Option Valuation with Volatility Components, Fat Tails, and Nonlinear Pricing Kernels

Kadir G. Babaglou, Peter Christoffersen, Steven L. Heston and Kris Jacobs

CREATES Research Paper 2015-55

Option Valuation with Volatility Components, Fat Tails, and Nonlinear Pricing Kernels[*]

Kadir Babaoğlu

Rotman School of Management

University of Toronto

Peter Christoffersen

University of Toronto,

CBS and CREATES

Steven Heston

Smith School of Business

University of Maryland

Kris Jacobs

University of Houston

and Tilburg University

November 18, 2014

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 second volatility factor is economically most important and improves option fit by 18% on average. A U-shaped pricing kernel improves the option fit by 17% on average, and more so for two-factor models. Fat tails improve option fit by just over 3% on average, and more so when a U-shaped pricing kernel is applied. Our results suggest that the three features we investigate are complements rather than substitutes.

JEL Classification: G12

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

---

[*]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

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). At least two important modeling approaches further improve the model's ability 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).[1] 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 approach 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.[2] 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 return and the options. This position is forcefully articulated by for example Bates (1996, 2003). Any deficiencies of the model in simultaneously modeling the underlying returns and option prices are by design not exclusively due to the specification of the driving process, but also to the specification of the 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 specification of the price of risk when the SV models are augmented with Poisson jumps.

Several papers including Ait-Sahalia and Lo (1998), Bakshi, Madan, and Panayotov (2010), Brown and Jackwerth (2012), Chabi-Yo (2012), Rosenberg and Engle (2002), and Shive and

---

[1]See for instance Chernov, Gallant, Ghysels, and Tauchen (2003) for a study of multiple volatility components in the underlying return series.

[2]See for instance Andersen, Benzoni, and Lund (2002), Bakshi, Cao, and Chen (1997), Bates (1996, 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.

Shumway (2006) have 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 wealth as well as volatility. The kernel is non-monotonic after projecting onto wealth, which is consistent with recent evidence. They conduct an estimation exercise that simultaneously fits returns and options, and show that the more general pricing kernel provides a superior fit.

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, nonlinear pricing kernels. These different model features ought to be complements rather than substitutes. The second volatility factor should improve the modeling of the 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 nonmonotonic 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. These questions are the subject of this paper. 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 second volatility factor is economically most important and improves option fit by 18% on average. A U-shaped pricing kernel improves the option fit by 17% on average across models, and more so for two-factor models. Fat tails improve option fit by just over 3% on average, and 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 the most general return dynamic we consider, with 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   The IG-GARCH(2,2) Model

We first introduce the IG-GARCH(2,2) model. We then show how the IG-GARCH(2,2) model can be transformed into a component model, and show how to specialize the model to nest

3

simpler cases.

2.1 Model Dynamic

The IG-GARCH(2,2) process is given by

$$\ln(S\,(t+\Delta)) = \ln(S(t)) + r + \mu h(t+\Delta) + \eta y(t+\Delta), \tag{1a}$$

$$h(t+\Delta) = w + b_1 h(t) + b_2 h(t-\Delta) + c_1 y(t) + c_2 y(t-\Delta) \tag{1b}$$
$$+a_1 h(t)^2/y(t) + a_2 h(t-\Delta)^2/y(t-\Delta),$$

where $y(t+\Delta)$ has an Inverse Gaussian distribution with degrees of freedom $h(t+\Delta)/\eta^2$. The Inverse Gaussian innovation and its reciprocal have the following conditional means

$$E_t[y(t+\Delta)] = h(t+\Delta)/\eta^2, \tag{2a}$$

$$E_t[1/y(t+\Delta)] = \eta^2/h(t+\Delta) + \eta^4/h(t+\Delta)^2. \tag{2b}$$

The dynamic (1a)-(1b) can be written in terms of zero-mean innovations as follows

$$\ln(S\,(t+\Delta)) = \ln(S(t)) + r + \tilde{\mu}h(t+\Delta) + (\eta y(t+\Delta) - h(t+\Delta)/\eta), \tag{3a}$$

$$h(t+\Delta) = \tilde{w} + \tilde{b}_1 h(t) + \tilde{b}_2 h(t-\Delta) + \upsilon_1(t) + \upsilon_2(t-\Delta), \tag{3b}$$

where

$$\tilde{\mu} = \mu + \eta^{-1}, \tag{4a}$$

$$\tilde{w} = w + a_1\eta^4 + a_2\eta^4, \tag{4b}$$

$$\tilde{b}_i = b_i + c_i/\eta^2 + a_i\eta^2, \tag{4c}$$

$$\upsilon_i(t) = c_i y(t) + a_i h(t)^2/y(t) - c_i h(t)/\eta^2 - a_i\eta^2 h(t) - a_i\eta^4. \tag{4d}$$

The conditional means of return and variance are given by

$$E_t[\ln(S\,(t+\Delta))] = \ln(S(t)) + r + \tilde{\mu}h(t+\Delta), \tag{5a}$$

$$E_t[h(t+2\Delta) = \tilde{w} + \tilde{b}_1 h(t+\Delta) + c_2 y(t) + a_2 h(t)^2/y(t). \tag{5b}$$

2.2 Equivalence with a Component Model

Motivated by Engle and Lee (1999), we now transform the IG-GARCH(2,2) into a component model that nests Christoffersen, Jacobs, Ornthanalai and Wang (2008). To this end we define

the long-run component $q(t)$ of the variance process (3b) as

$$q(t) = \frac{-\rho_1 \widetilde{w}}{(1-\rho_1)(\rho_2 - \rho_1)} + \frac{\rho_2}{\rho_2 - \rho_1} h(t) + \frac{\widetilde{b}_2}{\rho_2 - \rho_1} h(t-\Delta) + \frac{1}{\rho_2 - \rho_1} v_2(t-\Delta), \qquad (6)$$

where $v_2(t)$ is given by (4d), and where ρ_1 and ρ_2 are the smaller and larger roots of the quadratic equation

$$\rho^2 - \widetilde{b}_1 \rho - \widetilde{b}_2 = 0$$

respectively, which are the eigenvalues of the transition equation (1b). 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 (1a)-(1b) yields the IG-GARCH component model which we will denote IG-GARCH(C) below

$$\ln(S(t+\Delta)) = \ln(S(t)) + r + \mu h(t+\Delta) + \eta y(t+\Delta), \qquad (7a)$$

$$h(t+\Delta) = q(t+\Delta) + \rho_1(h(t) - q(t)) + v_h(t), \qquad (7b)$$

$$q(t+\Delta) = w_q + \rho_2 q(t) + v_q(t), \qquad (7c)$$

or equivalently,

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

where σ^2 is the unconditional variance, and

$$\sigma^2 = \frac{\widetilde{w}}{(1-\rho_1)(1-\rho_2)} \qquad w_q = \frac{\widetilde{w}}{1-\rho_1}$$

$$a_h = -\frac{\rho_1}{\rho_2 - \rho_1} a_1 - \frac{1}{\rho_2 - \rho_1} a_2 \qquad a_q = \frac{\rho_2}{\rho_2 - \rho_1} a_1 + \frac{1}{\rho_2 - \rho_1} a_2$$

$$c_h = -\frac{\rho_1}{\rho_2 - \rho_1} c_1 - \frac{1}{\rho_2 - \rho_1} c_2 \qquad c_q = \frac{\rho_2}{\rho_2 - \rho_1} c_1 + \frac{1}{\rho_2 - \rho_1} c_2$$

$$v_i(t) = c_i y(t) + a_i h(t)^2 / y(t) - c_i h(t) / \eta^2 - a_i \eta^2 h(t) - a_i \eta^4.$$

The unit root condition, $\rho_2 = 1$, corresponds to the restriction $\widetilde{b}_2 = 1 - \widetilde{b}_1$. The expression for σ^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

$$a_1 = a_h + a_q \qquad a_2 = -\rho_2 a_h - \rho_1 a_q$$

$$\widetilde{b}_1 = \rho_1 + \rho_2 \qquad \widetilde{b}_2 = -\rho_1 \rho_2$$

$$c_1 = c_h + c_q \qquad c_2 = -\rho_2 c_h - \rho_1 c_q$$

This proves that the IG-GARCH(2,2) model is equivalent to the component model (7a)-(7c). In the IG-GARCH(1,1) special case studied in Christoffersen, Heston and Jacobs (2006), the

long-run component in (7c) is effectively removed from the return dynamics.

The component structure helps interpreting the model. The coefficients of the lagged variables (long- or short-run component) 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 Gaussian Limit of the Component Model

While the IG-GARCH process in (1b) looks nonstandard, the Gaussian limit has more familiar dynamics. Consider the normalization of the innovation to the return process in (1a),

$$z(t) = \frac{\eta y(t) - h(t)/\eta}{\sqrt{h(t)}}. \tag{8}$$

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 η approaches zero, then the IG-GARCH(2,2) process (3a)-(3b) converges weakly to the Heston-Nandi (2000) GARCH(2,2) process:

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

where the limit is taken as follows

$$\begin{aligned} \tilde{w} &= \omega - \alpha_1 - \alpha_2, \\ a_i &= \alpha_i/\eta^4, \\ b_i &= \beta_i + \alpha_i\gamma_i^2 + 2\alpha_i\gamma_i/\eta - 2\alpha_i/\eta^2, \\ c_i &= \alpha_i(1 - 2\eta\gamma_i). \end{aligned}$$

Written in component form, the limit (keeping $\tilde{w}$ fixed) is

$$\begin{aligned} h(t+\Delta) &= q(t+\Delta) + \rho_1(h(t) - q(t)) + \nu_h(t), \\ q(t+\Delta) &= \omega_q + \rho_2 q(t) + \nu_q(t), \end{aligned} \tag{11a}$$

6

or equivalently,

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

where σ^2 is the unconditional variance, and

$$\alpha_h = -\frac{\rho_1}{\rho_2 - \rho_1}\alpha_1 - \frac{1}{\rho_2 - \rho_1}\alpha_2 \quad \alpha_q = \frac{\rho_2}{\rho_2 - \rho_1}\alpha_1 + \frac{1}{\rho_2 - \rho_1}\alpha_2,$$
$$\gamma_h = -\frac{\rho_1\alpha_1\gamma_1 + \alpha_2\gamma_2}{(\rho_2 - \rho_1)\alpha_h} \qquad \gamma_q = \frac{\rho_2\alpha_1\gamma_1 + \alpha_2\gamma_2}{(\rho_2 - \rho_1)\alpha_q}$$
$$\nu_i(t) = \alpha_i[(z(t) - \gamma_i\sqrt{h(t)})^2 - 1 - \gamma_i^2 h(t)],$$

where ρ_1 and ρ_2 are the respective smaller and larger roots of the quadratic equation

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

One can invert these component parameters to recover the GARCH(2,2) parameters

$$\alpha_1 = \alpha_h + \alpha_q \qquad \alpha_2 = -\rho_2\alpha_h - \rho_1\alpha_q$$
$$\gamma_1 = \frac{\alpha_h\gamma_h + \alpha_q\gamma_q}{\alpha_1} \qquad \gamma_2 = \frac{-\rho_2\alpha_h\gamma_h - \rho_1\alpha_q\gamma_q}{\alpha_2}$$
$$\beta_1 = \rho_1 + \rho_2 - \alpha_1\gamma_1^2 \quad \beta_2 = -\rho_1\rho_2 - \alpha_2\gamma_2^2.$$

Our inverse Gaussian component model in (7a)-(7c) thus corresponds in the limit to the component model of Christoffersen et al. (2008).

