td>0.172</td>
</tr>
<tr>
<td>$30 \leq T &lt; 80$</td>
<td>0.159</td>
</tr>
<tr>
<td>$80 \leq T &lt; 180$</td>
<td>0.163</td>
</tr>
<tr>
<td>$180 \leq T \leq 250$</td>
<td>0.167</td>
</tr>
<tr>
<td>Across Maturities</td>
<td>0.165</td>
</tr>
</tbody>
</table>
Maximum Likelihood estimations on returns

<table>
<thead>
<tr>
<th>Innovations Type</th>
<th>Gaussian Innovations</th>
<th>Gaussian Mixture Innovations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NGARCH Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>$1.467 \cdot 10^{-6}$</td>
<td>$1.204 \cdot 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>$(4.882 \cdot 10^{-9})$</td>
<td>$(4.362 \cdot 10^{-9})$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$0.0523$</td>
<td>$0.0466$</td>
</tr>
<tr>
<td></td>
<td>$(6.145 \cdot 10^{-5})$</td>
<td>$(1.484 \cdot 10^{-4})$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$0.8279$</td>
<td>$0.8317$</td>
</tr>
<tr>
<td></td>
<td>$(3.596 \cdot 10^{-4})$</td>
<td>$(3.674 \cdot 10^{-4})$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$1.4599$</td>
<td>$1.5814$</td>
</tr>
<tr>
<td></td>
<td>$(2.468 \cdot 10^{-3})$</td>
<td>$(4.209 \cdot 10^{-3})$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$0.0209$</td>
<td>$0.0035$</td>
</tr>
<tr>
<td></td>
<td>$(3.378 \cdot 10^{-4})$</td>
<td>$(1.606 \cdot 10^{-4})$</td>
</tr>
<tr>
<td><strong>Distribution Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>$0.9222, 0.0778$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(1.228 \cdot 10^{-3})$</td>
<td></td>
</tr>
<tr>
<td>$h$</td>
<td>$0.8486, 1.8099$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(1.019 \cdot 10^{-3})$</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>$(0.0804, -0.9534)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(7.199 \cdot 10^{-4})$</td>
<td></td>
</tr>
<tr>
<td>Skewness</td>
<td>0</td>
<td>$-0.2809$</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3</td>
<td>3.6198</td>
</tr>
<tr>
<td>Normalized logL</td>
<td>3.1276</td>
<td>3.1376</td>
</tr>
</tbody>
</table>

Table 3: Joint likelihood estimation of the Gaussian and Gaussian mixture NGARCH models using 2,520 daily log-returns of the S&P500 index over the period December 22nd, 1999–December 30th, 2009. The standard errors for the distribution parameters correspond to the estimation errors of the first components of $p$, $h$, and $m$ since the values of $p_2$, $m_2$, $h_2$ are determined by the relations $p_1 + p_2 = 1$ and (3.26). The implied skewness and kurtosis of the GARCH innovations are also reported.
Non-affine GARCH option pricing models, variance dependent kernels, and diffusion limits

<table>
<thead>
<tr>
<th>Innovations Type</th>
<th>Gaussian Innovations</th>
<th>Gaussian Mixture Innovations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pricing Kernel</td>
<td>( \theta_t^{(2)} = 0 )</td>
<td>( \theta_t^{(2)} = \theta^{(2)}/\sigma_t^2 )</td>
</tr>
<tr>
<td>GARCH Parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \alpha_0 )</td>
<td>1.209 \cdot 10^{-4}</td>
<td>6.426 \cdot 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>(1.146 \cdot 10^{-8})</td>
<td>(6.342 \cdot 10^{-9})</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.3264</td>
<td>0.3869</td>
</tr>
<tr>
<td></td>
<td>(3.697 \cdot 10^{-4})</td>
<td>(5.839 \cdot 10^{-5})</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.3607</td>
<td>0.2000</td>
</tr>
<tr>
<td></td>
<td>(3.667 \cdot 10^{-4})</td>
<td>(8.481 \cdot 10^{-5})</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.4439</td>
<td>0.6108</td>
</tr>
<tr>
<td></td>
<td>(8.162 \cdot 10^{-4})</td>
<td>(1.022 \cdot 10^{-4})</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>0.1238</td>
<td>0.1606</td>
</tr>
<tr>
<td></td>
<td>(2.303 \cdot 10^{-5})</td>
<td>(2.490 \cdot 10^{-6})</td>
</tr>
<tr>
<td>Distribution Parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p )</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( h )</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( m )</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Skewness</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Variance Premium Parameter (( \theta_2 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>-0.328</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(4.028 \cdot 10^{-5})</td>
</tr>
<tr>
<td>Normalized logL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (logL(\hat{L}))</td>
<td>2.0982</td>
<td>2.2802</td>
</tr>
<tr>
<td>From Returns (logL(\hat{L}_R))</td>
<td>1.4400</td>
<td>1.4774</td>
</tr>
<tr>
<td>From Options (logL(\hat{L}_O))</td>
<td>0.6582</td>
<td>0.8028</td>
</tr>
</tbody>
</table>

