Option Valuation with Volatility Components,
Fat Tails, and Non-Monotonic Pricing Kernels*

Kadir Babaoğlu  Peter Christoffersen
Rotman School of Management  University of Toronto,
University of Toronto  CBS and CREATEES
Steven Heston  Kris Jacobs
Smith School of Business  Bauer College of Business
University of Maryland  University of Houston
September 27, 2016

Abstract

We nest multiple volatility components, fat tails and a U-shaped pricing kernel in a single option model and compare their contribution to describing returns and option data. All three features lead to statistically significant model improvements. A U-shaped pricing kernel is economically most important and improves option fit by 17% on average and more so for two-factor models. A second volatility component improves the option fit by 9% on average. Fat tails improve option fit by just over 4% on average, but more so when a U-shaped pricing kernel is applied. Overall these three model features are complements rather than substitutes: the importance of one feature increases in conjunction with the others.

JEL Classification: G12

Keywords: Volatility components; fat tails; jumps; pricing kernel.

*We are grateful to our EFA discussant Fulvio Pegoraro, Alex Potapchik and Lars Stentoft for very helpful remarks. Christoffersen acknowledges financial assistance from Bank of Canada, GRI and SSHRC. Correspondence to: Steven Heston, R.H. Smith School of Business, University of Maryland, 4447 Van Munching Hall, College Park, MD 20742; Tel: (301) 405-9686; E-mail: sheston@rhsmith.umd.edu.
1 Introduction

By accounting for heteroskedasticity and volatility clustering, empirical studies on option valuation substantially improve on the Black-Scholes (1973) model prices through the parametric modeling of stochastic volatility (SV), see for example Heston (1993) and Bakshi, Cao, and Chen (1997). The literature has focused on two improvements to capture the stylized facts in the data. First, by accounting for more than one volatility component, the model becomes more flexible and its modeling of the term structure of volatility improves. This approach is advocated by Duffie, Pan, and Singleton (2000) and implemented on option prices by, among others, Bates (2000), Christoffersen, Heston, and Jacobs (2009), and Xu and Taylor (1994).\footnote{See for instance Chernov, Gallant, Ghysels, and Tauchen (2003) for a study of multiple volatility components in the underlying return series.} Christoffersen, Jacobs, Ornthanalai, and Wang (2008) propose a discrete-time GARCH option valuation model with two volatility components which has more structure, by modeling total volatility as evolving around a stochastic long-run mean.

The second modeling improvement that reliably improves model fit is to augment stochastic volatility with jumps in returns and/or volatility. A large number of studies have implemented this approach.\footnote{See for instance Andersen, Benzoni, and Lund (2002), Bakshi, Cao, and Chen (1997), Bates (1996a, 2000), Broadie, Chernov, and Johannes (2007), Chernov and Ghysels (2000), Eraker (2004), Jones (2003), and Pan (2002), for studies that estimate SV models with jumps using options and/or return data.} Intuitively, the advantage offered by jump processes is that they allow for conditional nonnormality, and therefore for instantaneous skewness and kurtosis. In discrete-time modeling, an equivalent approach uses innovations that are conditionally non-Gaussian. Examples of this approach are Christoffersen, Heston, and Jacobs (2006), who use Inverse Gaussian innovations, and Barone-Adesi, Engle, and Mancini (2008) who take a nonparametric approach.

The studies cited above demonstrate convincingly that these two modeling approaches improve model fit for both the option prices and the underlying returns. However, the most important challenge faced by these models is the simultaneous modeling of the underlying returns and the options. This position is forcefully articulated by for example Bates (1996b, 2003). Andersen, Fusari, and Todorov (2015) address this by fitting realized (physical) volatility together with option prices, but this still leaves open the question of a pricing kernel that links the observed “physical” measure and the risk-neutral measure inherent in option prices. In particular, deficiencies in a model’s ability to simultaneously describe returns and option prices may not exclusively be due to the specification of the driving process, but could also be caused by a misspecified price of risk, or equivalently the pricing kernel.

The literature focuses on pricing kernels that depend on wealth, originating in the seminal work of Brennan (1979) and Rubinstein (1976). Liu, Pan, and Wang (2004) discuss the specifi-
cation of the price of risk when SV models are augmented with Poisson jumps. Several papers, including Ait-Sahalia and Lo (1998), Jackwerth (2000), Rosenberg and Engle (2002), Bakshi, Madan, and Panayotov (2010), Brown and Jackwerth (2012), and Chabi-Yo (2012), have documented deviations from and explored extensions to the traditional log-linear pricing kernel. In recent work, Christoffersen, Heston, and Jacobs (2013) specify a more general pricing kernel that depends on volatility as well as wealth. The kernel is non-monotonic after projecting onto wealth, which is consistent with recent evidence by Cuesdeanu and Jackwerth (2015).

Christoffersen, Heston, and Jacobs (2013) show that the more general pricing kernel provides a superior fit to option prices and returns.

The literature thus suggests at least three important improvements on the benchmark SV option pricing model. First, multiple volatility components; second, conditional nonnormality or jumps; and third, non-monotonic pricing kernels. It is important to nest these features within a common framework in order to have a “horserace” comparison of their importance. In addition, examining these features jointly shows how they interact in describing returns and options. Ideally these different model features ought to be complements rather than substitutes.

The second volatility factor should improve the modeling of the volatility term structure, and therefore the valuation of options of different maturities, and long-maturity options in particular. Non-Gaussian innovations should prove most useful to capture the moneyness dimension for short-maturity options, which is usually referred to as the smirk. The non-monotonic pricing kernel has an entirely different purpose, because its relevance lies in the joint modeling of index returns and options, rather than the modeling of options alone.

However, the existing literature does not contain any evidence on whether these model features are indeed complements when confronted with the data. The literature does also not address the question of which model feature is statistically and economically most significant. This paper is the first to address this issue by comparing the three features within a nested model. We conduct an extensive empirical evaluation of the three model features using returns data, using options data, and finally using a sequential estimation exercise. We find that all three model features lead to statistically significant model improvements. A U-shaped pricing kernel is economically most important and improves option fit by 17% on average and more so for two-factor models. A second volatility factor improves the option fit by 9% on average. Fat tails improve option fit by just over 4% on average, but more so when a U-shaped pricing kernel is applied. Our results suggest that the three features are complements rather than substitutes.

The paper proceeds as follows. Section 2 introduces a class of GARCH dynamics for option

\[3\] Linn, Shive, and Shumway (2014) argue that the finding of a nonmonotone pricing kernel could be an artifact of the econometric method used. Cuesdeanu and Jackwerth (2015) show that the finding of a nonmonotone kernel is robust across a range of econometric techniques.
valuation. The most general return dynamic we consider has non-normal innovations and two
variance components, one of which is a stochastic long-run mean. We also derive the Gaussian
limit of this return process. Section 3 discusses the risk-neutralization of this process. Section 4
discusses data and estimation, and Section 5 presents the empirical results. Section 6 concludes.

2 A Class of GARCH Dynamics for Option Valuation

This section introduces a general class of GARCH dynamics for index option valuation. The
most general model we consider is a two-factor fat-tail GARCH model with dynamics that may
appear nonstandard. We therefore first introduce two better-known models, and subsequently
introduce the more general IG-GARCH(2,2) model. We also show how the IG-GARCH(2,2)
model can be transformed into a component model and demonstrate how it nests the simpler
cases.

One can use observable state variables to value options in any dynamic model. For example,
one might use implied volatilities extracted from option prices. Alternatively, one might use a
filtering technique such as the particle filter, or rely on realized volatility computed from intraday
returns as in Andersen, Fusari, and Todorov (2015). We choose a GARCH model because we
want to assess whether option prices are consistent with observed returns. In this framework
filtering is straightforward, which facilitates investigating the relationship between option prices
and return dynamics. The GARCH approach allows us to impose economic restrictions based on
observed returns, without an auxiliary filter that is separate from the assumptions of the option
model. The limitation of a GARCH approach is that it does not allow one-step-ahead volatility
to evolve independently of returns. This is not a significant problem in practice, because the
model allows innovations in variance to be imperfectly correlated with daily (or higher frequency)
returns.

2.1 The GARCH(1,1) Model

The first model we consider is the Heston-Nandi (2000) Gaussian GARCH(1,1) process:

\[
\ln(S(t+\Delta)) = \ln(S(t)) + r + \tilde{\mu}h(t+\Delta) + \sqrt{h(t+\Delta)}z(t+\Delta),
\]
\[
h(t+\Delta) = \omega + \beta h(t) + \alpha(z(t) - \gamma \sqrt{h(t)})^2,
\]

where \(z(t)\) has a standard normal distribution. This model allows for quasi-closed form val-
uation of European options, and has therefore been estimated and tested in several empirical
applications.\footnote{See for example, Hsieh and Ritchken (2005), Barone-Adesi, Engle and Mancini (2008), and Christoffersen, Jacobs, Ornthanalai and Wang (2008).}

\section{The GARCH(2,2) Model and a Component Model}

A straightforward generalization of the Heston-Nandi (2000) GARCH(1,1) dynamic in (1a)-(1b) is the following GARCH(2,2) process with normal innovations:

\begin{align}
\ln(S(t+\Delta)) &= \ln(S(t)) + r + \tilde{\mu}h(t+\Delta) + \sqrt{h(t+\Delta)}z(t+\Delta), \\
h(t+\Delta) &= \omega + \beta_1 h(t) + \beta_2 h(t-\Delta) \\
&\quad + \alpha_1(z(t) - \gamma_1\sqrt{h(t)})^2 + \alpha_2(z(t-\Delta) - \gamma_2\sqrt{h(t-\Delta)})^2,
\end{align}

(2a)

The GARCH(2,2) model is not typically used in empirical work. However, building on Engle and Lee (1999), by imposing some parameter restrictions it can be written as the component model of Christoffersen et al. (2008):

\begin{align}
h(t+\Delta) &= q(t+\Delta) + \rho_1(h(t) - q(t)) + \nu_h(t), \\
q(t+\Delta) &= \omega_q + \rho_2 q(t) + \nu_q(t).
\end{align}

(3a)

where

\begin{align*}
\nu_i(t) &= \alpha_i[(z(t) - \gamma_i\sqrt{h(t)})^2 - 1 - \gamma_i^2 h(t)] \quad i = h, q, \\
\gamma_h &= -\rho_1\frac{\alpha_1\gamma_1 + \alpha_2\gamma_2}{(\rho_2 - \rho_1)\alpha_h}, \\
\gamma_q &= \frac{\rho_2\alpha_1\gamma_1 + \alpha_2\gamma_2}{(\rho_2 - \rho_1)\alpha_q}, \\
\alpha_h &= -\frac{\rho_1}{\rho_2 - \rho_1}\alpha_1 - \frac{1}{\rho_2 - \rho_1}\alpha_2, \\
\alpha_q &= \frac{\rho_2}{\rho_2 - \rho_1}\alpha_1 + \frac{1}{\rho_2 - \rho_1}\alpha_2,
\end{align*}

and $\rho_1$ and $\rho_2$ are the respective smaller and larger roots of the quadratic equation

$$\rho^2 - (\beta_1 + \alpha_1\gamma_1^2)\rho - \beta_2 - \alpha_2\gamma_2^2 = 0,$$

The dynamic for the long-run component can equivalently be expressed as

$$q(t+\Delta) = \sigma^2 + \rho_2(q(t) - \sigma^2) + \nu_q(t),$$

where $\sigma^2$ is the unconditional variance.