3   The Risk-Neutral Model and Option Valuation

To value options, we introduce the pricing kernel and the resulting risk-neutral IG-GARCH(2,2) process. We then elaborate on the relationships between the risk-neutral and physical parameters. Finally 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 Christoffersen, Heston, and Jacobs (2013), where

$$M(t + \Delta) = M(t)\left(\frac{S(t + \Delta)}{S(t)}\right)^\phi \exp(\delta_0 + \delta_1 h(t + \Delta) + \xi h(t + 2\Delta)). \tag{12}$$

Christoffersen, 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 refer we to it as the nonlinear pricing kernel. If $\xi > 0$, the pricing kernel is U-shaped in returns.

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

$$
\begin{aligned}
s_y &= 1 - 2c_1\xi - 2\eta\phi, \\
s_h &= \sqrt{1 - 2a_1\xi\eta^4 s_y^{-3/2}}.
\end{aligned}
\tag{13}
$$

Inserting these definitions into the IG-GARCH(2,2) dynamics in (1) yields the risk-neutral process

$$
\ln(S(t+\Delta)) = \ln(S(t)) + r + \mu^* h^*(t+\Delta) + \eta^* y^*(t+\Delta),
\tag{14}
$$
$$
h^*(t+\Delta) = w^* + b_1 h^*(t) + b_2 h^*(t-\Delta) + c_1^* y(t) + c_2^* y(t-\Delta)
$$
$$
+ a_1^* h^*(t)^2 / y^*(t) + a_2^* h^*(t-\Delta)^2 / y^*(t-\Delta),
$$

where

$$
h^*(t) = s_h h(t), \quad y^*(t) = s_y y(t),
\tag{15a}
$$
$$
\mu^* = \mu/s_h, \quad \eta^* = \eta/s_y, \quad w^* = s_h w,
\tag{15b}
$$
$$
a_i^* = s_y a_i / s_h, \quad c_i^* = s_h c_i / s_y.
\tag{15c}
$$

The risk-neutral process is IG-GARCH because the innovation $y^*(t+\Delta)$ has an Inverse Gaussian distribution under the risk-neutral probabilities. 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 B.

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

$$
\mu^* = \frac{\sqrt{1 - 2\eta^*} - 1}{\eta^{*2}}.
\tag{16}
$$

This imposes an equivalent restriction between the physical parameters μ and the preference parameters ϕ and ξ

$$
\mu = s_h \frac{\sqrt{1 - 2\eta/s_y} - 1}{\eta^2/s_y^2} = \sqrt{1 - 2a_1\xi\eta^4} \frac{\sqrt{1 - 2c_1\xi - 2\eta\phi - 2\eta} - \sqrt{1 - 2c_1\xi - 2\eta\phi}}{\eta^2}.
\tag{17}
$$

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

$$s_y = \frac{(\frac{1}{2}\mu^2\eta^4 + (1 - 2a_1\xi\eta^4)\eta)^2}{(1 - 2a_1\xi\eta^4)\mu^2\eta^4}, \tag{18}$$

$$s_h = \frac{\mu\eta^2}{s_y^2(\sqrt{1 - 2\eta/s_y} - 1)}. \tag{19}$$

Because s_h is now a function only of s_y and physical parameters, this demonstrates that we can write (15a)-(15c) as a function of the physical parameters and one additional reduced form parameter, namely either s_y or ξ.

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 η approaches zero, $\tilde{\mu}^* = \mu^* + \eta^{*-1}$ approaches $-\frac{1}{2}$. Also in this limit, $s_h = s_y^{-1}$ as seen from equation (13). 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 (6), 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. We implement this special case in our empirical study.

3.4 Option Valuation

Option valuation with this model is straightforward. Put options can be valued using put-call parity. Following Heston-Nandi (2000), the value of a call option at time t with strike price X maturing at T is equal to

$$Call(S(t), h(t + \Delta), X, T) = S(t) \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{X^{-i\varphi} g_t^*(i\varphi + 1, T)}{i\varphi} \right] d\varphi \right) \tag{22}$$
$$- X \exp^{-r(T-t)} \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{X^{-i\varphi} g_t^*(i\varphi, T)}{i\varphi} \right] d\varphi \right).$$

where $g_t^*(\varphi, T)$ is the conditional generating function for the risk-neutral process in (14). The conditional generating function $g_t(\varphi, T)$ is given by:

$$g_t(\varphi, T) = E_t[S(T)^\phi] = S(t)^\phi \exp(A(t) + B(t)h(t + \Delta) + C(t)q(t + \Delta)], \tag{23}$$

where

$$A(T) = B(T) = C(T) = 0, \tag{24a}$$

$$A(t) = A(t + \Delta) + \phi r + (w_q - a_h \eta^4 - a_q \eta^4) B(t + \Delta) + (w_q - a_q \eta^4) C(t + \Delta) \tag{24b}$$
$$- \frac{1}{2} \ln(1 - 2(a_h + a_q) \eta^4 B(t + \Delta) - 2a_q \eta^4 C(t + \Delta)),$$

$$B(t) = \phi \mu + (\rho_1 - (c_h + c_q) \eta^{-2} - (a_h + a_q) \eta^2) B(t + \Delta) - (c_q \eta^{-2} + a_q \eta^2) C(t + \Delta) + \eta^{-2} - \tag{24c}$$

$$\frac{\sqrt{(1 - 2(a_q + a_h) \eta^4 B(t + \Delta) - 2a_q \eta^4 C(t + \Delta))(1 - 2\eta \phi - 2(c_q + c_h) B(t + \Delta) - 2c_q C(t + \Delta))}}{\eta^2},$$

$$C(t) = (\rho_2 - \rho_1) B(t + \Delta) + \rho_2 C(t + \Delta). \tag{24d}$$

This recursive definition requires computing equations (24b-24d) day-by-day with the terminal condition in (24a) and then integrating $g_t(\varphi, T)$ as in (22). All the parameters in equations (24b-24d) are risk-neutral.

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 1, 1996 through December 31, 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 31, 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 31, 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.

Table 1 also presents 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%. 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 Sections 2 and 3. 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 date using the same, internally consistent, set of parameters.

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 then 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

$$
\begin{aligned}
f\left(R(t)|h(t)\right) \;=\; & \frac{h(t)\,|\eta|^{-3}}{\sqrt{2\pi[R(t)-r-\tilde{\mu}h(t)]^{3}\eta^{-3}}} \times \\
& \exp\left(-\frac{1}{2}\left(\sqrt{\frac{R(t)-r-\tilde{\mu}h(t)}{\eta}}-\frac{h(t)}{\eta^{2}}\sqrt{\frac{\eta}{R(t)-r-\tilde{\mu}h(t)}}\right)^{2}\right).
\end{aligned}
$$

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

$$
\ln L^{R} \propto -\sum_{t=1}^{T}\left\{\ln(f(R(t)|h(t)))\right\}. \tag{25}
$$

We can therefore obtain the physical parameters Θ by estimating

$$
\Theta_{Return} = \arg\max_{\Theta}\ln L^{R}. \tag{26}
$$

The estimation results are contained in Table 2.

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

$$
\varepsilon_{i} = \left(Call_{i}^{Mkt} - Call_{i}^{Mod}\right)/BSV_{i}^{Mkt},
$$

where $Call_{i}^{Mkt}$ represents the market price of the ith option, $Call_{i}^{Mod}$ represents the model price,

and BSV_i^{Mkt} 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

$$\ln L^O \propto -\frac{1}{2} \sum_{i=1}^{N} \left\{ \ln \left(s_\varepsilon^2 \right) + \varepsilon_i^2 / s_\varepsilon^2 \right\}. \tag{27}$$

where we can concentrate out s_ε^2 using the sample analogue $\hat{s}_\varepsilon^2 = \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i^2$. 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.[3] We the obtain the risk-neutral parameters Θ^* based on options data by estimating

$$\Theta_{Option}^* = \arg \max_{\Theta^*} \ln L^O. \tag{28}$$

The estimation results are contained in Table 3.

Note that both estimation exercises mentioned above ignore the specification of the pricing kernel, and are therefore uninformative about the choice between the linear and nonlinear pricing kernels. We thus conduct a third estimation exercise where we sequentially estimate the nonlinear pricing kernel parameter, ξ, on options only, keeping all the physical parameters from (26) fixed. We thus estimate

$$\xi_{Seq} = \arg \max_{\xi} \ln L^O. \tag{29}$$

The estimation results are contained in Table 4. Sequential estimation is of course only conducted for the models with nonlinear 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 nonlinear (or variance-dependent) pricing kernel. We can quantify the contribution of these features in separately explaining the time series of

---

[3]See for instance Carr and Wu (2007) and Trolle and Schwartz (2009) for applications of BSV^{Mkt} weighted option errors.

returns and the cross-section of option prices, as well as returns and options together, which we do in our 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 nonlinear 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 2-6 present the empirical results. Table 2 presents estimation results for the estimation 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 on first returns and then options. Table 4 also reports improvement in fit for the nonlinear 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, and two have no variance dynamics.

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 each case, the improvement provided by the second factor is dramatically higher, with a likelihood improvement exceeding $6,000$ in Table 3. This shows that the most important feature in accurately modelling option prices is correct specification of the volatility dynamics.

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 149.1, and for the fat-tailed case the corresponding number is 138.2. 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 η 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 any reasonable confidence level.

5.2 Sequential Estimation of the Nonlinear Pricing Kernel Parameter

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 nonlinear pricing kernel parameter ξ by fitting options. Table 4 reports risk-neutral values of all parameters, but only ξ is estimated from options.

The penultimate column in Panel B of Table 4 reports the option likelihoods for the four dynamic models with nonlinear pricing kernel. The last column in Panel B shows the difference between the option likelihood for optimal ξ 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 nonlinear 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 $6,822$ in the single factor Gaussian model, and $8,459$ in the corresponding Inverse Gaussian model. In case of the two-factor models, the improvements are even higher: The nonlinear kernel improves the two-factor likelihoods by $9,049$ in the Gaussian model and $11,169$ 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 nonlinear 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. To this extent, complex modelling of risk premia also 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 these features are not estimated in Table 4, 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 ξ 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 ξ parameter and indeed for a wide range of values around the optimum. In Figure 1 the linear pricing kernel corresponds to the left-most point on the curves where $\xi = 0$.

We have now compared model fit across three dimensions: One versus two volatility components, normal versus IG innovations, and linear versus nonlinear pricing kernels. Our results show that the data favors the model we suggest in this paper that has IG innovations, two variance components, and a nonlinear pricing kernel. Next we investigate the models' properties in more detail.

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 0.97 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 (ρ_2 is around 0.99), but the persistence of the short-run factor, ρ_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 plots 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 C.

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 variance in the component models jump to a higher level than do the GARCH(1,1) variances. Similarly, conditional standard deviation of variance (conditional correlation between returns and variance) of the component models in

Figure 3, is higher in level (higher in negative levels) 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

$$Var_t(T) = \partial^2 \ln g_t(\varphi, T)/\partial \varphi^2|_{\varphi=0}, \tag{30}$$

$$Skew_t(T) = \frac{\partial^3 \ln g_t(\varphi, T)/\partial \varphi^3|_{\varphi=0}}{(\partial^2 \ln g_t(\varphi, T)/\partial \varphi^2|_{\varphi=0})^{3/2}}, \tag{31}$$

$$Kurt_t(T) = \frac{\partial^4 \ln g_t(\varphi, T)/\partial \varphi^4|_{\varphi=0}}{(\partial^2 \ln g_t(\varphi, T)/\partial \varphi^2|_{\varphi=0})^2} - 3. \tag{32}$$

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.