Table 4: Joint likelihood estimation of the model and pricing kernel parameters using the 2009 options dataset. The standard errors for the distribution parameters correspond to the estimation errors of the first components of \( p \), \( h \), and \( m \) since the values of \( p_2, m_2, h_2 \) are determined by the relations \( p_1 + p_2 = 1 \) and \( (3.26) \). The pricing kernels are the conditional Esscher transform obtained from \((3.1)\) by taking \( \theta_t^{(2)} = 0 \) (no variance premium) and a variance dependent SDF obtained from \((3.1)\) by taking \( \theta_t^{(2)} = \theta^{(2)}/\sigma_t^2 \).
Figure 1: Dependence of the joint log-likelihood on the variance premium parameter. All the model and distribution parameters have been kept fixed and set to the optimal values reported in Table 4 and it is only the variance premium parameter $\theta^{(2)}$ that varies. The top (respectively, bottom) panel refers to the case with Gaussian innovations (respectively, Gaussian mixture).
### Table 5: Results of the out of sample pricing exercise.

The module marked “In-Sample Error” reports the average IVRMSE committed each Wednesday at the time of pricing the options that have been used to optimize the options likelihood with respect to $\theta^{(2)}$ the very same Wednesday. The module marked “Next Day Pricing Error” reports the average IVRMSE committed each Thursday using the models whose parameters have been estimated the preceding day. The module marked “Next Week Pricing Error” reports the average IVRMSE committed each Wednesday using the models whose parameters have been estimated the Wednesday of the preceding week.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>In-Sample Error</th>
<th>Next Day Pricing Error</th>
<th>Next Week Pricing Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T &lt; 30$</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

The table shows the implied volatility root mean squared error (IVRMSE) with sequentially estimated parameters for different maturity periods, moneyness, and model types. The results are categorized by maturities ($T < 30$, $30 < T < 80$, $80 < T < 180$, $180 < T < 250$), and across maturities. Each entry represents the average IVRMSE committed at different stages of the pricing exercise.
Table 6: Classification table for the Hull-White model

<table>
<thead>
<tr>
<th>Case</th>
<th>( \tilde{v}(\ell) )</th>
<th>( \tilde{v}(r) )</th>
<th>( \tilde{v}_b(\ell) )</th>
<th>( \tilde{v}_b(r) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) ( a' &gt; \frac{\zeta^2}{2} )</td>
<td>( \varpi \leq 0 )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td></td>
<td>( \varpi &gt; 0 )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>(II) ( a' = \frac{\zeta^2}{2} )</td>
<td>( \varpi \leq 0 )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td></td>
<td>( \varpi &gt; 0 )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>(III) ( a' &lt; \frac{\zeta^2}{2} )</td>
<td>( \varpi \leq 0 )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
</tr>
<tr>
<td></td>
<td>( \varpi &gt; 0 )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
<td>( \infty )</td>
</tr>
</tbody>
</table>

Table 7: Classification table for the model (4.6)

<table>
<thead>
<tr>
<th>Case</th>
<th>( \tilde{v}(\ell) )</th>
<th>( \tilde{v}(r) )</th>
<th>( \tilde{v}_b(\ell) )</th>
<th>( \tilde{v}_b(r) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I) ( a' &gt; \frac{\zeta^2}{2} )</td>
<td>( \varpi \leq 0 )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td></td>
<td>( \varpi &gt; 0 )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>(II) ( a' = \frac{\zeta^2}{2} )</td>
<td>( \varpi \leq 0 )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td></td>
<td>( \varpi &gt; 0 )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
<td>( \infty )</td>
</tr>
<tr>
<td>(III) ( a' &lt; \frac{\zeta^2}{2} )</td>
<td>( \varpi \leq 0 )</td>
<td>( \infty )</td>
<td>( \infty )</td>
<td>undetermined</td>
</tr>
<tr>
<td></td>
<td>( \varpi &gt; 0 )</td>
<td>( \infty )</td>
<td>( &lt; \infty )</td>
<td>undetermined</td>
</tr>
</tbody>
</table>