The component parameters can also be inverted to recover the GARCH(2,2) parameters

\begin{align*}
\alpha_1 &= \alpha_h + \alpha_q, \\
\alpha_2 &= -\rho_2\alpha_h - \rho_1\alpha_q, \\
\gamma_1 &= \alpha_h\gamma_h + \alpha_q\gamma_q \quad \gamma_2 = -\frac{\rho_2\alpha_1\gamma_1 + \alpha_2\gamma_2}{\alpha_1}, \\
\beta_1 &= \rho_1 + \rho_2 - \alpha_1\gamma_1^2, \\
\beta_2 &= -\rho_1\rho_2 - \alpha_2\gamma_2^2.
\end{align*}
The component structure helps interpreting the model. The coefficients of the lagged variables (in the long- and short-run components) are the roots of the process’ characteristic equation. These parameters are more informative about the process than the parameters in the GARCH(2,2) model, which facilitates estimation including the identification of appropriate parameter starting values.

2.3 The IG-GARCH(1,1) Model

Another generalization of the Heston-Nandi (2000) GARCH(1,1) dynamic in (1a)-(1b) is the IG-GARCH(1,1) process in Christoffersen, Heston, and Jacobs (2006), given by:

\[
\begin{align*}
\ln(S(t + \Delta)) &= \ln(S(t)) + r + \mu h(t + \Delta) + \eta y(t + \Delta), \\
h(t + \Delta) &= w + b_1 h(t) + c_1 y(t) + a_1 h(t)^2/y(t),
\end{align*}
\]

where \(y(t + \Delta)\) has an Inverse Gaussian distribution with degrees of freedom \(h(t + \Delta)/\eta^2\). Note that while \(y(t + \Delta)\) is a positive random variable, returns are shifted by \(\mu h(t + \Delta)\) and can have both negative and positive values.

The Inverse Gaussian innovation and its reciprocal have the following conditional means

\[
\begin{align*}
E_t[y(t + \Delta)] &= h(t + \Delta)/\eta^2, \\
E_t[1/y(t + \Delta)] &= \eta^2/h(t + \Delta) + \eta^4/h(t + \Delta)^2.
\end{align*}
\]

The dynamic (4a)-(4b) can be written in terms of zero-mean innovations as follows

\[
\begin{align*}
\ln(S(t + \Delta)) &= \ln(S(t)) + r + \tilde{\mu} h(t + \Delta) + \sqrt{h(t + \Delta)} z(t + \Delta), \\
h(t + \Delta) &= \tilde{w} + \tilde{b}_1 h(t) + v_1(t),
\end{align*}
\]

where

\[
\begin{align*}
\tilde{\mu} &= \mu + \eta^{-1}, \\
\tilde{w} &= w + a_1 \eta^4, \\
\tilde{b}_1 &= b_1 + c_1/\eta^2 + a_1 \eta^2, \\
z(t) &= \frac{\eta y(t) - h(t)/\eta}{\sqrt{h(t)}}, \\
v_1(t) &= c_3 y(t) + a_1 h(t)^2/y(t) - c_1 h(t)/\eta^2 - a_1 \eta^2 h(t) - a_1 \eta^4.
\end{align*}
\]
The conditional means of return and variance are given by

$$E_t[\ln(S(t + \Delta) / S(t))] = r + \tilde{\mu}h(t + \Delta),$$  \hspace{1cm} (8a)
$$E_t[h(t + 2\Delta)] = \tilde{\nu} + \tilde{b}_1h(t + \Delta).$$  \hspace{1cm} (8b)

The advantage of the IG-GARCH(1,1) process in (4a)-(4b) over the GARCH(1,1) process in (1a)-(1b) is that the innovation is non-normal, thus allowing for conditional skewness and kurtosis. Because the Inverse Gaussian distribution converges to the normal distribution, the Heston-Nandi (2000) dynamic is nested by the specification of Christoffersen, Heston, and Jacobs (2006).

### 2.4 The IG-GARCH Component Model

We now combine the two generalizations of the Heston-Nandi (2000) model, the Inverse Gaussian innovations and the component structure. Consider the IG-GARCH(2,2) process given by:

$$\ln(S(t + \Delta)) = \ln(S(t)) + r + \mu h(t + \Delta) + \eta y(t + \Delta),$$  \hspace{1cm} (9a)
$$h(t + \Delta) = w + b_1h(t) + b_2h(t - \Delta) + c_1y(t) + c_2y(t - \Delta)$$
$$+ a_1h(t)^2/y(t) + a_2h(t - \Delta)^2/y(t - \Delta),$$  \hspace{1cm} (9b)

The dynamic (9a)-(9b) can be written in terms of zero-mean innovations as follows

$$\ln(S(t + \Delta)) = \ln(S(t)) + r + \tilde{\mu}h(t + \Delta) + \sqrt{h(t + \Delta)}z(t + \Delta),$$  \hspace{1cm} (10a)
$$h(t + \Delta) = \tilde{\nu} + \tilde{b}_1h(t) + \tilde{b}_2h(t - \Delta) + v_1(t) + v_2(t - \Delta),$$  \hspace{1cm} (10b)

where

$$\tilde{\mu} = \mu + \eta^{-1},$$  \hspace{1cm} (11a)
$$\tilde{\nu} = w + a_1\eta^4 + a_2\eta^4,$$  \hspace{1cm} (11b)
$$\tilde{b}_i = b_i + c_i/\eta^2 + a_i\eta^2,$$  \hspace{1cm} (11c)
$$z(t) = \frac{\eta y(t) - h(t)/\eta}{\sqrt{h(t)}}$$  \hspace{1cm} (11d)
$$v_i(t) = c_iy(t) + a_ih(t)^2/y(t) - c_ih(t)/\eta^2 - a_i\eta^2h(t) - a_i\eta^4.$$  \hspace{1cm} (11e)

Note that by incorporating the lagged return innovation $y(t - \Delta)$ and the reciprocals $1/y(t)$ and $1/y(t - \Delta)$, the change in variance is imperfectly correlated with the return.
The conditional means of return and variance are given by

\[
E_t[\ln(S(t + \Delta) / S(t))] = r + \tilde{\mu}h(t + \Delta),
\]
\[
E_t[h(t + 2\Delta)] = \tilde{\mu}h(t + \Delta) + c_2y(t) + a_2h(t)^2 / y(t).
\]

We now transform the IG-GARCH(2,2) into a component model that nests Christoffersen, Jacobs, Ornthanalai and Wang (2008). Define the long-run component \(q(t)\) of the variance process (10b) as

\[
q(t) = \frac{-\rho_1 \tilde{\mu}}{(1 - \rho_1)(\rho_2 - \rho_1)} + \frac{\rho_2}{\rho_2 - \rho_1}h(t) + \frac{\tilde{\mu}}{\rho_2 - \rho_1}h(t - \Delta) + \frac{1}{\rho_2 - \rho_1}v_2(t - \Delta),
\]

where \(v_2(t)\) is given by (11e), and where \(\rho_1\) and \(\rho_2\) are the respective smaller and larger roots of the quadratic equation

\[
\rho^2 - \tilde{\mu}_1\rho - \tilde{\mu}_2 = 0,
\]

which are the eigenvalues of the transition equation (9b). The short-run component is the deviation of variance from its long-run mean, \(h(t) - q(t)\). Substituting these into the IG-GARCH(2,2) dynamics (9a)-(9b) yields the IG-GARCH component model which we will denote IG-GARCH(C) below

\[
\ln(S(t + \Delta)) = \ln(S(t)) + r + \mu h(t + \Delta) + \eta y(t + \Delta),
\]
\[
h(t + \Delta) = q(t + \Delta) + \rho_1(h(t) - q(t)) + v_h(t),
\]
\[
q(t + \Delta) = w_q + \rho_2q(t) + v_q(t),
\]

or equivalently,

\[
q(t + \Delta) = \sigma^2 + \rho_2(q(t) - \sigma^2) + v_q(t),
\]

where \(\sigma^2\) is the unconditional variance, and

\[
\sigma^2 = \frac{\tilde{\mu}}{(1 - \rho_1)(1 - \rho_2)}, \quad w_q = \frac{\tilde{\mu}}{1 - \rho_1},
\]
\[
a_h = -\frac{\rho_1}{\rho_2 - \rho_1}a_1 - \frac{1}{\rho_2 - \rho_1}a_2 \quad a_q = \frac{\rho_2}{\rho_2 - \rho_1}a_1 + \frac{1}{\rho_2 - \rho_1}a_2
\]
\[
c_h = -\frac{\rho_1}{\rho_2 - \rho_1}c_1 - \frac{1}{\rho_2 - \rho_1}c_2 \quad c_q = \frac{\rho_2}{\rho_2 - \rho_1}c_1 + \frac{1}{\rho_2 - \rho_1}c_2
\]
\[
v_i(t) = c_i y(t) + a_i h(t)^2 / y(t) - c_i h(t) / \eta^2 - a_i \eta^2 h(t) - a_i \eta^4.
\]

The unit root condition, \(\rho_2 = 1\), corresponds to the restriction \(\tilde{\mu}_2 = 1 - \tilde{\mu}_1\). The expression for \(\sigma^2\) shows that total variance persistence in the component model is simply

\[
1 - (1 - \rho_1)(1 - \rho_2) = \rho_2 + \rho_1(1 - \rho_2).
\]
The component parameters can also be inverted to get the IG-GARCH(2,2) parameters

\[ a_1 = a_h + a_q \]
\[ a_2 = -\rho_2 a_h - \rho_1 a_q \]
\[ b_1 = \rho_1 + \rho_2 \]
\[ b_2 = -\rho_1 \rho_2 \]
\[ c_1 = c_h + c_q \]
\[ c_2 = -\rho_2 c_h - \rho_1 c_q \]

This proves that the IG-GARCH(2,2) model is equivalent to the component model (14a)-(14c). In the IG-GARCH(1,1) special case studied in Christoffersen, Heston and Jacobs (2006), the long-run component in (14c) is effectively removed from the return dynamics.

### 2.5 The Gaussian Limit of the IG Model

We now show formally how the Gaussian models are nested by the Inverse Gaussian models. Consider the normalization of the innovation to the return process in (9a),

\[
z(t) = \frac{\eta y(t) - h(t)/\eta}{\sqrt{h(t)}}.
\]

This normalized Inverse Gaussian innovation converges to a Gaussian distribution as the degrees of freedom, \( h(t)/\eta^2 \), approach infinity. If we fix \( z(t) \) and \( h(t) \), and take the limit as \( \eta \) approaches zero, then the IG-GARCH(2,2) process (10a)-(10b) converges weakly to the Heston-Nandi (2000) GARCH(2,2) process in (2a):

\[
\ln(S(t + \Delta)) = \ln(S(t)) + r + \tilde{\mu} h(t + \Delta) + \sqrt{h(t + \Delta)} z(t + \Delta),
\]
\[
h(t + \Delta) = \omega + \beta_1 h(t) + \beta_2 h(t - \Delta)
\]
\[
+ \alpha_1 (z(t) - \gamma_1 \sqrt{h(t)})^2 + \alpha_2 (z(t - \Delta) - \gamma_2 \sqrt{h(t - \Delta)})^2,
\]

where the limit is taken as follows

\[
\tilde{\mu} = \omega - \alpha_1 - \alpha_2,
\]
\[
a_i = \alpha_i / \eta^2,
\]
\[
b_i = \beta_i + \alpha_i \gamma_i^2 + 2 \alpha_i \gamma_i / \eta - 2 \alpha_i / \eta^2,
\]
\[
c_i = \alpha_i (1 - 2 \eta \gamma_i).
\]

Written in component form, the limit is given by (3a)-(3b).

Our Inverse Gaussian Component model in (14a)-(14c) thus corresponds in the limit to the component model of Christoffersen, Jacobs, Ornthanalai, and Wang (2008). Christoffersen, Heston, and Jacobs (2006) show that the Inverse Gaussian GARCH(1,1) model in (4a)-(4b) nests
the Heston-Nandi (2000) Gaussian GARCH(1,1) model in (1a)-(1b).

3 The Risk-Neutral Model and Option Valuation

To value options, we introduce the pricing kernel and the resulting risk-neutral dynamics. We then elaborate on the relationships between the risk-neutral and physical parameters. We first discuss the risk-neutralization for the most general IG-GARCH(2,2) process. Subsequently we discuss special cases nested by the most general specification.