Figure 4 highlights 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 hump-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 indicate that the effects of shocks last much longer in the component models.

Figure 5 and 6 repeat Figures 2 and 3 but now using the option-based parameters in Table 3 rather than the physical parameters in Table 2. The return-implied variance paths for the GARCH(1,1) and component models are very different in Figure 2, and the differences are substantial but less prominent in Figure 5 where we use the option-based estimates in Table 3.

Some important results can also 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.

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 in Table 2, Panel B uses the option-based estimates in Table 3, and Panel C uses the sequential estimates in 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 16.07% $(1 - 5.0723/6.0435)$ and 15.77% in Panel A, to 18.83% and 20.37% in Panel B, and finally 19.76% and 19.30% in Panel C. On average the improvement from adding a second volatility factor is 18.35%. The improvement from adding a second volatility factor is largest in Panels B and C which do not restrict the pricing kernel to be linear. 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 1.33% and 0.98% in Panel A, to 3.4% and 5.23% in Panel B, and 4.70% and 4.16% in Panel C. The overall improvement from adding fat tails is 3.3% 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 which do not restrict the pricing kernel to be linear. 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 14.01% $(1 - 5.1966/6.0435)$ for the GARCH(1,1) model, 16.95% 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 17.30%. 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

18

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 every moneyness category. This is not evident a priori because the models are fit to all options at once and not to individual 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 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 ξ 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 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 close to zero overall. 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 but one category of RMSE results which is short term options in Panel B. 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, and the improvement due to the second volatility component is actually more pronounced than the one due to the Inverse Gaussian innovation.

The increases in likelihood due to fat-tailed innovations are much smaller than those due to the second volatility factor. This observation from all of the estimations 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, as we do in returns-based estimation results, we 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 in 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) ϕ. In the nonlinear pricing kernel the computation of risk-aversion is slighly 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

$$RRA(t) \equiv -S(t)\frac{U''(S(t))}{U'(S(t))} = -S(t)\frac{M'(t)}{M(t)} = -S(t)\frac{\partial \ln(M(t))}{\partial S(t)}, \qquad (33)$$

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

From (12) we have that

$$\frac{\partial \ln\left(M\left(t\right)\right)}{\partial S\left(t\right)} = \frac{\phi}{S\left(t\right)} + \xi \frac{\partial h\left(t+\Delta\right)}{\partial S\left(t\right)}. \tag{34}$$

In the Gaussian model we have

$$\frac{\partial h\left(t+\Delta\right)}{\partial S\left(t\right)} = \frac{\partial h\left(t+\Delta\right)}{\partial z\left(t\right)} \frac{\partial z\left(t\right)}{\partial S\left(t\right)} = \frac{2\alpha_1\left(z(t) - \gamma_1\sqrt{h(t)}\right)}{\sqrt{h(t)}S\left(t\right)}. \tag{35}$$

Combining (34) and (35) we get a relative risk aversion of

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

Note as indicated above that the parameter ϕ does not in inself 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

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

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

$$\phi = -\left(\tilde{\mu} + \gamma_1\right)\left(1 - 2\alpha_1\xi\right) + \gamma_1 - \frac{1}{2} \approx 33.56,$$

so that we get

$$E\left[RRA\left(t\right)\right] \approx -33.56 + 2\alpha_1\xi\gamma_1 \approx 1.36.$$

This result shows that the nonlinear pricing kernel delivers reasonable coefficients of relative risk aversion, and furthermore that it is important not to rely on (the negative of) ϕ as a measure of RRA when using the nonlinear 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 second volatility factor is economically most important and improves option fit by 18% on average. A U-shaped pricing kernel improves the option fit

by 17% and more so for two-factor models. Fat tails improve option fit by just over 3% on average and more so when a U-shaped pricing kernel is applied. Our results suggest that the three features we investigate 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 important to evaluate them in a properly specified model that nests all these features.

Appendix A: Martingale Restrictions

6.1 Restriction with risk-free rate

Using bond prices are martingale,

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

where $B_\tau(t)$ is a bond with maturity τ 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 $\delta(t+\Delta) \equiv \delta_0 + \delta_1 h(t+\Delta)$. We assume that $B_\tau(t+\Delta)/B_\tau(t) \equiv \exp(r_f)$.

$$
\begin{aligned}
1 &= E_t\left[\exp(\ln\frac{M(t+\Delta)}{M(t)} + \ln\frac{B_\tau(t+\Delta)}{B_\tau(t)})\right] = E_t\left[\exp(\ln\frac{M(t+\Delta)}{M(t)} + r_f)\right] \\
&= E_t\left[\exp(\phi r(t+\Delta) + \delta_0 + \delta_1 h(t) + \xi h(t+2\Delta) + r_f)\right] \\
&= E_t\left[\exp\left(\begin{array}{c} \delta_0 + \delta_1 h(t) + \phi(r_f + \mu h(t+\Delta) + \eta y(t+\Delta)) + \\ \xi\left(w + b_1 h(t+\Delta) + c_1 y(t+\Delta) + a_1 h(t+\Delta)^2/y(t+\Delta)\right) + r_f \end{array}\right)\right] \\
&= \exp\left(\delta_0 + \delta_1 h(t) + \phi\mu h(t+\Delta) + (1+\phi)r_f + \xi(w + b_1 h(t+\Delta))\right) \\
&\quad \times E_t\left[\exp\left((\phi\eta + \xi c_1)y(t+\Delta) + \xi a_1 h(t+\Delta)^2/y(t+\Delta)\right)\right] \\
&= \exp\left(\delta_0 + \delta_1 h(t) + \phi\mu h(t+\Delta) + (1+\phi)r_f + \xi(w + b_1 h(t+\Delta))\right) \\
&\quad \times \exp\left(-\frac{1}{2}\ln(1 - 2\xi a_1\eta_4) + h(t+\Delta)/\eta^2 - h(t+\Delta)/\eta^2\sqrt{(1-2\xi a_1\eta_4)(1-2(\phi\eta+\xi c_1))}\right) \\
0 &= \delta_0 + \delta_1 h(t) + \phi\mu h(t+\Delta) + (1+\phi)r_f + \xi(w + b_1 h(t+\Delta)) \\
&\quad -\frac{1}{2}\ln(1 - 2\xi a_1\eta_4) + h(t+\Delta)/\eta^2 - h(t+\Delta)/\eta^2\sqrt{(1-2\xi a_1\eta_4)(1-2(\phi\eta+\xi c_1))} \\
0 &= \left[\delta_0 + (1+\phi)r_f + \xi w - \frac{1}{2}\ln(1 - 2\xi a_1\eta_4)\right] \\
&\quad + \left[\delta_1 + \phi\mu + \xi b_1 + \eta^{-2} - \eta^{-2}\sqrt{(1-2\xi a_1\eta_4)(1-2(\phi\eta+\xi c_1))}\right]h(t+\Delta)
\end{aligned}
$$

Since both terms above should be equal to zero in order to satisfy the equation,

$$\delta_0 = -(1+\phi)r_f - \xi w + \frac{1}{2}\ln(1 - 2\xi a_1\eta_4) \tag{36}$$

$$\delta_1 = -\phi\mu - \xi b_1 - \eta^{-2}\left(1 - \sqrt{(1-2\xi a_1\eta_4)(1-2(\phi\eta+\xi c_1))}\right) \tag{37}$$

6.2   Restriction with market returns

Using stock prices are martingale,

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

where $M(t+\Delta)/M(t)=(S(t+\Delta)/S(t))^{\phi}\exp(\delta(t+\Delta)+\xi h(t+2\Delta))$ and $\delta(t+\Delta)\equiv\delta_0+\delta_1 h(t+\Delta)$.

$$1=E_t\left[\exp(\ln\frac{M(t+\Delta)}{M(t)}+\ln\frac{S(t+\Delta)}{S(t)})\right]=E_t\left[\exp(\ln\frac{M(t+\Delta)}{M(t)}+r(t+\Delta))\right]$$

$$=E_t\left[\exp(\phi r(t+\Delta)+\delta_0+\delta_1 h(t+\Delta)+\xi h(t+2\Delta)+r(t+\Delta))\right]$$

$$=E_t\left[\exp\left(\begin{array}{c}\delta_0+\delta_1 h(t+\Delta)+(1+\phi)(r_f+\mu h(t+\Delta)+\eta y(t+\Delta))\\+\xi\left(w+b_1 h(t+\Delta)+c_1 y(t+\Delta)+a_1 h(t+\Delta)^2/y(t+\Delta)\right)\end{array}\right)\right]$$

$$=\exp\left(\delta_0+\delta_1 h(t+\Delta)+(1+\phi)(r_f+\mu h(t+\Delta))+\xi(w+b_1 h(t+\Delta))\right)$$

$$\times E_t\left[\exp\left((\eta+\phi\eta+\xi c_1)y(t+\Delta)+\xi a_1 h(t+\Delta)^2/y(t+\Delta)\right)\right]$$

$$=\exp\left(\begin{array}{c}-(1+\phi)r_f-\xi w+\frac{1}{2}\ln(1-2\xi a_1\eta_4)\\+(-\phi\mu-\xi b_1-\eta^{-2}[1-\sqrt{(1-2\xi a_1\eta_4)(1-2(\phi\eta+\xi c_1))}])h(t+\Delta)\end{array}\right)$$

$$\times\exp\left((1+\phi)(r_f+\mu h(t+\Delta))+\xi(w+b_1 h(t+\Delta))\right)$$

$$\times\exp\left(-\frac{1}{2}\ln(1-2\xi a_1\eta_4)+h(t+\Delta)/\eta^2-h(t+\Delta)/\eta^2\sqrt{(1-2\xi a_1\eta_4)(1-2(\eta+\phi\eta+\xi c_1))}\right)$$

$$=\exp\left(\begin{array}{c}\mu h(t+\Delta)+h(t+\Delta)/\eta^2\sqrt{(1-2\xi a_1\eta_4)(1-2(\phi\eta+\xi c_1))}\\-h(t+\Delta)/\eta^2\sqrt{(1-2\xi a_1\eta_4)(1-2(\eta+\phi\eta+\xi c_1))}\end{array}\right).$$

Therefore we can write μ in terms of other parameters,

$$\mu=\eta^{-2}\sqrt{(1-2\xi a_1\eta_4)}\left[\sqrt{1-2(\eta+\phi\eta+\xi c_1)}-\sqrt{1-2(\phi\eta+\xi c_1)}\right].$$

7   The Inverse Gaussian Risk-Neutral Distribution

The log-pricing kernel is,

$$\ln \frac{M(t+\Delta)}{M(t)} = \delta_0 + \delta_1 h(t+\Delta) + \phi \ln(S(t+\Delta)/S(t)) + \xi h(t+2\Delta) \tag{38}$$

$$= \delta_0 + \delta_1 h(t+\Delta) + \phi(r_f + \mu h(t+\Delta) + \eta y(t+\Delta))$$
$$+ \xi(w + bh(t+\Delta) + cy_{t+1} + ah(t+\Delta)^2/y(t+\Delta))$$
$$= \delta_0 + \delta_1 h(t+\Delta) + \phi r_f + \xi w + (\phi\mu + \xi b)h(t+\Delta)$$
$$+ (\phi\eta + \xi c)y(t+\Delta) + \xi ah(t+\Delta)^2/y(t+\Delta)$$

where δ_0 and δ_1 are,

$$\delta_0 = -(1+\phi)r_f - \xi w + \frac{1}{2}\ln(1 - 2\xi a_1 \eta_4)$$
$$\delta_1 = -\phi\mu - \xi b_1 - \eta^{-2}\left(1 - \sqrt{(1 - 2\xi a_1 \eta_4)(1 - 2(\phi\eta + \xi c_1))}\right).$$

Also, remind that

$$E_t\left[\exp\left(\alpha y(t+\Delta) + \beta/y(t+\Delta)\right)\right] = \frac{h(t+\Delta)/\eta^2}{\sqrt{h(t+\Delta)^2/\eta^4 - 2\beta}} \tag{39}$$
$$\times \exp\left(h(t+\Delta)/\eta^2 - \sqrt{(h(t+\Delta)^2/\eta^4 - 2\beta)(1 - 2\alpha)}\right)$$
$$= \frac{1}{\sqrt{1 - 2\beta h(t+\Delta)^{-2}\eta^4}}$$
$$\times \exp\left[h(t+\Delta)/\eta^2\left(1 - \sqrt{(1 - 2\beta h(t+\Delta)^{-2}\eta^4)(1 - 2\alpha)}\right)\right].$$

We can derive the MGF under risk-neutral measure as the following,

$$E_t^Q\left(e^{xy^*(t+\Delta)}\right) = E_t\left[\exp(r_f)\frac{M(t+\Delta)}{M(t)}\exp(xs_y y(t+\Delta))\right]$$

$$= E_t\left[\exp(r_f + \delta_0 + \delta_1 h(t+\Delta) + \phi\ln\frac{S(t+\Delta)}{S(t)} + \xi h(t+2\Delta) + xs_y y(t+\Delta))\right]$$

$$= \exp\left(r_f + \delta_0 + \phi r_f + \xi w + (\delta_1 + \phi\mu + \xi b)h(t+\Delta)\right)$$

$$\times E_t\left[\exp((\phi\eta + \xi c + xs_y)y(t+\Delta) + \xi a h(t+\Delta)^2/y(t+\Delta))\right]$$

$$= \exp\left[h(t+\Delta)/\eta^2\left(\begin{array}{c}\sqrt{(1-2\xi a\eta^4)(1-2(\phi\eta+\xi c_1))}\\ -\sqrt{(1-2\xi a\eta^4)(1-2\phi\eta-2\xi c-2xs_y)}\end{array}\right)\right]$$

$$= \exp\left(\begin{array}{c}h(t+\Delta)/\eta^2\sqrt{(1-2\xi a\eta^4)(1-2\phi\eta-2\xi c)}\\ \times\left[1-\sqrt{\frac{1-2\phi\eta-2\xi c-2xs_y}{1-2\phi\eta-2\xi c}}\right]\end{array}\right)$$

$$= \exp\left(h(t+\Delta)/\eta^2\sqrt{s_1 s_2}\left[1-\sqrt{1-2x\frac{s_y}{s_2}}\right]\right)$$

where $y^*(t) \equiv y(t)s_y$, $s_1 = 1 - 2\xi a\eta^4$ and $s_2 = 1 - 2\phi\eta - 2\xi c$. If we take the derivative of a MGF and evaluate it at 0, we get the expected value of the random variable,

$$E_{t-\Delta}^Q[y^*(t)] = \frac{dE_{t-\Delta}^Q[e^{xy^*(t)}]}{dx}\bigg|_{x=0} = \frac{d\exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1-\sqrt{1-2x\frac{s_y}{s_2}}\right]\right)}{dx}\bigg|_{x=0}$$

$$= \exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1-\sqrt{1-2x\frac{s_y}{s_2}}\right]\right)\frac{d\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1-\sqrt{1-2x\frac{s_y}{s_2}}\right]\right)}{dx}\bigg|_{x=0}$$

$$= \exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1-\sqrt{1-2x\frac{s_y}{s_2}}\right]\right)$$

$$\times \frac{d\left(h(t)/\eta^2\sqrt{s_1 s_2} - \sqrt{h(t)^2/\eta^4 s_1 s_2 - 2xh(t)^2/\eta^4 s_1 s_y}\right)}{dx}\bigg|_{x=0}$$

$$= \exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1-\sqrt{1-2x\frac{s_y}{s_2}}\right]\right)(1-2x)^{-1/2}h(t)/\eta^2\sqrt{s_1 s_2}\bigg|_{x=0}$$

$$= h(t)/\eta^2\sqrt{s_1 s_2}.$$

If we take the second derivative of a MGF, evaluate it at 0 and adjust with squared first moment,

then we get variance under the risk-neutral measure,

$$
\begin{aligned}
Var^Q_{t-\Delta}[y^*(t)] &= \frac{d^2 E^Q_{t-\Delta}[e^{xy^*(t)}]}{dx^2}\bigg|_{x=0} - \left(\frac{dE^Q_{t-\Delta}[e^{xy^*(t)}]}{dx}\bigg|_{x=0}\right)^2 \\
&= \frac{d\left[\exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1 - \sqrt{1 - 2x\frac{s_y}{s_2}}\right]\right)(1-2x)^{-1/2}h(t)/\eta^2\sqrt{s_1 s_2}\right]}{dx}\bigg|_{x=0} \\
&\quad - h(t)^2/\eta^4 s_1 s_2 \\
&= h(t)/\eta^2\sqrt{s_1 s_2}(1-2x)^{-1/2}\frac{d\left[\exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1 - \sqrt{1 - 2x\frac{s_y}{s_2}}\right]\right)\right]}{dx}\bigg|_{x=0} \\
&\quad + h(t)/\eta^2\sqrt{s_1 s_2}\exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1 - \sqrt{1 - 2x\frac{s_y}{s_2}}\right]\right)\frac{d(1-2x)^{-1/2}}{dx}\bigg|_{x=0} \\
&\quad - h(t)^2/\eta^4 s_1 s_2 \\
&= h(t)/\eta^2\sqrt{s_1 s_2}\exp\left(h(t)/\eta^2\sqrt{s_1 s_2}\left[1 - \sqrt{1 - 2x\frac{s_y}{s_2}}\right]\right)(1-2x)^{-3/2}\bigg|_{x=0} \\
&= h(t)/\eta^2\sqrt{s_1 s_2}. \tag{40}
\end{aligned}
$$

Remind the variance process, and define risk-neutral variance as $h(t+\Delta)^* \equiv h(t+\Delta)s_h$ and use previously defined $y^*(t) \equiv y(t)s_y$,

$$
\begin{aligned}
h(t+\Delta) &= w + bh(t) + cy(t) + ah(t)^2/y(t) \\
h(t+\Delta)s_h &= ws_h + bh(t)s_h + cy(t)s_h + ah(t)^2 s_h/y(t) \\
h^*(t+\Delta) &= ws_h + bh^*(t) + cy^*(t)s_h/s_y + a\frac{s_y}{s_h}(h^*(t))^2/y^*(t),
\end{aligned}
$$

and the returns,

$$
\log(S(t+\Delta)) = \log(S(t)) + r + \mu h^*(t+\Delta)/s_h + \eta y^*(t+\Delta)/s_y.
$$

27

Therefore,

$$\mu^* = \mu/s_h$$
$$\eta^* = \eta/s_y$$
$$w^* = w s_h$$
$$b^* = b$$
$$a^* = a s_y/s_h$$
$$c^* = c s_h/s_y$$
$$h^*(t)/(\eta^*)^2 = (h(t)/\eta^2)s_h s_y^2,$$

where the last equation should also equals eq (40), and implies that

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

Now consider the probability density function of the stock price, $S(t)$,

$$f_{t-\Delta}(S(t)) = f_{t-\Delta}(y(t))\left|\frac{\partial y(t)}{\partial S(t)}\right| \tag{41}$$

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

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

To find the risk-neutral dynamic, note that the risk-neutral density is the product of the physical density and the pricing kernel properly normalized as follows

$$\frac{M(t)}{M(t - \Delta)} = \exp(-r_f)\frac{f^*_{t-\Delta}(S(t))}{f_{t-\Delta}(S(t))}$$

$$f^*_{t-\Delta}(S(t)) = f_{t-\Delta}(S(t))\exp(r_f)M(t)/M(t - \Delta) \tag{42}$$

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

Using the pricing kernel definition in (12) and the IG-GARCH(1,1) return dynamic, we can

write

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

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

$$\times \exp\left[r_f + \delta_0 + \delta_1 h(t) + \phi \ln(S(t)/S(t-\Delta)) + \xi h(t+2\Delta)\right]$$

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

$$\times \exp\left[r_f + \delta_0 + \phi r_f + \xi w + (\delta_1 + \phi\mu + \xi b)h(t) + (\phi\eta + \xi c)y(t) + \xi a h(t)^2/y(t)\right]$$

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

$$\times \exp\left[h(t)/\eta^2\sqrt{(1-2\xi a_1\eta^4)(1-2\phi\eta-2\xi c_1)}\right]$$

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

Substituting the physical distribution from equation (41) and rearranging terms yields

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

$$\exp\left[-\frac{1}{2}\left(\sqrt{y(t)(1-2\phi\eta-2\xi c)} - \frac{h(t)/\eta^2\sqrt{(1-2\xi a\eta^4)(1-2\phi\eta-2\xi c)}}{\sqrt{y(t)(1-2\phi\eta-2\xi c)}}\right)^2\right]$$

$$= \frac{h(t)\sqrt{(1-2\xi a\eta^4)(1-2\phi\eta-2\xi c)^{-3}}(1-2\phi\eta-2\xi c)^3/\eta^3}{\sqrt{2\pi y(t)^3(1-2\phi\eta-2\xi c)^3}S(t)} \times$$

$$\exp\left[-\frac{1}{2}\left(\frac{\sqrt{y(t)(1-2\phi\eta-2\xi c)}}{-\frac{h(t)\sqrt{(1-2\xi a\eta^4)(1-2\phi\eta-2\xi c)^{-3}}(1-2\phi\eta-2\xi c)^2/\eta^2}{\sqrt{y(t)(1-2\phi\eta-2\xi c)}}}\right)^2\right].$$

This enables us to define the risk-neutral counterparts to $y(t)$, $h(t)$, and η 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 implicitly defined

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

as in the text. Using these definitions 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,

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

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

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

Appendix B: The Risk-Neutral Component Model

The component representation of the risk-neutral process (14) is given by

$$
\begin{aligned}
\ln(S(t+\Delta)) &= \ln(S(t)) + r + \tilde{\mu}^* h(t+\Delta) + (\eta^* y^*(t+\Delta) - h^*(t+\Delta)/\eta^*), \\
h^*(t+\Delta) &= q^*(t+\Delta) + \rho_1^*(h^*(t) - q^*(t)) + \nu_h^*(t), \\
q^*(t+\Delta) &= \sigma^{*2} + \rho_2^*(q^*(t) - \sigma^{*2}) + \nu_q^*(t),
\end{aligned}
$$

where

$$q^*\left(t\right) = \frac{-\rho_1^*\widetilde{w}^*}{(1-\rho_1^*)(\rho_2^*-\rho_1^*)} + \frac{\rho_2^*}{\rho_2^*-\rho_1^*}h^*(t) + \frac{\widetilde{b}_2^*}{\rho_2^*-\rho_1^*}h^*(t-\Delta) + \frac{1}{\rho_2^*-\rho_1^*}\upsilon_2^*(t-\Delta),$$

$$\widetilde{\mu}^* = \mu^* + \eta^{*-1} = \mu/s_h + s_y\eta^{-1},$$

$$\sigma^{*2} = \left(\frac{\rho_2^*}{1-\rho_2^*} - \frac{\rho_1^*}{1-\rho_1^*}\right)\frac{\widetilde{w}^*}{\rho_2^*-\rho_1^*},$$

$$\widetilde{w}^* = w^* + a_1^*\eta^{*4} + a_2^*\eta^{*4} = s_h w + \frac{a_1\eta^4 + a_2\eta^4}{s_h s_y^3},$$

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

$$\upsilon_q^*(t) = c_q^* y^*(t) + a_q^* h^*(t)^2/y^*(t) - c_q^* h^*(t)/\eta^{*2} - a_q^*\eta^{*2}h^*(t) - a_q^*\eta^{*4},$$

$$a_h^* = -\frac{\rho_1^*}{\rho_2^*-\rho_1^*}a_1^* - \frac{1}{\rho_2^*-\rho_1^*}a_2^*,$$

$$c_h^* = -\frac{\rho_1^*}{\rho_2^*-\rho_1^*}c_1^* - \frac{1}{\rho_2^*-\rho_1^*}c_2^*,$$

$$a_q^* = \frac{\rho_2^*}{\rho_2^*-\rho_1^*}a_1^* + \frac{1}{\rho_2^*-\rho_1^*}a_2^*,$$

$$c_q^* = \frac{\rho_2^*}{\rho_2^*-\rho_1^*}c_1^* + \frac{1}{\rho_2^*-\rho_1^*}c_2^*,$$

$$\widetilde{b}_i^* = b_i + s_h s_y c_i/\eta^2 + \frac{a_i\eta^2}{s_h s_y},$$

and where ρ_1^* and ρ_2^* are the smaller and larger respective roots of the equation $\rho^{*2} - \widetilde{b}_1^*\rho - \widetilde{b}_2^* = 0$.