3.1 Risk-Neutralization

For the purpose of option valuation we need to derive the risk-neutral dynamics from the physical dynamics and pricing kernel. Risk-neutralization is more complicated for the Inverse Gaussian distribution than for the Gaussian distribution. We implement a volatility-dependent pricing kernel following Christofersten, Heston, and Jacobs (2013), where

$$M(t + \Delta) = M(t) \left( \frac{S(t + \Delta)}{S(t)} \right)^\theta \exp(\delta_0 + \delta_1 h(t + \Delta) + \xi h(t + 2\Delta)).$$

Recent evidence by Cuesdeanu and Jackwerth (2015) suggests that the pricing kernel may be a non-monotonic function of returns. Accordingly, Christofersten, Heston, and Jacobs (2013) show that in a GARCH framework, the log-kernel is a nonlinear and non-monotonic function of the path of spot returns. Henceforth we refer to it as the non-monotonic pricing kernel. If $\xi > 0$, the pricing kernel is U-shaped in returns. In Appendix A we show that the risk-free and the risky assets both satisfy the martingale restriction under the pricing kernel in equation (18).

In Appendix B we show that the scaled return innovation $s_y(t)$ is distributed Inverse Gaussian under the risk-neutral measure with variance $s_h(t)$, where

$$s_y = 1 - 2c_1\xi - 2\eta\phi,$$
$$s_h = \sqrt{1 - 2a_1\xi\eta^4s_y^{-3/2}}.$$  \(19\)

Inserting these definitions into the IG-GARCH(2,2) dynamics in (9) yields the risk-neutral

\[\text{We are grateful to our EFA discussant Fulvio Pegoraro for helping us clarify this derivation.}\]
\[
\ln(S(t + \Delta)) = \ln(S(t)) + r + \mu^* h^*(t + \Delta) + \eta^* y^*(t + \Delta),
\]
\[
h^*(t + \Delta) = w^* + b_1 h^*(t) + b_2 h^*(t - \Delta) + c_1^* y(t) + c_2^* y(t - \Delta) + a_1^* h^*(t)^2/y^*(t) + a_2^* h^*(t - \Delta)^2/y^*(t - \Delta),
\]

where

\[
h^*(t) = s_h h(t), \quad y^*(t) = s_y y(t),
\]
\[
\mu^* = \frac{\mu}{s_h}, \quad \eta^* = \frac{\eta}{s_y}, \quad w^* = s_h w,
\]
\[
a_1^* = \frac{s_y a_1}{s_h}, \quad c_1^* = \frac{s_h c_1}{s_y}.
\]

The risk-neutral return process is IG-GARCH because the innovation \(y^*(t + \Delta)\) has an Inverse Gaussian distribution under the risk-neutral measure. Notice that \(b_1\) and \(b_2\) are identical in the physical and risk-neutral processes. The risk-neutral process can also be written as a component model, the details are in Appendix C.

### 3.2 Preference Parameters and Risk-Neutral Parameters

Note that the risk-neutralization is specified for convenience in terms of the two reduced-form preference parameters \(s_h\) and \(s_y\). It is worth emphasizing that in fact only one extra parameter is required to convert physical to risk-neutral parameters. The martingale restriction for the risk-neutral dynamics is given by

\[
\mu^* = \frac{\sqrt{1 - 2\eta^2} - 1}{\eta^2}.
\]

This imposes the following restriction between the physical parameters \(\mu\) and the preference parameters \(\phi\) and \(\xi\)

\[
\mu = s_h \frac{\sqrt{1 - 2\eta/s_y} - 1}{\eta^2/s_y^2} = \sqrt{1 - 2a_1 \xi \eta^2} \frac{\sqrt{1 - 2c_1 \xi - 2\eta \phi - 2\eta} - \sqrt{1 - 2c_1 \xi - 2\eta \phi}}{\eta^2}.
\]

Given the physical parameters and the value of \(\xi\) (or \(s_y\), we can thus recover the value of the risk aversion parameter \(\phi\) (or \(s_h\)). In other words, it takes only one additional parameter to convert between physical and risk-neutral parameters. To see this, alternatively re-write these
restrictions as

\[ s_y = \frac{(\frac{1}{2} \mu^2 \eta^4 + (1 - 2a_1 \xi \eta^4) \eta^2)}{(1 - 2a_1 \xi \eta^4) \mu^2 \eta^4}, \]  

(24)

\[ s_h = \frac{\mu \eta^2}{s_y^2(\sqrt{1 - 2\eta/s_y} - 1)}. \]  

(25)

Because \( s_h \) is now a function only of \( s_y \) and physical parameters, this demonstrates that we can write (21a)-(21c) as a function of the physical parameters and one additional parameter, either the reduced form parameter \( s_y \) or the preference parameter \( \xi \).

3.3 Nested Option Models

The full risk-neutral valuation model has two components with inverse-Gaussian innovations. This model contains a number of simpler models as special cases. First consider the Gaussian limit of the risk-neutral dynamic. In the limit, as \( \eta \) approaches zero, \( \tilde{\mu}^* = \mu^* + \eta^{*-1} \) approaches \( -\frac{1}{2} \). Also in this limit, \( s_h = s_y^{-1} \) as seen from equation (19). The risk-neutral process therefore converges to

\[
\ln(S(t + \Delta)) = \ln(S(t)) + r - \frac{1}{2} h^*(t + \Delta) + \sqrt{h^*(t + \Delta)}z^*(t + \Delta),
\]

\[
h^*(t + \Delta) = \omega^* + \beta_1 h^*(t) + \beta_2 h^*(t - \Delta)
\]

\[+ \alpha_1^*(z(t) - \gamma_1^* \sqrt{h^*(t)})^2 + \alpha_2^*(z(t - \Delta) - \gamma_2^* \sqrt{h^*(t - \Delta)})^2, \]

where

\[
z^*(t + \Delta) = \frac{z(t + \Delta)}{\sqrt{s_h}} + (\frac{\tilde{\mu}}{\sqrt{s_h}} + \frac{\sqrt{s_h}}{2}) \sqrt{h^*(t + \Delta)},
\]

\[
\omega^* = s_h \omega, \quad \alpha_i^* = s_h^2 \alpha_i, \quad \gamma_i^* = \frac{\gamma_i + \tilde{\mu}}{s_h} + \frac{1}{2}.
\]

This is the GARCH(2,2) generalization of the risk-neutral version of the Gaussian GARCH(1,1) model studied in Christoffersen, Heston, and Jacobs (2013). Following our previous analysis in equation (13), one may alternatively express this as the risk-neutral Gaussian component model.

Further setting \( \xi = 0 \), or equivalently \( s_h = 1 \), we retrieve the GARCH(2,2) version of the Heston-Nandi (2000) model.

Finally, the risk-neutral versions of the GARCH(1,1) models are obtained straightforwardly by setting the appropriate parameters to zero, similar to the restrictions for the physical dynamics discussed in Section 2.
3.4 Option Valuation

Option valuation with this model is straightforward. Following Heston and Nandi (2000), the value of a European call option at time $t$ with strike price $X$ maturing at $T$ is equal to

$$
\text{Call}(S(t), h(t + \Delta), X, T) = S(t) \left( \frac{1}{2} + \frac{\exp^{-r(T-t)}}{\pi} \int_{0}^{\infty} \text{Re} \left[ \frac{X^{-i\varphi}g^*_i(i\varphi, 1, T)}{i\varphi S(t)} \right] d\varphi \right) - X \exp^{-r(T-t)} \left( \frac{1}{2} + \frac{1}{\pi} \int_{0}^{\infty} \text{Re} \left[ \frac{X^{-i\varphi}g^*_i(i\varphi, T)}{i\varphi} \right] d\varphi \right).$$

(28)

where $g^*_i(\varphi, T)$ is the conditional generating function for the risk-neutral process in (20). The conditional generating function $g_t(\varphi, T)$ under the physical measure is given by:

$$
g_t(\varphi, T) = E_t[S(T)^\varphi] = S(t)^\varphi \exp (A(t) + B(t)h(t + \Delta) + C(t)q(t + \Delta)),

(29)

where

$$
A(T) = B(T) = C(T) = 0,

(30)
$$

$$
A(t) = A(t + \Delta) + \varphi r + (w_q - a_h\eta^4 - a_q\eta^4)B(t + \Delta) + (w_q - a_q\eta^4)C(t + \Delta)

- \frac{1}{2} \ln(1 - 2(a_h + a_q)\eta^4B(t + \Delta) - 2a_q\eta^4C(t + \Delta)),

(31)
$$

$$
B(t) = \varphi \mu + (\rho_1 - (c_h + c_q)\eta^{-2} - (a_h + a_q)\eta^2)B(t + \Delta) - (c_q\eta^{-2} + a_q\eta^2)C(t + \Delta) + \eta^{-2} - \sqrt{(1 - 2(a_q + a_h)\eta^4B(t + \Delta) - 2a_q\eta^4C(t + \Delta))(1 - 2\eta\varphi - 2(c_q + c_h)B(t + \Delta) - 2c_qC(t + \Delta))}\eta^2

C(t) = (\rho_2 - \rho_1)B(t + \Delta) + \rho_2C(t + \Delta).

(32)
$$

This recursive definition requires computing equations (31-33) period-by-period with the terminal condition in (30) and then integrating $g_t(\varphi, T)$ as in (28). Note that in equations (31)-(33) risk-neutral parameters should be used when valuing options. Note also that we have supplied the conditional generating function for the IG-GARCH(C) model. The corresponding functions for the nested models can be obtained as special cases of this function using the results above. Put options can be valued using put-call parity.

Armed with the formulas for computing option values, we are now ready to embark on an empirical investigation of our model.
4 Data and Estimation

4.1 Data

Our empirical analysis uses out-of-the-money S&P500 call and put options for the January 10, 1996 through December 26, 2012 period with a maturity between 14 and 365 days. We apply the filters proposed by Bakshi, Cao, and Chen (1997) as well as other consistency checks. Rather than using a short time series of daily option data, we use an extended time period, but we select option contracts for one day per week only. This choice is motivated by two constraints. On the one hand, it is important to use as long a time period as possible, in order to be able to identify key aspects of the model including volatility persistence. See for instance Broadie, Chernov, and Johannes (2007) for a discussion. On the other hand, despite the numerical efficiency of our model, the optimization problems we conduct are very time-intensive, because we use very large panels of option contracts. Selecting one day per week over a long time period is therefore a useful compromise. We use Wednesday data, because it is the day of the week least likely to be a holiday. It is also less likely than other days such as Monday and Friday to be affected by day-of-the-week effects. Moreover, following the work of Dumas, Fleming and Whaley (1998) and Heston and Nandi (2000), several studies have used a long time series of Wednesday contracts. The first Wednesday available in the OptionMetrics database is January 10, 1996, and so our sample is January 10, 1996 through December 26, 2012.

Panel A in Table 1 presents descriptive statistics for the return sample. The return sample is constructed from the S&P500 index returns. The return sample dates from January 1, 1990 through December 31, 2012. The standard deviation of returns, at 18.61%, is substantially smaller than the average option-implied volatility, at 22.47%. The higher moments of the return sample are consistent with return data in most historical time periods, with a small negative skewness and substantial excess kurtosis. Table 1 also presents descriptive statistics for the return sample from January 10, 1996 through December 26, 2012, which matches the option sample. In comparison to the 1990-2012 sample, the standard deviation is somewhat higher, and average returns are somewhat lower. Average skewness and kurtosis in 1996-2012 are quite similar to the 1990-2012 sample.

Panels B and C of Table 1 present descriptive statistics for the option data by moneyness and maturity. Moneyness is defined as the implied futures price $F$ divided by strike price $X$. When $F/X$ is smaller than one, the contract is an out-of-the-money (OTM) call, and when $F/X$ is larger than one, the contract is an OTM put. The out-of-the-money put prices are converted into call prices using put-call parity. The sample includes a total of 29,022 option contracts with an average mid-price of 41.63 and average implied volatility of 22.47% as noted above. The implied volatility is largest for the OTM put options in Panel B, reflecting the well-known
volatility smirk in index options. The implied volatility term structure in Panel C is roughly flat on average during the sample period.