Appendix C: 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

$$h(t+\Delta) = q(t+\Delta) + \rho_1\left[h(t) - q(t)\right] + v_h(t)$$
$$q(t+\Delta) = w_q + \rho_2 q(t) + v_q(t)$$
$$v_h(t) = c_h\left[y(t) - \delta(t)\right] + a_h h(t)^2\left[1/y(t) - 1/\delta(t) - 1/\delta(t)^2\right]$$
$$v_q(t) = c_q\left[y(t) - \delta(t)\right] + a_q h(t)^2\left[1/y(t) - 1/\delta(t) - 1/\delta(t)^2\right].$$

Conditional variance of variance is given by

$$
\begin{aligned}
h(t+2\Delta) - E_t[h(t+2\Delta)] &= q(t+2\Delta) + \rho_1\left[h(t+\Delta) - q(t+\Delta)\right] + v_h(t+\Delta)\\
&\quad - E_t\left[q(t+2\Delta) + \rho_1\left[h(t+\Delta) - q(t+\Delta)\right] + v_h(t+\Delta)\right]\\
&= w_q + \rho_2 q(t) + v_q(t) + v_h(t+\Delta) - E_t[w_q + \rho_2 q(t) + v_q(t)]\\
&= v_q(t) + v_h(t+\Delta)\\
&= (c_h + c_q)\left[y(t+\Delta) - \delta(t+\Delta)\right]\\
&\quad + (a_h + a_q)h(t+\Delta)^2\left[1/y(t+\Delta) - 1/\delta(t+\Delta) - 1/\delta(t+\Delta)^2\right]
\end{aligned}
$$

$$Var_t[h(t+2\Delta)] = E_t \left[\begin{array}{c} (c_h + c_q)\left[y(t+\Delta) - \delta(t+\Delta)\right] \\ +(a_h + a_q)h(t+\Delta)^2\left[1/y(t+\Delta) - 1/\delta(t+\Delta) - 1/\delta(t+\Delta)^2\right]^2 \end{array} \right]$$

$$= (c_h + c_q)^2 E_t\left[(y(t+\Delta) - \delta(t+\Delta))^2\right]$$

$$+ 2(a_h + a_q)(c_h + c_q)h(t+\Delta)^2 E_t \left[\begin{array}{c} (y(t+\Delta) - \delta(t+\Delta))\times \\ (1/y(t+\Delta) - 1/\delta(t+\Delta) - 1/\delta(t+\Delta)^2) \end{array} \right]$$

$$+ (a_h + a_q)^2 h(t+\Delta)^4 E_t\left[(1/y(t+\Delta) - 1/\delta(t+\Delta) - 1/\delta(t+\Delta)^2)^2\right]$$

$$= (c_h + c_q)^2 Var_t[y(t+\Delta)] + 2(a_h + a_q)(c_h + c_q)h(t)^2 Cov_t[y(t+\Delta), 1/y(t+\Delta)]$$

$$+ (a_h + a_q)^2 h(t)^4 Var_t[1/y(t+\Delta)]$$

$$= (c_h + c_q)^2 \delta(t+\Delta) - 2(a_h + a_q)(c_h + c_q)h(t)^2/\delta(t+\Delta)$$

$$+ (a_h + a_q)^2 h(t)^4 (1/\delta(t+\Delta)^3 + 2/\delta(t+\Delta)^4)$$

$$= (c_h + c_q)^2 h(t+\Delta)/\eta^2 - 2(a_h + a_q)(c_h + c_q)\eta^2 h(t)$$

$$+ (a_h + a_q)^2 \eta^6 h(t) + 2(a_h + a_q)^2 \eta^8.$$

We thus can write

$$Std_t[h(t+2\Delta)] = \sqrt{2(a_h + a_q)^2 \eta^8 + \left[(c_h + c_q)/\eta - (a_h + a_q)\eta^3\right]^2 h(t+\Delta)}$$

Consider now the innovation to returns

$$R(t+\Delta) - E_t[R(t+\Delta)] = r + (\mu + 1/\eta)h(t+\Delta) + \eta(y(t+\Delta) - h(t+\Delta)/\eta^2)$$
$$+ E_t[r + (\mu + 1/\eta)h(t+\Delta) + \eta(y(t+\Delta) - h(t+\Delta)/\eta^2)],$$

so that

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

We can now compute

$$Cov_t[R(t+\Delta), h(t+2\Delta)] = E_t[\eta(y(t+\Delta) - \delta(t+\Delta))((c_h + c_q)(y(t+\Delta) - \delta(t+\Delta))$$
$$+ (a_h + a_q)h(t+\Delta)^2(1/y(t+\Delta) - 1/\delta(t+\Delta) - 1/\delta(t+\Delta)^2))]$$

$$= (c_h + c_q)\eta E_t\left[(y(t+\Delta) - \delta(t+\Delta))^2\right]$$

$$+ (a_h + a_q)\eta h(t+\Delta)^2 \times$$

$$E_t\left[(y(t+\Delta) - \delta(t+\Delta))(1/y(t+\Delta) - 1/\delta(t+\Delta) - 1/\delta(t+\Delta)^2)\right]$$

Taking expectations yields

$$
\begin{aligned}
Cov_t[R(t+\Delta), h(t+2\Delta)] &= (c_h + c_q)\eta Var_t[y(t+\Delta)] \\
&\quad + (a_h + a_q)\eta h(t+\Delta)^2 Cov_t[y(t+\Delta), 1/y(t+\Delta)] \\
&= (c_h + c_q)/\eta h(t+\Delta) - (a_h + a_q)\eta^3 h(t+\Delta) \\
Corr_t[R(t+\Delta), h(t+2\Delta)] &\equiv \frac{Cov_t[R(t+\Delta), h(t+2\Delta)]}{\sqrt{Var_t[R(t+\Delta)]Var_t[h(t+2\Delta)]}} \\
&= \frac{((c_h + c_q)/\eta - (a_h + a_q)\eta^3)h(t+\Delta)}{\sqrt{h(t+\Delta)[2(a_h + a_q)^2\eta^8 + (c/\eta - (a_h + a_q)\eta^3)^2 h(t+\Delta)]}} \\
&= \frac{[(c_h + c_q)/\eta - (a_h + a_q)\eta^3]\sqrt{h(t+\Delta)}}{\sqrt{2(a_h + a_q)^2\eta^8 + [(c_h + c_q)/\eta - (a_h + a_q)\eta^3]^2 h(t+\Delta)}}.
\end{aligned}
$$

From this we can define the desired leverage correlation

$$
\begin{aligned}
Corr_t[R(t+\Delta), h(t+2\Delta)] &\equiv \frac{Cov_t[R(t+\Delta), h(t+2\Delta)]}{\sqrt{Var_t[R(t+\Delta)]Var_t[h(t+2\Delta)]}} \\
&= \frac{((c_h + c_q)/\eta - (a_h + a_q)\eta^3)h(t+\Delta)}{\sqrt{h(t+\Delta)[2(a_h + a_q)^2\eta^8 + (c/\eta - (a_h + a_q)\eta^3)^2 h(t+\Delta)]}} \\
&= \frac{[(c_h + c_q)/\eta - (a_h + a_q)\eta^3]\sqrt{h(t+\Delta)}}{\sqrt{2(a_h + a_q)^2\eta^8 + [(c_h + c_q)/\eta - (a_h + a_q)\eta^3]^2 h(t+\Delta)}}.
\end{aligned}
$$

References

[1] Ait-Sahalia, Y., and A. Lo (1998), Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, Journal of Finance, 53, 499-547.

[2] Andersen, T.G., L. Benzoni, and J. Lund (2002), Estimating Jump-Diffusions for Equity Returns, Journal of Finance, 57, 1239-1284.

[3] Bakshi, G., C. Cao, and Z. Chen (1997), Empirical Performance of Alternative Option Pricing Models, Journal of Finance, 52, 2003-2049.

[4] Bakshi, G., D. Madan, and G. Panayotov (2010), Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels, Journal of Financial Economics, 97, 130-154.

[5] Barone-Adesi, G., R. F. Engle, and L. Mancini (2008), A GARCH Option Pricing Model with Filtered Historical Simulation, Review of Financial Studies, 21, 1223-1258.

[6] Bates, D. (1996a), Testing Option Pricing Models, in Handbook of Statistics, Statistical Methods in Finance, G.S. Maddala and C.R. Rao (eds.), 567-611. Amsterdam: Elsevier.

[7] Bates, D. (1996b), Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, Review of Financial Studies, 9, 69-107.

[8] Bates, D. (2000), Post-'87 Crash Fears in the S&P 500 Futures Option Market, Journal of Econometrics, 94, 181-238.

[9] Black, F., and M. Scholes (1973), The Pricing of Options and Corporate Liabilities, Journal of Political Economy, 81, 637-659.

[10] Brennan, M. (1979), The Pricing of Contingent Claims in Discrete-Time Models, Journal of Finance, 34, 53-68.

[11] Broadie, M., M. Chernov, and M. Johannes (2007), Model Specification and Risk Premiums: Evidence from Futures Options, Journal of Finance, 62, 1453-1490.

[12] Brown, D., and J. Jackwerth (2012), The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory, Contemporary Studies in Economic and Financial Analysis, 94, 155-183.

[13] Carr, P., and L. Wu (2007), Stochastic Skew in Currency Options, Journal of Financial Economics, 86, 213-247.

[14] Chabi-Yo, F., (2012), Pricing Kernels with Stochastic Skewness and Volatility Risk, Management Science, 58, 624-640.

[15] Chernov, M., R. Gallant, E. Ghysels, and G. Tauchen (2003), Alternative Models for Stock Price Dynamics, Journal of Econometrics, 116, 225–257.

[16] Chernov, M., and E. Ghysels (2000), A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation, Journal of Financial Economics, 56, 407-458.

[17] Christoffersen, P., S. Heston, and K. Jacobs (2006), Option Valuation with Conditional Skewness, Journal of Econometrics, 131, 253-284.

[18] Christoffersen, P., S. Heston, and K. Jacobs (2009), The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well, Management Science, 55, 1914-1932.