4.2 Estimation

We now present a detailed empirical investigation of the model outlined in Section 2. We can separately evaluate the model’s ability to describe return dynamics and to fit option prices. But the model’s ability to capture the differences between the physical and risk-neutral distributions requires fitting both return and option data using the same, internally consistent, set of parameters.

We first use an estimation exercise that fits options and returns separately. We also employ sequential estimation following Broadie, Chernov, and Johannes (2007), who first estimate each model on returns only and subsequently assess the fit of each model to option prices in a second step where only risk-premium parameters are estimated. This procedure is also used by Christoffersen, Heston, and Jacobs (2013) in the context of a Gaussian GARCH(1,1) model with a quadratic pricing kernel.

First consider returns. In the Inverse Gaussian case, the conditional density of the daily return is

\[
f(R(t)|h(t)) = \frac{h(t) \eta^{-3}}{\sqrt{2\pi(R(t) - r - \hat{\mu}h(t))^3\eta^{-3}}} \times \\
\exp \left( -\frac{1}{2} \left( \frac{R(t) - r - \hat{\mu}h(t)}{\eta} - \frac{h(t)}{\eta^2} \sqrt{\frac{\eta}{R(t) - r - \hat{\mu}h(t)}} \right)^2 \right).\]

The return log-likelihood is summed over all return dates.

\[
\ln L^R \propto \sum_{t=1}^{T} \{ \ln(f(R(t)|h(t))) \}. \tag{34}
\]

We can therefore obtain the physical parameters \( \Theta \) by estimating

\[
\Theta_{Return} = \arg \max_{\Theta} \ln L^R. \tag{35}
\]

Now consider the options data. Define the Black-Scholes Vega (BSV) weighted option valuation errors as

\[
\varepsilon_i = \left( \text{Call}_i^{Mkt} - \text{Call}_i^{Mod} \right) / \text{BSV}_i^{Mkt},
\]

where \( \text{Call}_i^{Mkt} \) represents the market price of the \( i^{th} \) option, \( \text{Call}_i^{Mod} \) represents the model price,
and $BSV_{i}^{Mix}$ represents the Black-Scholes vega of the option (the derivative with respect to volatility) at the market implied level of volatility. Assume these disturbances are i.i.d. normal so that the option log-likelihood is

$$\ln L^O \propto -\frac{1}{2} \sum_{i=1}^{N} \{\ln \left(s^2_{e} + \varepsilon^2_{i}/s^2_{e}\right)\}. \quad (36)$$

where we can concentrate out $s^2_{e}$ using the sample analogue $s^2_{e} = \frac{1}{N} \sum_{i=1}^{N} \varepsilon^2_{i}$. We use the term structure of interest rates from OptionMetrics when pricing options.

The vega-weighted option errors are very useful because it can be shown that they are an approximation to implied volatility based errors, which have desirable statistical properties. Unlike implied volatility errors, they do not require Black-Scholes inversion of model prices at every step in the optimization, which is very costly in large scale empirical estimation exercises such as ours.\(^6\) We obtain the risk-neutral parameters $\Theta^*$ based on options data by estimating

$$\Theta^*_{Opt} = \arg \max_{\Theta^*} \ln L^O. \quad (37)$$

Note that both estimation exercises mentioned above ignore the specification of the pricing kernel, and are therefore uninformative about the choice between the log-linear and non-monotonic pricing kernels. We thus conduct a third estimation exercise where we sequentially estimate the non-monotonic pricing kernel parameter, $\xi$, on options only, keeping all the physical parameters from (35) fixed. We thus estimate

$$\xi_{Seq} = \arg \max_{\xi} \ln L^O. \quad (38)$$

Sequential estimation is of course only conducted for the models with non-monotonic pricing kernels. Our sequential estimation approach follows that in Broadie, Chernov, and Johannes (2007) and Christoffersen, Heston, and Jacobs (2013).

## 5 Empirical Results

Because our specification nests several models, it allows for a comparison of the relative importance of model features. Specifically, we can compare the contribution of a second stochastic volatility factor, fat-tailed innovations, and a non-monotonic (or variance-dependent) pricing kernel. We can quantify the contribution of these features in separately explaining the time se-

\(^6\)See for instance Carr and Wu (2007) and Trolle and Schwartz (2009) for applications of $BSV_{i}^{Mix}$ weighted option errors.
ries of returns and the cross-section of option prices, as well as in explaining returns and options together, which we do in a sequential estimation exercise.

While a horserace based on model fit is of interest, it is also relevant to verify whether the different model features are complements rather than substitutes. In theory this should be the case: the second volatility factor should improve the modeling of the term structure of volatility, and therefore the valuation of options of different maturities, especially long-maturity options. In contrast, the fat-tailed IG innovation should prove most useful to capture the moneyness dimension for short-maturity out-of-the-money options, which is usually referred to as the smirk. The non-monotonic pricing kernel has an entirely different purpose, because its relevance lies in the joint modeling of index returns and options, rather than the modeling of options alone.

Tables 2-7 present the empirical results. Table 2 presents estimation results using returns data. The results include parameter estimates and log-likelihoods, as well as several implications of the parameter estimates such as moments and persistence. Table 3 presents similar results for the estimation based on option data, and Table 4 does the same for the sequential estimation based first on returns and subsequently on options. Table 4 also reports the improvement in fit for the non-monotonic pricing kernel over the linear pricing kernel in terms of log-likelihood values. Tables 5 and 6 provide more details on the models’ fit across moneyness and maturity categories for the three estimation exercises in Tables 2-4.

5.1 Fitting Returns and Fitting Options

We will organize our initial discussion around the measures of fit (i.e. log-likelihood values) for the different models contained in Table 2 (return fitting) and Table 3 (option fitting). We have results for the fit of six models in these tables. Of these six models, three have Gaussian innovations and three are characterized by fat-tailed Inverse Gaussian innovations. Two models have two variance factors, two have one factor. For comparison we also estimate two models that have no variance dynamics, which we refer to as homoskedastic models.

The most highly parameterized two-factor model with fat tails fits the returns and options data best, as can be seen in Tables 2 and 3, while the most restrictive single factor Gaussian model fits worst, which is not surprising in an in-sample exercise.

All the two-factor models have substantially higher likelihood values than all the one-factor models. The two-factor models have three more parameters than the corresponding one-factor models, and two times the difference in the log-likelihoods is asymptotically distributed chi-square with three degrees of freedom. The 99.9% p-level for this test is 16.3. In the case of the option-based estimation in Table 3, the improvement provided by the second factor is very large, with a likelihood improvement of approximately 5,000. This suggests that the most
important feature in accurately modelling option prices is the correct specification of the volatility
dynamics. This finding confirms the results in Andersen, Fusari, and Todorov (2015), where the
biggest improvement in fit also stems from a second factor.

The inclusion of a second factor also significantly improves the return fit in Table 2. For
example, for the Gaussian case, twice the difference in the log-likelihood between the two-factor
and one-factor models is 100, and for the fat-tailed case the corresponding number is 71. These
test statistics are highly significant. We conclude that a second factor is important in describing
the underlying returns as well as option prices.

When comparing IG versus Gaussian models, Tables 2 and 3 show that adding the single
parameter \( \eta \) in the IG models increases the return and option likelihoods substantially. In Table
3 the likelihood improvements are again in the thousands. The improvements in the return
likelihoods in Table 2 are less dramatic but still statistically significant at conventional confidence
levels.

Notice also that the magnitude of improvement contributed by the fat-tailed IG feature de-
pends on the other model features. For the option-based estimation in Table 3, the improvement
in option log-likelihood is 4,314 for the homoskedastic case, 1,853 for the one-factor GARCH,
and 1,575 for the two-factor GARCH. For the return-based estimation in Table 2, the improve-
ments in log-likelihood are 3.5 for the homoskedastic case, 48.1 for the one-factor GARCH, and
33.6 for the two-factor GARCH. The IG feature improves the fit more in the case of the simpler
one-factor model than in the case of the two-factor model. Non-normal innovations and a second
volatility component are therefore to some extent substitutes in model specification.

5.2 Sequential Estimation of the Non-Monotonic Pricing Kernel Pa-
rameter

Table 2 contains return-based estimates of the physical distributions. Table 3 contains option-
based estimates of the risk-neutral distribution. Neither table is informative about the pricing
kernel. In Table 4 we therefore use the physical parameter estimates from Table 2 and estimate
only the non-monotonic pricing kernel parameter \( \xi \) by fitting options. Table 4 reports risk-neutral
values of all parameters, but only \( \xi \) is estimated from options.

The penultimate column in Panel B of Table 4 reports the option likelihoods for the four
dynamic models with non-monotone pricing kernel. The last column in Panel B shows the
difference between the option likelihood for optimal \( \xi \) and that for \( \xi = 0 \), where the options are
valued using the risk-neutralized parameters from Table 2.

The increase in option log-likelihood when allowing for a non-monotonic pricing kernel and
adding just a single parameter is again in the thousands.
Table 4 shows that the log-likelihood increase due to the more general pricing kernel is 6,644 in the single factor Gaussian model, and 9,548 in the corresponding Inverse Gaussian model. In case of the two-factor models, the improvements are even higher: The non-monotonic kernel improves the two-factor likelihoods by 9,336 in the Gaussian model and 11,180 in the Inverse Gaussian model.

We conclude that the importance of modeling a more general pricing kernel depends on the models’ ability to capture the tails of the distribution. The richer dynamics of two-factor models allow them to better fit the fat tails, and a non-monotonic pricing kernel captures this property by allowing the model’s physical parameters to fit the returns and risk-neutral parameters to fit options in the same model. Complex modelling of risk premia complements adequate modelling of return dynamics.

Table 4 is also interesting in that it shows that the two key conclusions from Tables 2 and 3 still obtain: Allowing for inverse Gaussian innovations improves the fit, as does allowing for a second variance component. Note that in Table 4 these conclusions are based on option fit but use return-based estimates, which shows that these findings are not merely in-sample phenomena.

Figure 1 complements Table 4 by plotting the implied volatility RMSE percentages (top panel) and log-likelihood values (bottom panel) for different values of the $\xi$ parameter in the models we consider. Figure 1 shows that the IG-GARCH component model we propose has lower RMSE and higher log-likelihood values for the optimal $\xi$ parameter and indeed for a wide range of values around the optimum. The linear pricing kernel corresponds to the left-most point on the curves where $\xi = 0$.

5.3 Capturing Dynamics in Higher Moments

Examination of the parameter estimates in Tables 2-4 reveals the main reason for the superior performance of the two-factor models. For the returns-based estimation in Table 2, the persistence of the single factor estimates is 0.98 at a daily frequency for the Gaussian and the Inverse Gaussian model. For the two-factor models, the long-run factor is always very persistent ($\rho_2$ is around 0.99), but the persistence of the short-run factor, $\rho_1$, is 0.71 in the Gaussian model and 0.74 in the Inverse Gaussian model. The single-factor models are forced to compromise between slow and fast mean reversion, leading to a deterioration in fit in some parts of the sample.

Figures 2 and 3 provide additional perspective on the differences between the GARCH(1,1) and component models. Figure 2 plots the spot variance for all models using the return-based estimates. Figure 3 also uses the return-based estimates to plot conditional (“leverage”) correlation between returns and variance, $Corr_t[R(t + \Delta), h(t + 2\Delta)]$, which is informative about the third moment dynamics, and conditional standard deviation of variance, $\sqrt{Var_t[h(t + 2\Delta)]}$,
which is informative about the fourth moment dynamics. The formulas used for these conditional moments are contained in Appendix D.