[19] Christoffersen, P., S. Heston, and K. Jacobs (2013), Capturing Option Anomalies with a Variance-Dependent Pricing Kernel, Review of Financial Studies, 26, 1963-2006.

[20] Christoffersen, P., K. Jacobs, C. Ornthanalai, and Y. Wang (2008), Option Valuation with Long-Run and Short-Run Volatility Components, Journal of Financial Economics, 90, 272-297.

[21] Duffie, D., J. Pan, and K. Singleton (2000), Transform Analysis and Asset Pricing for Affine Jump-diffusions, Econometrica, 68, 1343-1376.

[22] Dumas, B., J. Fleming, and R. Whaley (1998), Implied Volatility Functions: Empirical Tests, Journal of Finance, 53, 2059-2106.

[23] Engle, R., and G. Lee (1999), A Long-Run and Short-Run Component Model of Stock Return Volatility, in Cointegration, Causality and Forecasting: Festschrift in Honour of Clive W. J. Granger, edited by R. Engle and H. White, Oxford University Press.

[24] Eraker, B. (2004), Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, Journal of Finance, 59, 1367-1403.

[25] Heston, S. (1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, Review of Financial Studies, 6, 327-343.

[26] Heston, S., and S. Nandi (2000), A Closed-Form GARCH Option Pricing Model, Review of Financial Studies, 13, 585-626.

[27] Jackwerth, J. (2000), Recovering Risk Aversion from Option Prices and Realized Returns. Review of Financial Studies, 13, 433–51.

[28] Jones, C. (2003), The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets, Journal of Econometrics, 116, 181-224.

[29] Liu, J., J. Pan, and T. Wang (2004), An Equilibrium Model of Rare Event Premia, Review of Financial Studies, 18, 131-164.

[30] Pan, J. (2002), The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, Journal of Financial Economics, 63, 3-50.

[31] Rosenberg, J. and R. Engle (2002), Empirical Pricing Kernels, Journal of Financial Economics 64, 341-372.

[32] Rubinstein, M. (1976), The Valuation of Uncertain Income Streams and the Pricing of Options, Bell Journal of Economics, 7, 407-425.

[33] Shive, S., and T. Shumway (2006), Is the Pricing Kernel Monotonic?, Working Paper, Notre Dame University.

[34] Taylor, S., and X. Xu (1994), The Term Structure of Volatility Implied by Foreign Exchange Options, Journal of Financial and Quantitative Analysis, 29, 57–74.

[35] Trolle, A., and E. Schwartz (2009), Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives, Review of Financial Studies, 22, 4423-4461.

Figure 1. RMSE and Option Likelihood Values versus ξ.
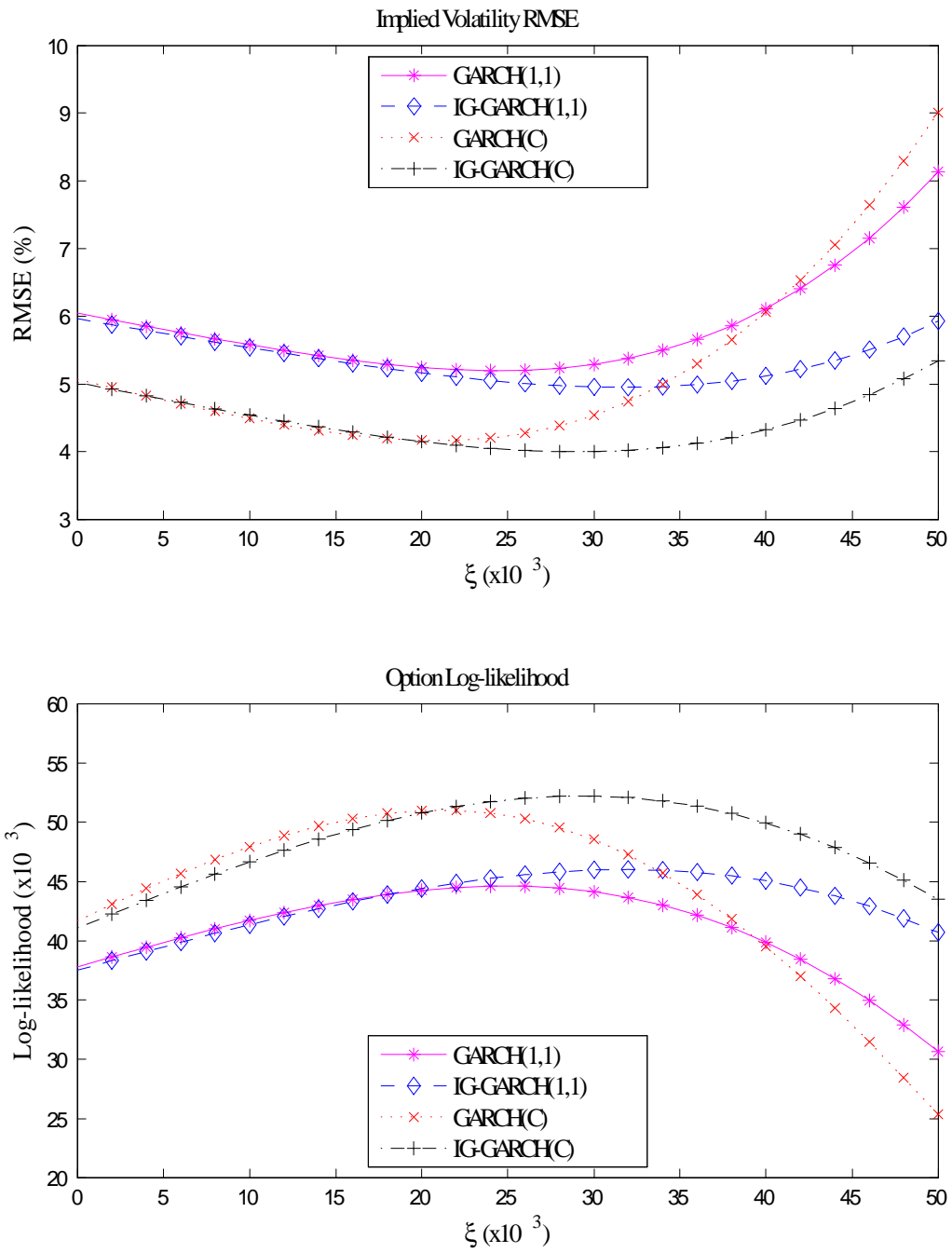
Notes to Figure: We plot the RMSE (top panel) and the option likelihood function (bottom panel) as a function of the nonlinear pricing kernel parameter, ξ. All other parameter values are fixed at their optimal values from Table 2.

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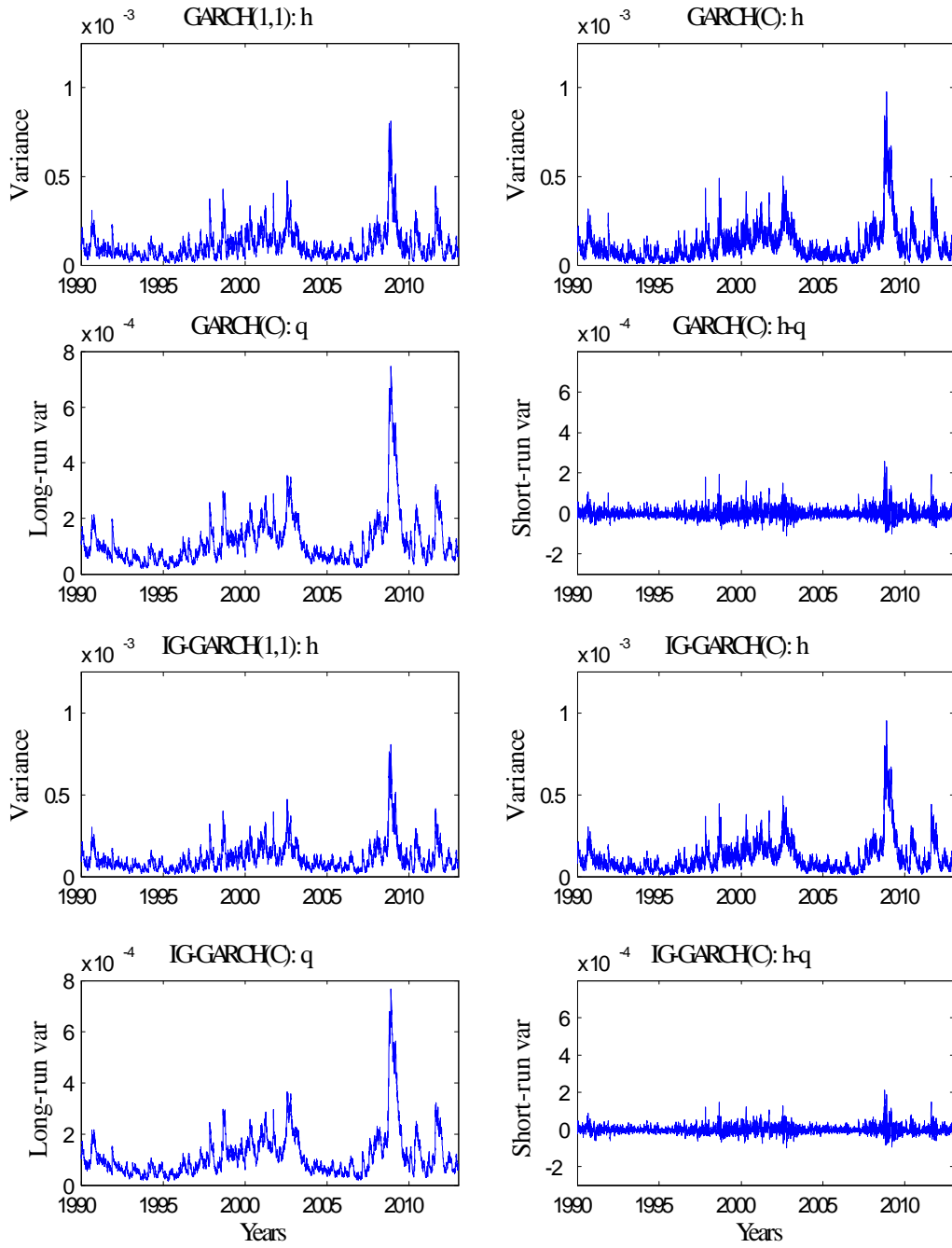
Notes to Figure: For each model we plot the spot variance components over time. The parameter values are obtained from MLE on returns in Table 2.