In Figure 2, we can see that component model total variance (i.e. $h(t)$) is more variable and has the ability to increase faster than the GARCH(1,1), thanks to its short-run component (i.e. $h(t) - q(t)$). During the recent financial crisis the variances in the component models jump to a higher level than do the GARCH(1,1) variances. Consistent with this finding, the conditional standard deviation of variance (conditional correlation between returns and variance) of the component models in Figure 3, is higher in level (more negative) and more noisy than those of GARCH(1,1) models.

Figure 4 graphs the term structure of variance, skewness and kurtosis using the derivatives of the moment generating function. Variance, skewness and kurtosis are defined by

\[ \text{Var}_t(T) = \frac{\partial^2 \ln g_t(\varphi, T)}{\partial \varphi^2}|_{\varphi = 0}, \]  
(39)

\[ \text{Skew}_t(T) = \frac{\partial^3 \ln g_t(\varphi, T)}{\partial \varphi^3}|_{\varphi = 0} \left( \frac{\partial^2 \ln g_t(\varphi, T)}{\partial \varphi^2}|_{\varphi = 0} \right)^{3/2}, \]  
(40)

\[ \text{Kurt}_t(T) = \frac{\partial^4 \ln g_t(\varphi, T)}{\partial \varphi^4}|_{\varphi = 0} \left( \frac{\partial^2 \ln g_t(\varphi, T)}{\partial \varphi^2}|_{\varphi = 0} \right)^2 - 3. \]  
(41)

The plots in the first column of Figure 4 show variance normalized by unconditional variance of each model, the second column shows skewness and the third column shows kurtosis. Each row corresponds to a different model. The initial variance is set to twice the unconditional model variance in the solid lines and the initial variance is set to one-half the unconditional variance in the dashed lines. For the component models we set the long-run variance component, $q(t)$ equal to three-quarters of total variance, $h(t)$. We use the return-based parameters in Table 2 to plot Figure 4.

The left-side panels in Figure 4 highlight the differences between the GARCH(1,1) and component models. The impact of the current conditions on the future variance is much larger for the component models, and this is of course due to the persistence of the long-run component. For the GARCH(1,1) model, the conditional variance converges much quicker to the long-run variance.

Figure 4 also shows that the term structures of skewness and kurtosis in the models differ between one-factor and component models. The one-factor models generate strongly humped-shaped term structures whereas the component models do so to a much lesser degree.

Figure 4 confirms that the Gaussian and Inverse Gaussian models do not differ much in the term structure dimension, and also indicates that the effects of shocks last much longer in the component models.

Figures 5 and 6 repeat Figures 2 and 3 but uses the option-based parameters in Table 3 rather
than the physical parameters in Table 2. In Figure 5, the variance paths for the GARCH(1,1) and component models are very different compared to the return-implied paths in Figure 2. Note in particular that the short-run component in the component models strongly differs between Figures 2 and 5. In Figure 6, the time path of the conditional standard deviation of variance in the right-side panels is rather similar to the one from Figure 3, but this is not the case for the conditional correlation in the left-side panels.

Most model implications can be easily understood by inspecting the parameter estimates in Tables 2-4. In the case of the risk-neutral estimates from options in Table 3, a first important conclusion is that the component models are more persistent than the GARCH(1,1) model, but the differences are smaller than in the case of the return-based estimates in Table 2. As a result, the impact of the current conditions on the future variance is larger for the component models, but the differences with the GARCH(1,1) model are larger for the return-based estimates. Second, results are always very similar for the Gaussian and Inverse Gaussian models, which is not surprising. Third, and most importantly, the risk-neutral dynamics are more persistent than physical dynamics. As a result, the impact of the current conditions on the future variance is much larger for the option-implied risk-neutral estimates, regardless of the model.

When estimating the models using returns and options sequentially in Table 4, the persistence of the models, and consequently the impact of the current conditions on the future variance, is close to the physical persistence based on returns in Table 2 since we fix the physical parameters in this estimation to the optimized returns-based parameter estimates.

5.4 The Relative Importance of Model Features for Option RMSE

We now perform an assessment of the relative importance of the three model features for option fitting. To this end consider the “All” RMSE in the last column of Table 5 which contains the implied volatility root mean squared error across all options. Panel A uses the return-based estimates from Table 2, Panel B uses the option-based estimates from Table 3, and Panel C uses the sequential estimates from Table 4.

The last column in Table 5 enables us to make six pairwise comparisons of GARCH(1,1) and component GARCH(C) models. The improvement from adding a second volatility factor ranges from 4.87% \((1 - 5.0694/5.3289)\) and 3.8% in Panel A, to 16.24% and 15.45% in Panel B, and finally 9.71% and 6.6% in Panel C. On average the improvement from adding a second volatility factor is 9.45%. The improvement from adding a second volatility factor is largest in Panels B and C where the non-monotonic pricing kernel affects the results. The second volatility component and the U-shaped pricing kernel thus appear to be complements.

The last column in Table 5 also enables us to compute six pairwise comparisons of GARCH
versus IG-GARCH models. The IV-RMSE improvement from adding fat tails ranges from 2.12% and 1.02% in Panel A, to 6.18% and 5.29% in Panel B, and 7.38% and 4.20% in Panel C. The overall improvement from adding fat tails is 4.4% and thus considerably lower than from adding a second volatility factor. The improvement from adding fat tails is again largest in Panels B and C where the non-monotonic pricing kernel affects the results. Fat tails and a U-shaped pricing kernel thus also appear to be complements rather than substitutes.

Finally, comparing Panels C and A in Table 5 allows us assess the importance of a U-shaped versus a linear pricing kernel. The improvement from allowing for a U-shaped kernel is 13.39% (1 – 4.6155/5.3289) for the GARCH(1,1) model, 18.05% for the IG-GARCH(1,1) model, 17.80% for the GARCH(C) model, and 20.44% for the IG-GARCH(C) model. On average the improvement is 17.42%. The improvement from allowing for a U-shaped kernel is larger for IG than for Gaussian GARCH models, and it is larger for two-factor than for single-factor models which again suggests that the three features we investigate are complements rather than substitutes.

5.5 Capturing Smiles and Smirks

In Tables 5 and 6 we further investigate the model option fit across the moneyness and maturity categories defined in Table 1. Tables 5 and 6 report implied volatility RMSE and bias (in percent) by moneyness, and maturity, respectively.

Table 5 shows that the IG-GARCH(C) model we propose fits the data best in almost all moneyness categories. Not surprisingly, all models have most difficulty fitting the deep in-the-money calls (corresponding to deep out-of-the-money puts) which are very expensive. It is also not surprising that the fit in Panel B is almost always better than in Panel C which in turn is better than in Panel A. In Panel B, the option fit drives all the parameter estimates, in Panel C only $\xi$ is estimated on options, whereas in Panel A no parameters are fitted to option prices. Again, the most important conclusion from Table 5 is the the IG-GARCH(C) model performs well regardless of implementation and moneyness category.

Panel A of Table 5 also shows that the large RMSEs are largely driven by bias. The bias is defined as market IV less model IV. Positive numbers thus indicate that the model underprices options on average. Panel A shows that the models with linear pricing kernel estimated on returns only have large positive biases in every moneyness category. In Panel B, where all parameters are estimated on options, the bias is much closer to zero. In Panel C the bias is much smaller than in Panel A but it is still fairly large for deep in-the-money calls.

Table 6 reports the implied volatility RMSE and bias by maturity. The IG-GARCH component model now performs the best in all categories. Table 6 also shows that all models tend to
underprice options (i.e. positive bias) at most maturities except for the very long-dated options.

Tables 5 and 6 indicate that the fat-tailed Inverse Gaussian distribution is also helpful in fitting the data. Fat-tailed innovations increase the values of short-term out-of-the-money options, whereas two-factor dynamics increase the tails and values of long-term out-of-the-money options. Tables 5 and 6 demonstrate that these model features are to some extent complementary.

The increases in likelihood due to fat-tailed innovations are much smaller than those due to the second volatility factor. This observation is consistent across estimation exercises and is confirmed by inspecting stylized facts. Figure 5 indicates that the variance paths are very similar for the models with Gaussian and Inverse Gaussian innovations for the option-based estimation results. However, this is unsurprising and not necessarily very relevant for the purpose of option valuation. Models with very similar variance paths can greatly differ with respect to their (conditional) third and fourth moments, and these model properties are of critical importance for option valuation, and for capturing smiles and smirks in particular. Therefore, we again look at conditional correlation and standard deviation of variance paths for the options-based estimations in Figure 6, which indicates substantial differences between the conditional correlation and standard deviation of variance paths for the Gaussian and Inverse Gaussian models. However, perhaps somewhat surprisingly, Figures 5 and 6 clearly indicate that the differences between the GARCH(1,1) and component models are actually larger than the differences between the Gaussian and Inverse Gaussian models in this dimension. This is surprising because a priori we expect the second factor to be more important for term structure modeling, as confirmed by Figure 4. The conditional moments in Figures 5 and 6 are more important for the modeling of smiles and smirks, and a priori we expect the modeling of the conditional innovation to be more important in this dimension. However, it seems that the second volatility factor is also of first-order importance in this dimension.

Figure 7 further illustrates the component model’s flexibility. We plot model-based implied volatility smiles using our proposed IG component model and the parameter values from Table 4. The total spot volatility, $\sqrt{h(t)}$, is fixed at 25% per year in all panels. In the top panel, the long run volatility factor, $\sqrt{q(t)}$ is set to 20%, in the middle panel it is set to 25%, and in the bottom the top panel it is set to 30%. We also show the IG-GARCH(1,1) model for reference. It is of course the same across the three panels. Figure 7 shows that the second volatility factor gives the model a great deal of flexibility in modeling the implied volatility smile.

5.6 Model-Implied Relative Risk Aversion

When using the standard log-linear pricing kernel, the coefficient of relative risk aversion is simply (the negative of) $\phi$. In the non-monotonic pricing kernel the computation of risk-aversion
is slightly more involved and we therefore provide some discussion here.

Assume a representative agent with utility function $U(S(t))$ then the one-period coefficient of relative risk aversion can be written

$$RRA(t) = -S(t) \frac{U''(S(t))}{U'(S(t))} = -S(t) \frac{M'(t)}{M(t)} = -S(t) \frac{\partial \ln (M(t))}{\partial S(t)},$$

where we have used the insight of Jackwerth (2000) to link risk aversion to the pricing kernel. From (18) we have that

$$\frac{\partial \ln (M(t))}{\partial S(t)} = \frac{\phi}{S(t)} + \xi \frac{\partial h(t + \Delta)}{\partial S(t)}.$$  (43)

In the Gaussian model we have

$$\frac{\partial h(t + \Delta)}{\partial S(t)} = \frac{\partial h(t + \Delta)}{\partial z(t)} \frac{\partial z(t)}{\partial S(t)} = \frac{2\alpha_1}{\sqrt{h(t)S(t)}} \left(z(t) - \gamma_1 \sqrt{h(t)}\right).$$  (44)

Combining (43) and (44) we get a relative risk aversion of

$$RRA(t) = -\phi - \frac{2\alpha_1 \xi}{\sqrt{h(t)}} \left(z(t) - \gamma_1 \sqrt{h(t)}\right).$$

Note as indicated above that the parameter $\phi$ does not in itself capture relative risk aversion unless $\xi = 0$ which corresponds to the linear pricing kernel.

Using the law of iterated expectations we can now compute the expected $RRA$ as

$$E[RRA(t)] = -\phi + 2\alpha_1 \gamma_1.$$

Using the GARCH(1,1) parameter estimates in Tables 2 and 4, and the results in Appendix B of Christoffersen, Heston and Jacobs (2013), we get

$$\phi = -(\tilde{\mu} + \gamma_1) \left(1 - 2\alpha_1 \xi\right) + \gamma_1 - \frac{1}{2} \approx 20.26,$$

so that we get

$$E[RRA(t)] \approx -20.26 + 2\alpha_1 \gamma_1 \approx 1.44.$$  

This result shows that the non-monotonic pricing kernel delivers reasonable coefficients of relative risk aversion, and furthermore that it is important not to rely on (the negative of) $\phi$ as a measure of $RRA$ when using the non-monotonic pricing kernel. Determining which equilibrium models are consistent with our pricing kernel is an interesting question that we leave for future
work.

6 Conclusion

We find that multiple volatility factors, fat-tailed return innovations, and a variance-dependent pricing kernel all provide economically and statistically significant improvements in describing S&P500 returns and option prices. A U-shaped pricing kernel is economically most important and improves option fit by 17% on average and more so for two-factor models. A second volatility factor improves the option fit by 9% on average. Fat tails improve option fit by just over 4% on average, and more so when a U-shaped pricing kernel is applied. Our results show that overall these three features are complements rather than substitutes. This indicates that while proper specification of volatility dynamics is quantitatively most important in option models, the interdependent explanatory power of different features make it essential to evaluate them in a properly specified model that nests all of these features.
Appendix A: Martingale Restrictions

A.1 Restrictions Implied by the Risk-free Asset

We first impose on the pricing kernel that the risk-free bond price is a martingale under the risk-neutral measure. We need

\[ E_t \left[ \frac{M(t + \Delta)}{M(t)} B_r(t + \Delta) \right] = B_r(t) \]

where \( B_r(t) \) is a bond with maturity \( \tau \) at time \( t \) and \( M(t+\Delta)/M(t) = (S(t+\Delta)/S(t))^{\phi} \exp(\delta(t + \Delta) + \xi h(t + 2\Delta)) \) where \( \delta(t + \Delta) \equiv \delta_0 + \delta_1 h(t + \Delta) \). WLOG we assume that the risk-free rate is constant so that \( B_r(t+\Delta)/B_r(t) \equiv \exp(r_f) \). We can now write

\[ 1 = E_t [\exp(\phi r(t + \Delta) + \delta_0 + \delta_1 h(t) + \xi h(t + 2\Delta) + r_f)] \]  \hspace{1cm} (45)

The martingale restriction in equation (45) implies that we need to impose the following parameter restrictions on the pricing kernel,

\[ \delta_0 = -(1 + \phi) r_f - \xi w + \frac{1}{2} \ln(1 - 2\xi a_1 \eta_4) \]  \hspace{1cm} (46)

\[ \delta_1 = -\phi \mu - \xi b_1 - \eta^{-2} \left( 1 - \sqrt{(1 - 2\xi a_1 \eta_4)(1 - 2(\phi \eta + \xi c_1))} \right) \]  \hspace{1cm} (47)

where we have used the following property of the IG distribution

\[ E_t[\exp (\alpha y(t + \Delta) + \beta / y(t + \Delta))] = \]  \hspace{1cm} (48)

\[ \frac{1}{\sqrt{1 - 2\beta h(t + \Delta) - 2\eta^4}} \times \exp \left[ h(t + \Delta)/\eta^2 \left( 1 - \sqrt{(1 - 2\beta h(t + \Delta)-2\eta^4)(1 - 2\alpha)} \right) \right]. \]

A.2 Restrictions Implied by the Risky Asset

We next impose on the pricing kernel that the risky stock is a martingale under the risk-neutral measure. We now need

\[ E_t \left[ \frac{M(t + \Delta)}{M(t)} S(t + \Delta) \right] = S(t) \]

We can thus write

\[ 1 = E_t [\exp(\phi r(t + \Delta) + \delta_0 + \delta_1 h(t + \Delta) + \xi h(t + 2\Delta) + r(t + \Delta))] \]
Taking logs this condition implies the following restriction on \( \mu \),

\[
\mu = \eta^{-2} \sqrt{(1 - 2\alpha_1 \eta_4)} \left[ \sqrt{1 - 2(\eta + \phi \eta + \xi c)} - \sqrt{1 - 2(\phi \eta + \xi c)} \right],
\]

where we have used equation (48).

**Appendix B: The Risk-Neutral Distribution**

Consider the physical probability density function of the IG stock price

\[
f_{t-\Delta}(S(t)) = f_{t-\Delta}(y(t)) \left| \frac{\partial y(t)}{\partial S(t)} \right| \quad (49)
\]

\[
= \frac{h(t)/\eta^3}{\sqrt{2\pi y(t)^3 S(t)}} \exp \left( - \frac{1}{2} \left[ \sqrt{y(t)} - \frac{h(t)/\eta^2}{\sqrt{y(t)}} \right]^2 \right)
\]

To find the risk-neutral dynamic, we use the price kernel as follows

\[
f_{t-\Delta}^*(S(t)) = f_{t-\Delta}(S(t)) \exp(r_f) M(t)/M(t - \Delta),
\]

where \( M(t - \Delta) \) is \( (t - \Delta) \)-measurable.

Using the pricing kernel definition in (18) and the IG-GARCH(1,1) return dynamic, we can write

\[
f_{t-\Delta}^*(S(t)) = f_{t-\Delta}(S(t)) \exp[r_f + \delta_0 + \delta_1 h(t) + \phi \ln(S(t)/S(t - \Delta)) + \xi h(t + 2\Delta)]
\]

\[
= \frac{h(t)/\eta^3 \sqrt{1 - 2\alpha \eta^4}}{\sqrt{2\pi y(t)^3 S(t)}} \exp \left[ - \frac{1}{2} \left( \sqrt{(1 - 2\phi \eta - 2\xi c) y(t)} - \frac{h(t)/\eta^2}{\sqrt{y(t)}} \sqrt{1 - 2\alpha \eta^4} \right)^2 \right]
\]

Substituting the physical distribution from equation (49) and rearranging terms yields

\[
f_{t-\Delta}^*(S(t)) = \frac{h(t) \sqrt{(1 - 2\xi \alpha \eta^4)(1 - 2\phi \eta - 2\xi c)^{-3}(1 - 2\phi \eta - 2\xi c)^3 / \eta^3}}{\sqrt{2\pi y(t)^3 S(t)}} \times
\]

\[
\exp \left[ - \frac{1}{2} \left( \sqrt{y(t)}(1 - 2\phi \eta - 2\xi c) \right)^2 \right].
\]
This enables us to define the risk-neutral counterparts to $y(t), h(t),$ and $\eta$ by

$$y^*(t) = y(t)(1 - 2\phi\eta - 2\xi c) = y(t)s_y,$$
$$h^*(t) = h(t)\sqrt{(1 - 2\xi a\eta^4)(1 - 2\phi\eta - 2\xi c)^{-3}} = h(t)s_h,$$
$$\eta^* = \eta/(1 - 2\phi\eta - 2\xi c) = \eta/s_y,$$

where we have used the definitions

$$s_y = 1 - 2\phi\eta - 2\xi c,$$
$$s_h = \sqrt{1 - 2\xi a\eta^4s_y^{-3/2}},$$

as in the text. Using these mappings yields the risk neutral density

$$f_{t-\Delta}(S(t)) = \frac{h^*(t)/|\eta^*|^3}{\sqrt{2\pi(y^*(t))^3}S(t)} \exp\left[-\frac{1}{2}\left(\sqrt{y^*(t)} - \frac{h^*(t)/|\eta^*|^2}{\sqrt{y^*(t)}}\right)^2\right].$$

So that,

$$f_{t-\Delta}(y^*(t)) = f_{t-\Delta}(S(t)) \left| \frac{\partial S(t)}{\partial y^*(t)} \right| = f_{t-\Delta}(S(t)) |S(t) \times (-\eta^*)|$$

$$= \frac{h^*(t)/|\eta^*|^2}{\sqrt{2\pi(y^*(t))^3}} \exp\left[-\frac{1}{2}\left(\sqrt{y^*(t)} - \frac{h^*(t)/|\eta^*|^2}{\sqrt{y^*(t)}}\right)^2\right].$$

Therefore $y^*(t)$ is distributed Inverse-Gaussian, and we can write,

$$y^*(t) \sim IG\left(\frac{h^*(t)}{(\eta^*)^2}\right).$$

Using the physical return process and the above mappings we can write the risk-neutral return process as

$$\ln(S(t + \Delta)) = \ln(S(t)) + r + \mu h^*(t + \Delta)/s_h + \eta y^*(t + \Delta)/s_y$$
$$h^*(t + \Delta) = w s_h + b h^*(t) + c y^*(t) s_h/s_y + a \frac{s_y}{s_h} h^*(t)^2 / y^*(t),$$
or equivalently

\[
\ln(S(t + \Delta)) = \ln(S(t)) + r + \mu^* h^*(t + \Delta) + \eta^* y^*(t + \Delta)
\]

\[
h^*(t + \Delta) = w^* + bh^*(t) + c^* y^*(t) + a^* h^*(t)^2 / y^*(t),
\]

where we have used the parameter mapping in equation (21b) and (21c).

**Appendix C: The Risk-Neutral Component Model**

The component representation of the risk-neutral process (20) is given by

\[
\ln(S(t + \Delta)) = \ln(S(t)) + r + \mu^* h(t + \Delta) + (\eta^* y(t + \Delta) - h^*(t + \Delta) / \eta^*),
\]

\[
h^*(t + \Delta) = q^*(t + \Delta) + \rho_1^* (h^*(t) - q^*(t)) + \nu_{h}(t),
\]

\[
q^*(t + \Delta) = \sigma^{*2} + \rho_2^* (q^*(t) - \sigma^{*2}) + \nu_{q}(t),
\]

where

\[
q^*(t) = \frac{-\rho_1^* \bar{w}^*}{(1 - \rho_1^*)(\rho_2^* - \rho_1^*)} + \frac{\rho_2^*}{\rho_2^* - \rho_1^*} h^*(t) + \frac{\tilde{b}_2^*}{\rho_2^* - \rho_1^*} h^*(t - \Delta) + \frac{1}{\rho_2^* - \rho_1^*} \nu_{2}(t - \Delta),
\]

\[
\tilde{\mu} = \mu^* + \eta^{* -1} = \mu / s_h + s_y \eta^{-1};
\]

\[
\sigma^{*2} = \left( \frac{\rho_2^*}{1 - \rho_2^*} - \frac{\rho_1^*}{1 - \rho_1^*} \right) \bar{w}^*;
\]

\[
\bar{w}^* = w^* + a_1^* \eta^{*4} + a_2^* \eta^{*4} = s_h w + \frac{a_1 \eta^4 + a_2 \eta^4}{s_h s_y^3},
\]

\[

\nu_{h}(t) = c_h^* y^*(t) + a_h^* h^*(t)^2 / y^*(t) - c_h^* h^*(t) / \eta^{*2} - a_h^* \eta^{*2} h^*(t) - a_h^* \eta^{*4},
\]

\[
\nu_{q}(t) = c_q^* y^*(t) + a_q^* h^*(t)^2 / y^*(t) - c_q^* h^*(t) / \eta^{*2} - a_q^* \eta^{*2} h^*(t) - a_q^* \eta^{*4}.
\]
\[ a^*_h = -\frac{\rho^*_1 a^*_1}{\rho^*_2 - \rho^*_1} - \frac{1}{\rho^*_2 - \rho^*_1} a^*_2, \]
\[ c^*_h = -\frac{\rho^*_1 c^*_1}{\rho^*_2 - \rho^*_1} - \frac{1}{\rho^*_2 - \rho^*_1} c^*_2, \]
\[ a^*_q = \frac{\rho^*_2}{\rho^*_2 - \rho^*_1} a^*_1 + \frac{1}{\rho^*_2 - \rho^*_1} a^*_2, \]
\[ c^*_q = \frac{\rho^*_2}{\rho^*_2 - \rho^*_1} c^*_1 + \frac{1}{\rho^*_2 - \rho^*_1} c^*_2, \]
\[ \tilde{b}^*_i = b_i + s_h c_i / \eta^2 + \frac{a_i \eta^2}{s_h s_y}, \]

and where \( \rho^*_1 \) and \( \rho^*_2 \) are the smaller and larger respective roots of the equation \( \rho^*_2 - \tilde{b}^*_i \rho - \tilde{b}^*_2 = 0 \).