Figure 3. Leverage Correlation and Volatility of Variance Using Return-Based Estimates
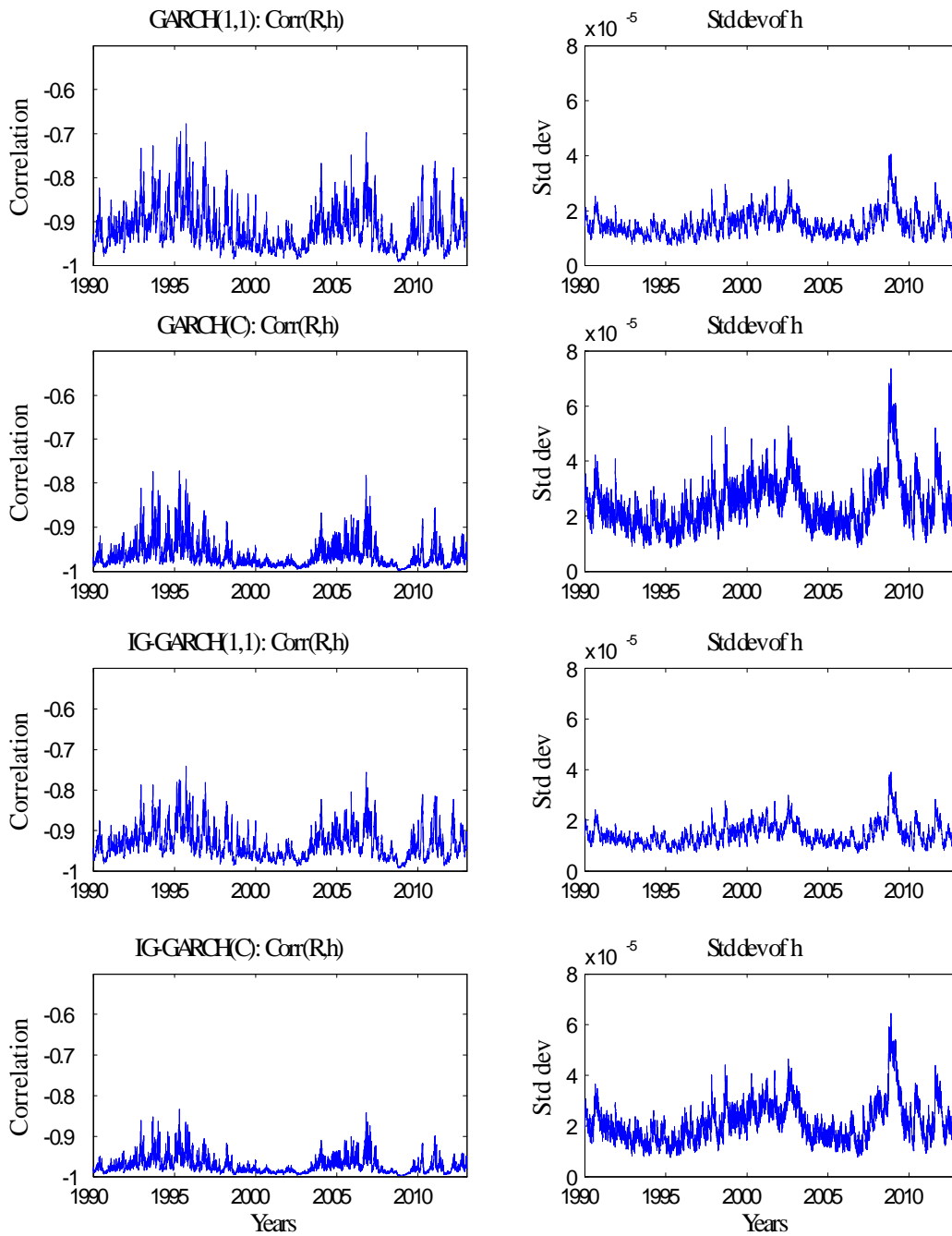
Notes to Figure: For each model we plot the conditional correlation and the conditional standard deviation of variance. In the left panels, we plot the conditional correlation between return and variance as implied by the models. In the right panels, we plot the conditional standard deviation of conditional 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.

Figure 4. Term Structure of Variance, Skewness and Kurtosis
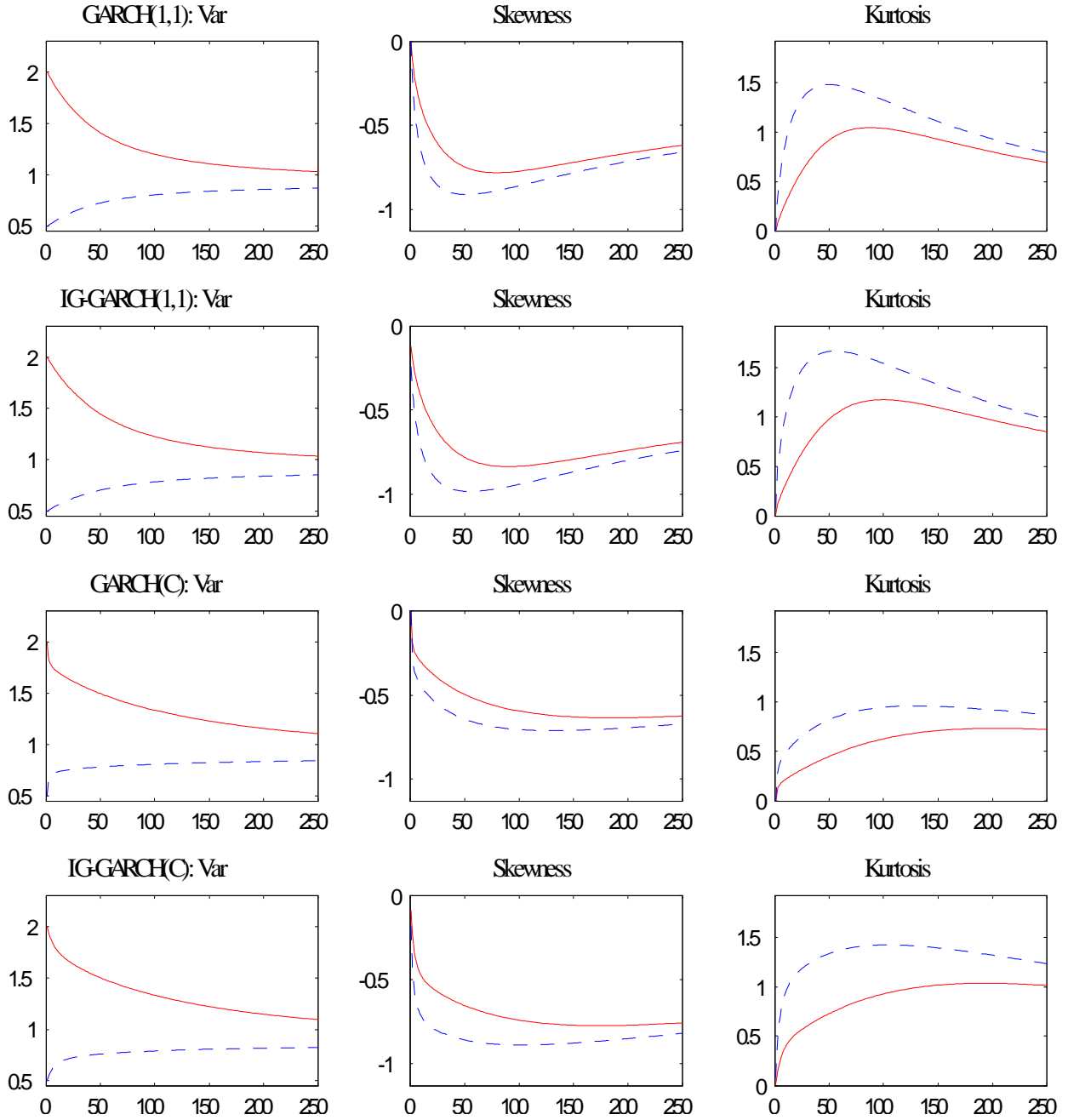
Notes to Figure: We plot the term structure of variance, skewness and excess kurtosis with high (solid) and low (dashed) initial variance for 1 though 250 trading days. Conditional variance is normalized by the unconditional variance, σ^2. For the low initial variance, the initial value of $q(t + \Delta)$ 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.

41

Figure 5. Spot Variance Paths Using Option-Based Estimates
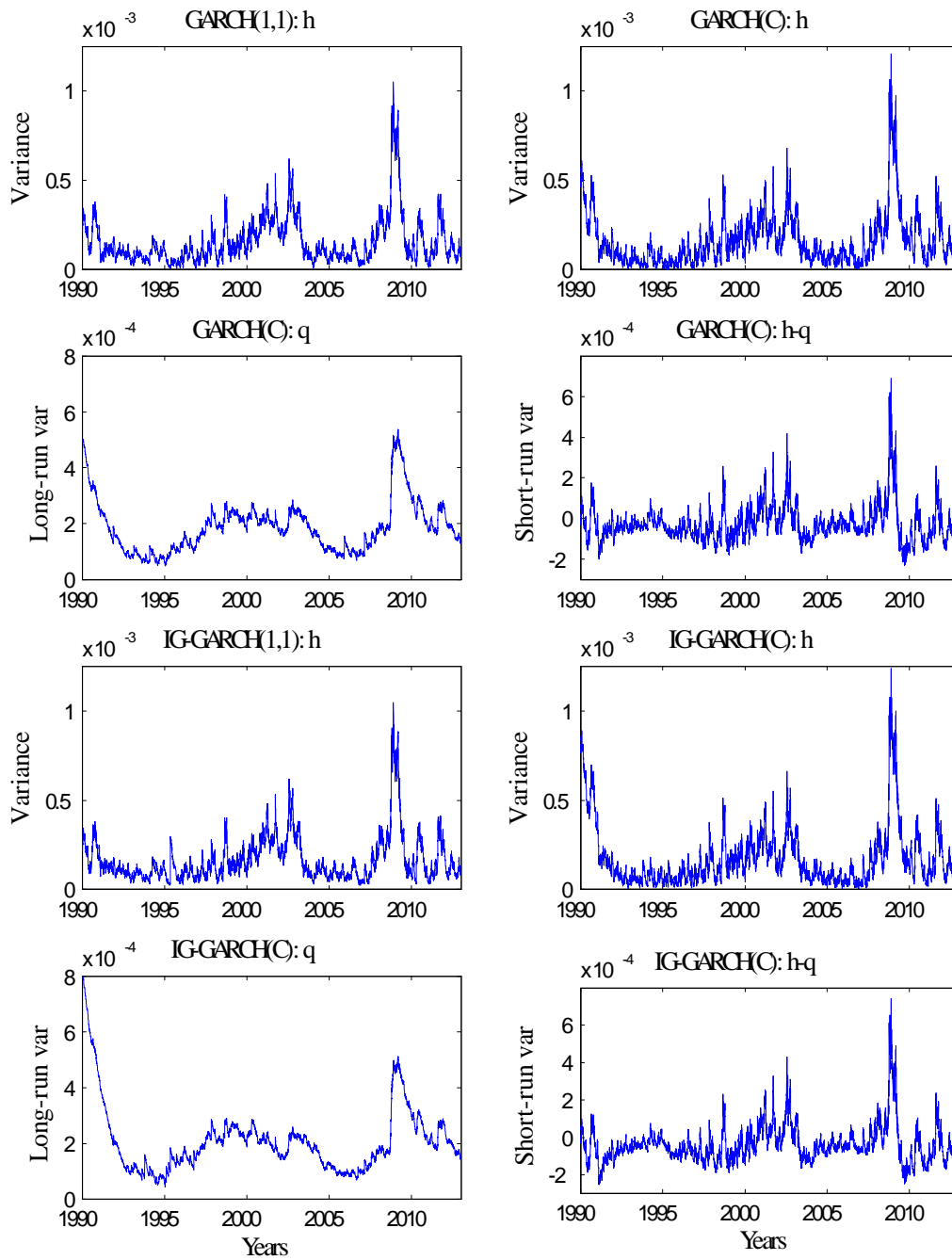
Notes to Figure: We plot the spot variance components over time. The parameter values are obtained from MLE on options in Table 3.