**Appendix D: Conditional Moments**

Consider the following basic definitions

\[ Var_t[h(t + 2\Delta)] \equiv E_t \left[ (h(t + 2\Delta) - E_t[h(t + 2\Delta)])^2 \right] \]
\[ Cov_t[R(t + \Delta), h(t + 2\Delta)] \equiv E_t \left[ (R(t + \Delta) - E_t[R(t + \Delta)]) (h(t + 2\Delta) - E_t[h(t + 2\Delta)]) \right] \]

where \( R(t + \Delta) \equiv \ln S(t + \Delta) - \ln S(t) \).

In this section, we only focus on the derivation of conditional correlation, and conditional standard deviation of variance for IG-GARCH(C) model, since derivations for other models are similar.

Recall that the standardized conditional moments of an Inverse Gaussian random variable \( y(t + 1) \) are given by:

\[ E_t[y(t + \Delta)] = \delta(t + \Delta) \]
\[ Var_t[y(t + \Delta)] = \delta(t + \Delta) \]
\[ E_t[1/y(t + \Delta)] = 1/\delta(t + \Delta) + 1/\delta(t + \Delta)^2 \]
\[ Var_t[1/(t + \Delta)] = 1/\delta(t + \Delta)^3 + 2/\delta(t + \Delta)^4 \]
\[ Cov_t[y(t + \Delta), 1/y(t + \Delta)] = -1/\delta(t + \Delta), \]

where the degree of freedom is defined by

\[ \delta(t + \Delta) = h(t + \Delta)/\eta^2. \]
The variance process is defined as

\[
\begin{align*}
    h(t + \Delta) &= q(t + \Delta) + \rho_1 [h(t) - q(t)] + v_h(t) \\
    q(t + \Delta) &= w_q + \rho_2 q(t) + v_q(t) \\
    v_h(t) &= c_h [y(t) - \delta(t)] + a_h h(t) [1/y(t) - 1/\delta(t) - 1/\delta(t)^2] \\
    v_q(t) &= c_q [y(t) - \delta(t)] + a_q h(t) [1/y(t) - 1/\delta(t) - 1/\delta(t)^2].
\end{align*}
\]

Conditional variance of variance is given by

\[
\begin{align*}
    \text{Var}_t[h(t + 2\Delta)] &= (c_h + c_q)^2 h(t + \Delta)/\eta^2 - 2(a_h + a_q)(c_h + c_q)\eta^2 h(t) \\
    &\quad + (a_h + a_q)^2 \eta^6 h(t) + 2(a_h + a_q)^2 \eta^8.
\end{align*}
\]

We thus can write

\[
\text{Std}_t[h(t + 2\Delta)] = \sqrt{2(a_h + a_q)^2 \eta^8 + [(c_h + c_q)/\eta - (a_h + a_q)\eta^3]^2 h(t + \Delta)}
\]

Consider now the innovation to returns

\[
R(t + \Delta) - E_t[R(t + \Delta)] = \eta(y(t + \Delta) - \delta(t + \Delta))
\]

We can then derive covariance and correlation

\[
\begin{align*}
    \text{Cov}_t[R(t + \Delta), h(t + 2\Delta)] &= (c_h + c_q)\eta \text{Var}_t[y(t + \Delta)] \\
    &\quad + (a_h + a_q)\eta h(t + \Delta)^2 \text{Cov}_t[y(t + \Delta), 1/y(t + \Delta)] \\
    &= (c_h + c_q)/\eta h(t + \Delta) - (a_h + a_q)\eta^3 h(t + \Delta) \\
    \text{Corr}_t[R(t + \Delta), h(t + 2\Delta)] &= \frac{\text{Cov}_t[R(t + \Delta), h(t + 2\Delta)]}{\sqrt{\text{Var}_t[R(t + \Delta)]\text{Var}_t[h(t + 2\Delta)]}} \\
    &= \frac{[(c_h + c_q)/\eta - (a_h + a_q)\eta^3] h(t + \Delta)}{\sqrt{2(a_h + a_q)^2 \eta^8 + [(c_h + c_q)/\eta - (a_h + a_q)\eta^3]^2 h(t + \Delta)}}.
\end{align*}
\]
References


Notes to Figure: We plot the RMSE (top panel) and the option likelihood function (bottom panel) as a function of the non-monotonic pricing kernel parameter, $\xi$. All other parameter values are fixed at their optimal values from Table 2.
Figure 2. Spot Variance Paths Using Return-Based Estimates

Notes to Figure: For each model we plot the spot variance components over time. The parameter values are obtained from MLE on returns in Table 2.
Figure 3. Leverage Correlation and Volatility of Variance Using Return-Based Estimates

Notes to Figure: For each model we plot the conditional correlation and the conditional standard deviation of variance. In the left panels, we plot the conditional correlation between return and variance as implied by the models. In the right panels, we plot the conditional standard deviation of conditional variance. The scales are identical across the rows of panels to facilitate comparison across models. The parameter values are obtained from MLE on returns in Table 2.
Notes to Figure: We plot the term structure of variance, skewness and excess kurtosis with high (solid) and low (dashed) initial variance for 1 through 250 trading days. Conditional variance is normalized by the unconditional variance, $\sigma^2$. For the low initial variance, the initial value of $q(t + \Delta)$ is set to $0.75\sigma^2$, and the initial value of $h(t + \Delta)$ is set to $0.5\sigma^2$. For the high initial variance, the initial value of $q(t + \Delta)$ is set to $1.75\sigma^2$, and the initial value of $h(t + \Delta)$ is set to $2\sigma^2$. The return-based parameter values from Table 2 are used.
Notes to Figure: We plot the spot variance components over time. The parameter values are obtained from MLE on options in Table 3.
Figure 6. Leverage Correlation and Volatility of Variance Using Option-Based Estimates

Notes to Figure: For each model we plot the conditional correlation and the conditional standard deviation of variance. In the left panels, we plot the conditional correlation between return and variance as implied by the models. In the right panels, we plot the conditional standard deviation of conditional variance. The scales are identical across the rows of panels to facilitate comparison across models. The parameter values are obtained from MLE on options in Table 3.
Figure 7. Model-Based Implied Volatility Smiles in IG-GARCH Component Model

Notes to Figure: We plot model-based implied volatility smiles for 30 days to maturity from the IG-GARCH(1,1) and IG-GARCH(C) models. Long-run volatility, $\sqrt{q(t)}$, is set to 20% (top) panel, 25% (middle panel), and 30% (bottom panel). Total volatility, $\sqrt{h(t)}$ is set to 25% in all panels. The parameter estimates from Table 4 are used to generate the model prices. Model implied volatilities are calculated by inverting the Black-Scholes formula on the model prices.
We present descriptive statistics for daily return data from January 2, 1990 through December 31, 2012, as well as for daily return data from January 10, 1996 through December 26, 2012. We use Wednesday closing options contracts from January 10, 1996 through December 26, 2012.

### Table 1: Returns and Options Data

#### Panel A. Return Characteristics (Annualized)

<table>
<thead>
<tr>
<th></th>
<th>1990-2012</th>
<th>1996-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6.06%</td>
<td>4.99%</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>18.61%</td>
<td>20.57%</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.228</td>
<td>-0.217</td>
</tr>
<tr>
<td>Excess kurtosis</td>
<td>8.461</td>
<td>7.235</td>
</tr>
</tbody>
</table>

#### Panel B. Option Data by Moneyness

<table>
<thead>
<tr>
<th>F/X ≤ .80</th>
<th>.80 &lt; F/X ≤ .90</th>
<th>.90 &lt; F/X ≤ 1.00</th>
<th>1.00 &lt; F/X ≤ 1.10</th>
<th>1.10 &lt; F/X ≤ 1.20</th>
<th>F/X &gt; 1.20</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Contrac</td>
<td>720</td>
<td>3,819</td>
<td>8,413</td>
<td>8,033</td>
<td>5,778</td>
<td>2,259</td>
</tr>
<tr>
<td>Average IV</td>
<td>23.11%</td>
<td>19.65%</td>
<td>18.79%</td>
<td>22.09%</td>
<td>27.03%</td>
<td>30.52%</td>
</tr>
<tr>
<td>Average Price</td>
<td>62.94</td>
<td>40.71</td>
<td>43.62</td>
<td>47.93</td>
<td>33.18</td>
<td>28.14</td>
</tr>
<tr>
<td>Average Spread</td>
<td>1.30</td>
<td>1.42</td>
<td>1.89</td>
<td>2.06</td>
<td>1.58</td>
<td>1.41</td>
</tr>
</tbody>
</table>

#### Panel C. Option Data by Maturity

<table>
<thead>
<tr>
<th>DTM ≤ 30</th>
<th>30 &lt; DTM ≤ 60</th>
<th>60 &lt; DTM ≤ 90</th>
<th>90 &lt; DTM ≤ 120</th>
<th>120 &lt; DTM ≤ 180</th>
<th>DTM &gt; 180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Contrac</td>
<td>2,771</td>
<td>6,127</td>
<td>4,565</td>
<td>2,720</td>
<td>4,019</td>
<td>8,820</td>
</tr>
<tr>
<td>Average IV</td>
<td>24.93%</td>
<td>23.33%</td>
<td>22.45%</td>
<td>22.93%</td>
<td>21.74%</td>
<td>21.32%</td>
</tr>
<tr>
<td>Average Price</td>
<td>18.06</td>
<td>26.83</td>
<td>31.88</td>
<td>39.29</td>
<td>48.54</td>
<td>61.92</td>
</tr>
<tr>
<td>Average Spread</td>
<td>0.94</td>
<td>1.31</td>
<td>1.59</td>
<td>1.87</td>
<td>1.97</td>
<td>2.29</td>
</tr>
</tbody>
</table>
### Table 2: Maximum Likelihood Estimation Results for Return Distribution

#### Panel A: Parameter Estimates

<table>
<thead>
<tr>
<th>Gaussian Models</th>
<th>( \hat{\mu} )</th>
<th>( \omega )</th>
<th>( \hat{\beta}_1 )</th>
<th>( \alpha_1 )</th>
<th>( \gamma_1 )</th>
<th>( \hat{\phi}_1 )</th>
<th>( \alpha_h )</th>
<th>( \gamma_h )</th>
<th>( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoskedastic</td>
<td>0.78</td>
<td>1.373E-04</td>
<td></td>
<td></td>
<td></td>
<td>(1.12E+0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>1.10</td>
<td>-1.396E-06</td>
<td>0.900</td>
<td>3.761E-06</td>
<td>145.7</td>
<td>(1.35E-7)</td>
<td>(7.84E-3)</td>
<td>(2.30E-7)</td>
<td>(1.02E+1)</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>1.26</td>
<td>1.473E-06</td>
<td>0.705</td>
<td>9.979E-07</td>
<td>840.6</td>
<td>(1.53E-7)</td>
<td>(3.13E-2)</td>
<td>(2.70E+6)</td>
<td>(1.52E-3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IG Models</th>
<th>( \hat{\mu} )</th>
<th>( w )</th>
<th>( b_1 )</th>
<th>( a_1 )</th>
<th>( c_1 )</th>
<th>( \hat{\phi}_1 )</th>
<th>( a_h )</th>
<th>( c_h )</th>
<th>( \eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoskedastic</td>
<td>0.78</td>
<td>1.372E-04</td>
<td>-21.82</td>
<td>3.190E+07</td>
<td>4.047E-06</td>
<td>(1.28E-7)</td>
<td>(3.88E+0)</td>
<td>(1.03E+7)</td>
<td>(2.15E-7)</td>
</tr>
<tr>
<td>IG-GARCH(1,1)</td>
<td>1.16</td>
<td>-1.469E-06</td>
<td>0.0979</td>
<td>3.761E-06</td>
<td>145.7</td>
<td>(1.35E-7)</td>
<td>(7.84E-3)</td>
<td>(2.30E-7)</td>
<td>(1.02E+1)</td>
</tr>
<tr>
<td>IG-GARCH(C)</td>
<td>1.23</td>
<td>1.393E-06</td>
<td>0.743</td>
<td>9.979E-07</td>
<td>840.6</td>
<td>(1.53E-7)</td>
<td>(3.13E-2)</td>
<td>(2.70E+6)</td>
<td>(1.52E-3)</td>
</tr>
</tbody>
</table>

#### Panel B: Model Properties and Likelihoods

<table>
<thead>
<tr>
<th>Gaussian Models</th>
<th>Return Mean</th>
<th>Annualized Volatility</th>
<th>Volatility Persistence</th>
<th>Uncond. Skewness</th>
<th>Uncond. Kurtosis</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoskedastic</td>
<td>5.99%</td>
<td>18.60%</td>
<td>0.000</td>
<td>3.000</td>
<td>17,548.0</td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>6.48%</td>
<td>17.00%</td>
<td>0.979387</td>
<td>0.015</td>
<td>4.750</td>
<td>18,781.4</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>6.91%</td>
<td>16.91%</td>
<td>0.996170</td>
<td>0.024</td>
<td>5.199</td>
<td>18,831.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoskedastic</td>
<td>5.99%</td>
<td>18.59%</td>
<td>-0.033</td>
<td>3.000</td>
<td>17,551.5</td>
<td></td>
</tr>
<tr>
<td>IG-GARCH(1,1)</td>
<td>6.61%</td>
<td>16.92%</td>
<td>0.982695</td>
<td>-0.152</td>
<td>4.775</td>
<td>18,829.5</td>
</tr>
<tr>
<td>IG-GARCH(C)</td>
<td>6.78%</td>
<td>16.81%</td>
<td>0.996802</td>
<td>-0.099</td>
<td>5.247</td>
<td>18,865.0</td>
</tr>
</tbody>
</table>

Parameter estimates are obtained by an MLE estimation on returns from January 2, 1990 through December 31, 2012. Data description can be found in Table 1. For each model we report parameter estimates, the maximum log-likelihood values and some model properties. Robust standard errors (based on the outer product of gradients) are in parantheses below the parameter estimates. We estimate six models. Each model has constant or time-varying volatility (which is either two components or one), Normal or IG innovations.
Table 3: Maximum Likelihood Estimation Results for Risk Neutral Distribution

Panel A. Parameter Estimates

<table>
<thead>
<tr>
<th>Gaussian Models</th>
<th>( \mu^{*} )</th>
<th>( \omega^{*} )</th>
<th>( \beta^{*} )</th>
<th>( \alpha^{*} )</th>
<th>( \gamma^{*} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoskedastic</td>
<td>-0.50</td>
<td>1.272E-04</td>
<td>0.823</td>
<td>2.931E-06</td>
<td>241.23</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>-0.50</td>
<td>-1.260E-06</td>
<td>1.17E-3</td>
<td>8.76E-9</td>
<td>(8.76E-1)</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>-0.50</td>
<td>2.877E-07</td>
<td>0.981</td>
<td>4.587E-07</td>
<td>2096.18</td>
</tr>
<tr>
<td>IG Models</td>
<td>( \mu^{*} )</td>
<td>( \omega^{*} )</td>
<td>( \beta^{*} )</td>
<td>( \alpha^{*} )</td>
<td>( \gamma^{*} )</td>
</tr>
<tr>
<td>Homoskedastic</td>
<td>-0.48</td>
<td>1.593E-04</td>
<td>2.50</td>
<td>4.931E+05</td>
<td>5.841E-06</td>
</tr>
<tr>
<td>IG-GARCH(1,1)</td>
<td>-0.50</td>
<td>-1.956E-06</td>
<td>9.41E-3</td>
<td>5.32E+3</td>
<td>(3.46E-8)</td>
</tr>
<tr>
<td>IG-GARCH(C)</td>
<td>-0.50</td>
<td>3.068E-07</td>
<td>0.984</td>
<td>4.313E+06</td>
<td>3.017E-06</td>
</tr>
</tbody>
</table>

Panel B: Model Properties and Likelihoods

<table>
<thead>
<tr>
<th>Gaussian Models</th>
<th>Return Mean</th>
<th>Annualized Volatility</th>
<th>Volatility Persistence</th>
<th>Uncond. Skewness</th>
<th>Uncond. Kurtosis</th>
<th>Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoskedastic</td>
<td>1.69%</td>
<td>17.90%</td>
<td>0.000</td>
<td>3.000</td>
<td>32,632.1</td>
<td></td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>0.21%</td>
<td>24.85%</td>
<td>0.993182</td>
<td>-0.017</td>
<td>5.113</td>
<td>53,971.2</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>-4.09%</td>
<td>38.44%</td>
<td>0.999990</td>
<td>-0.018</td>
<td>5.485</td>
<td>59,124.7</td>
</tr>
</tbody>
</table>

Parameter estimates are obtained by an MLE estimation on options from January 10, 1996 through December 26, 2012. Data description can be found in Table 1. For each model we report parameter estimates, the maximum log-likelihood values and some model properties. Robust standard errors (based on the outer product of gradients) are in parentheses below the parameter estimates. We estimate six models using only options data. Each model has constant or time-varying volatility (which is either two components or one), and Normal or IG innovations.
Parameter estimates are obtained by an sequential MLE estimation on options from January 10, 1990 through December 26, 2012. Data description can be found in Table 1. For each model we report parameter estimates, the maximum log-likelihood values and some model properties. Robust standard errors (based on the outer product of gradients) are in parantheses below the parameter estimates. We estimate preference parameter $\xi$ of four models using the returns-based parameters reported in Table 2 by applying the transformations (from physical to risk-neutral measure) mentioned in the appendix. Each model has constant or time-varying volatility (which is either two components or one), and Normal or IG innovations. The last column in Panel B reports the increase in log-likelihood going from a linear ($\xi=0$) to nonlinear pricing kernel.
We report implied volatility (IV) RMSE (values before parentheses) and bias (values inside parentheses) in percent by moneyness. The bias is defined as market IV less model IV. Panel A uses the parameter estimates from the return-based estimation in Table 2, Panel B uses the options-based estimates in Table 3, and Panel C uses the sequential estimates in Table 4.
Table 6: Implied Volatility RMSE and Bias by Maturity

Panel A. IV RMSE (Bias) by Maturity for Models Fitted to Returns Only

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM ≤ 30</th>
<th>30 &lt; DTM ≤ 60</th>
<th>60 &lt; DTM ≤ 90</th>
<th>90 &lt; DTM ≤ 120</th>
<th>120 &lt; DTM ≤ 180</th>
<th>DTM&gt;180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>4.6495</td>
<td>5.2143 (3.3967)</td>
<td>5.2047 (3.3524)</td>
<td>5.6930 (3.9329)</td>
<td>5.3252 (3.5045)</td>
<td>5.5520 (3.4450)</td>
<td>5.3289 (3.4274)</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>4.4964</td>
<td>5.0258 (4.4418)</td>
<td>4.9616 (3.3926)</td>
<td>5.3320 (3.8838)</td>
<td>5.0543 (3.5510)</td>
<td>5.2453 (3.5312)</td>
<td>5.0694 (3.4766)</td>
</tr>
<tr>
<td>IG-GARCH(1,1)</td>
<td>4.5423</td>
<td>5.0795 (3.5273)</td>
<td>5.0601 (3.5077)</td>
<td>5.4981 (4.0699)</td>
<td>5.2431 (3.6863)</td>
<td>5.4791 (3.6685)</td>
<td>5.2161 (3.5962)</td>
</tr>
<tr>
<td>IG-GARCH(C)</td>
<td>4.4689</td>
<td>4.9698 (3.5111)</td>
<td>4.8928 (3.4777)</td>
<td>5.2497 (3.9643)</td>
<td>5.0139 (3.6417)</td>
<td>5.2034 (3.6316)</td>
<td>5.0179 (3.5610)</td>
</tr>
</tbody>
</table>

Panel B. IV RMSE (Bias) by Maturity for Models Fitted to Options Only

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM ≤ 30</th>
<th>30 &lt; DTM ≤ 60</th>
<th>60 &lt; DTM ≤ 90</th>
<th>90 &lt; DTM ≤ 120</th>
<th>120 &lt; DTM ≤ 180</th>
<th>DTM&gt;180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>4.3171</td>
<td>4.2394 (1.3154)</td>
<td>3.7668 (0.8594)</td>
<td>3.6197 (1.0708)</td>
<td>3.4035 (0.6858)</td>
<td>3.4165 (0.1294)</td>
<td>3.7663 (0.7732)</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>4.0301</td>
<td>3.7421 (0.9144)</td>
<td>3.2046 (0.5466)</td>
<td>2.9483 (0.6318)</td>
<td>2.7796 (0.4902)</td>
<td>2.5290 (0.0589)</td>
<td>3.1545 (0.5035)</td>
</tr>
<tr>
<td>IG-GARCH(1,1)</td>
<td>4.0868</td>
<td>3.9773 (1.3381)</td>
<td>3.4912 (0.8563)</td>
<td>3.2858 (1.0133)</td>
<td>3.2160 (0.5547)</td>
<td>3.2401 (-0.0475)</td>
<td>3.5336 (0.7051)</td>
</tr>
<tr>
<td>IG-GARCH(C)</td>
<td>3.8771</td>
<td>3.6164 (1.2356)</td>
<td>3.0171 (0.7659)</td>
<td>2.7624 (0.6886)</td>
<td>2.5303 (0.4240)</td>
<td>2.3618 (-0.2367)</td>
<td>2.9875 (0.5460)</td>
</tr>
</tbody>
</table>

Panel C. IV RMSE (Bias) by Maturity for Models Fitted to Options Sequentially

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM ≤ 30</th>
<th>30 &lt; DTM ≤ 60</th>
<th>60 &lt; DTM ≤ 90</th>
<th>90 &lt; DTM ≤ 120</th>
<th>120 &lt; DTM ≤ 180</th>
<th>DTM&gt;180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>GARCH(1,1)</td>
<td>4.3879</td>
<td>4.6955 (1.7208)</td>
<td>4.5165 (1.1667)</td>
<td>4.7743 (1.5080)</td>
<td>4.4305 (0.7540)</td>
<td>4.7115 (0.1848)</td>
<td>4.6155 (1.0174)</td>
</tr>
<tr>
<td>GARCH(C)</td>
<td>4.0662</td>
<td>4.2872 (1.6855)</td>
<td>4.0786 (1.0979)</td>
<td>4.1949 (1.3008)</td>
<td>3.9768 (0.5875)</td>
<td>4.2529 (-0.1007)</td>
<td>4.1672 (0.8692)</td>
</tr>
<tr>
<td>IG-GARCH(1,1)</td>
<td>4.0903</td>
<td>4.3269 (1.6860)</td>
<td>4.1321 (1.0911)</td>
<td>4.2528 (1.3719)</td>
<td>4.1266 (0.6011)</td>
<td>4.4375 (-0.0544)</td>
<td>4.2747 (0.8913)</td>
</tr>
<tr>
<td>IG-GARCH(C)</td>
<td>3.9210</td>
<td>4.1112 (1.6930)</td>
<td>3.8617 (1.0860)</td>
<td>3.9197 (1.2588)</td>
<td>3.8005 (0.5184)</td>
<td>4.1033 (-0.2402)</td>
<td>3.9924 (0.8146)</td>
</tr>
</tbody>
</table>

We report implied volatility (IV) RMSE (values before parentheses) and bias (values inside parentheses) in percent by maturity. The bias is defined as market IV less model IV. Panel A uses the parameter estimates from the return-based estimation in Table 2, Panel B uses the options-based estimates in Table 3, and Panel C uses the sequential estimates in Table 4.