Figure 6. Leverage Correlation and Volatility of Variance Using Option-Based Estimates
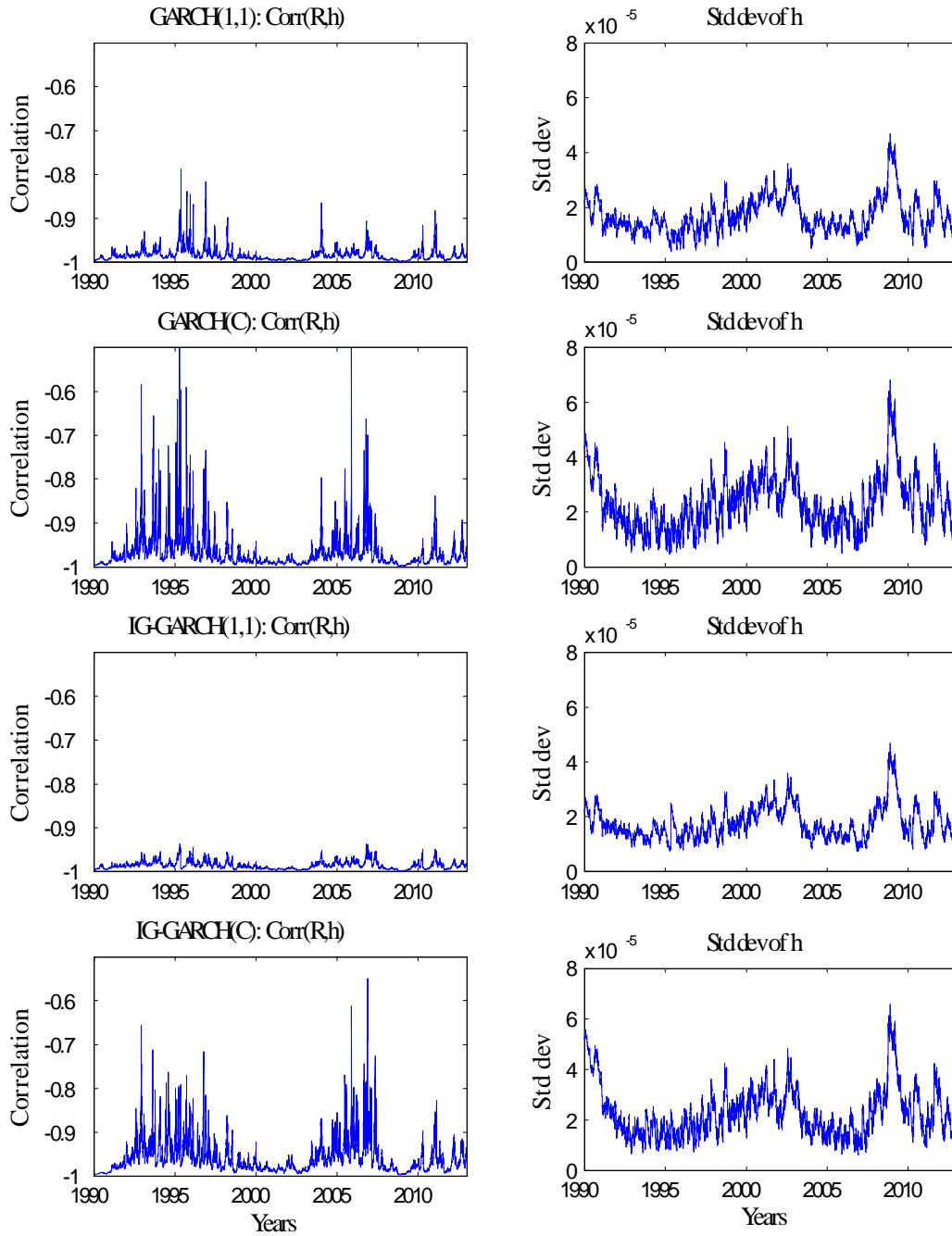
Notes to Figure: For each model we plot the conditional correlation and the conditional standard deviation of variance. In the left panels, we plot the conditional correlation between return and variance as implied by the models. In the right panels, we plot the conditional standard deviation of conditional 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
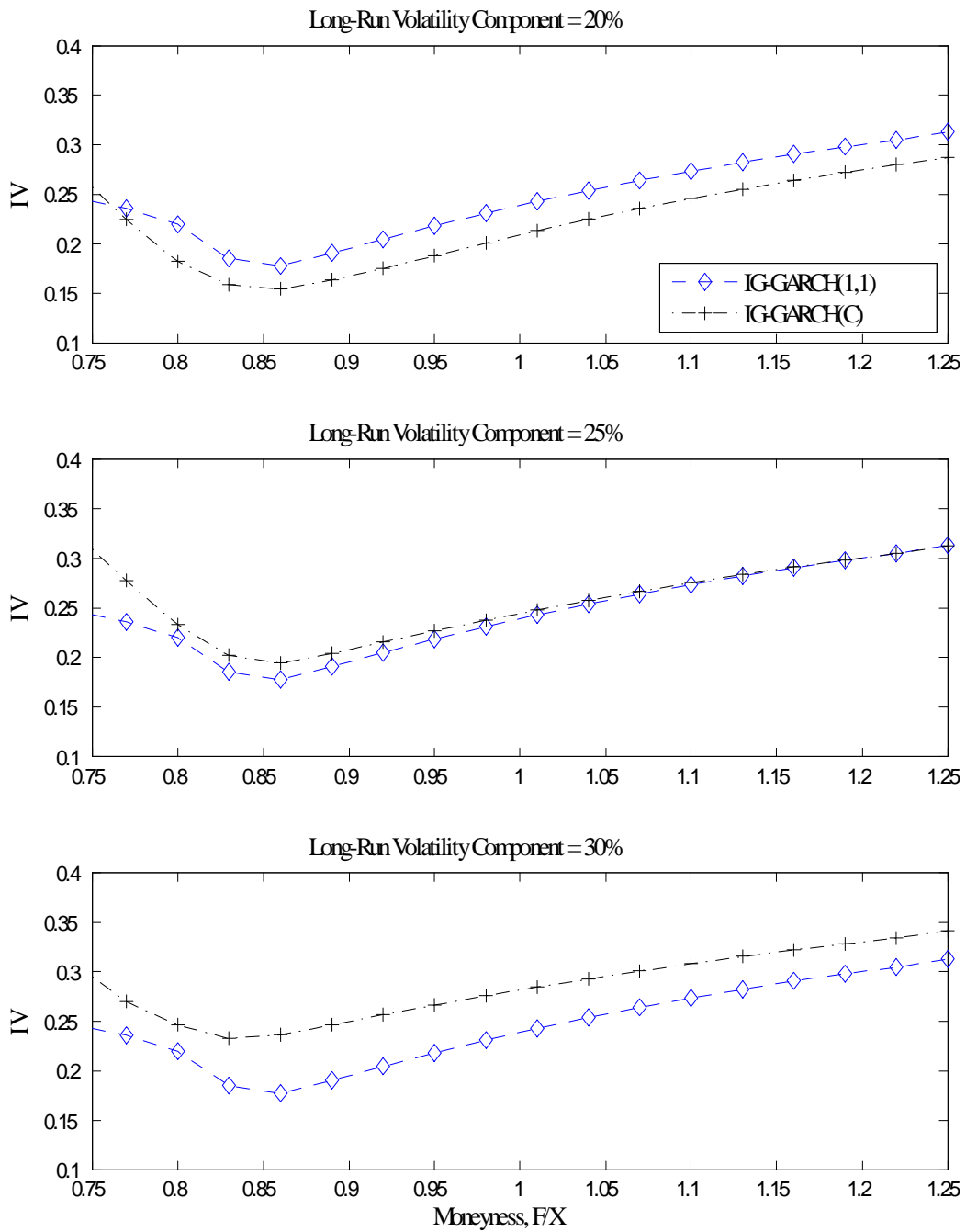
Notes to Figure: We plot model-based implied volatility smiles for 30 days to maturity from the IG-GARCH(1,1) and IG-GARCH(C) models. Long-run volatility, $\sqrt{q(t)}$, is set to 20% (top) panel, 25% (middle panel), and 30% (bottom panel). Total volatility, $\sqrt{h(t)}$ is set to 25% in all panels. 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.

Table 1. Returns and Options Data

Panel A: Return Characteristics (Annualized)

	1990-2012	1996-2012
Mean	6.06%	4.99%
Std. deviation	18.61%	20.57%
Skewness	-0.228	-0.217
Excess kurtosis	8.461	7.235

Panel B. Option Data by Moneyness

	F/X ≤ .80	.80 < F/X ≤ .90	.90 < F/X ≤ 1.00	1.00 < F/X ≤ 1.10	1.10 < F/X ≤ 1.20	F/X>1.20	All
Number of Contracts	720	3,819	8,413	8,033	5,778	2,259	29,022
Average IV	23.11%	19.65%	18.79%	22.09%	27.03%	30.52%	22.47%
Average Price	62.94	40.71	43.62	47.93	33.18	28.14	41.63
Average $ Spread	1.30	1.42	1.89	2.06	1.58	1.41	1.76

Panel C. Option Data by Maturity

	DTM ≤ 30	30 < DTM ≤ 60	60 < DTM ≤ 90	90 < DTM ≤ 120	120 < DTM ≤ 180	DTM>180	All
Number of Contracts	2,771	6,127	4,565	2,720	4,019	8,820	29,022
Average IV	24.93%	23.33%	22.45%	22.93%	21.74%	21.32%	22.47%
Average Price	18.06	26.83	31.88	39.29	48.54	61.92	41.63
Average $ Spread	0.94	1.31	1.59	1.87	1.97	2.29	1.76

We present descriptive statistics for daily return data from January 1, 1990 through December 31, 2012, as well as for daily return data from January 10, 1996 through December 31, 2012. We use Wednesday closing options contracts from January 10, 1996 to December 31, 2012.

Table 2. Maximum Likelihood Estimation on Returns

Panel A: Parameter Estimates

Gaussian Models

	$\tilde{\mu}$	ω	β	α	γ	ρ_2	α_q	γ_q
Homoskedastic	0.779	1.373E-04						

	$\tilde{\mu}$	ω	β	α	γ
GARCH(1,1)	1.059	5.653E-18	0.836	3.823E-06	184.2

	$\tilde{\mu}$	ω_q	ρ_1	α_h	γ_h	ρ_2	α_q	γ_q
GARCH(C)	1.264	1.473E-06	0.705	9.979E-07	840.6	0.987	2.832E-06	118.7

Inverse Gaussian Models

	μ	w							η
Homoskedastic	7819.3	1.372E-04							-1.279E-04

	μ	w	b	a	c				η
IG-GARCH(1,1)	1685.3	2.423E-15	-19.77	2.617E+07	4.061E-06				-5.938E-04

	μ	w_q	ρ_1	a_h	c_h	ρ_2	a_q	c_q	η
IG-GARCH(C)	2239.1	1.393E-06	0.743	2.247E+06	6.987E-07	0.988	7.101E+07	3.153E-06	-4.469E-04

Panel B: Model Properties and Likelihoods

	Return Mean	Annualized Volatility	Volatility Persistence	Uncond. Skewness	Uncond. Kurtosis	Return Log-Likelihood
Gaussian Models						
Homoskedastic	5.99%	18.60%		0.000	3.000	17,548
GARCH(1,1)	6.28%	16.79%	0.9658	0.015	4.750	18,755
GARCH(C)	6.91%	16.91%	0.9962	0.024	5.199	18,829
Inverse Gaussian Models						
Homoskedastic	5.99%	18.59%		-0.033	3.000	17,552
IG-GARCH(1,1)	6.45%	16.64%	0.9704	-0.152	4.775	18,794
IG-GARCH(C)	6.78%	16.81%	0.9968	-0.099	5.247	18,863

Parameter values are obtained from ML estimation on returns from 1990 to 2012. For each model we report parameter estimates, the maximum log-likelihood values and various model properties. We estimate six models. Each model has constant or time-varying volatility (GARCH(1,1) or component GARCH), and Normal or IG innovations.

Table 3. Maximum Likelihood Estimation On Options

Panel A: Parameter Estimates

Gaussian Models

Homoskedastic	$\tilde{\mu}^*$	ω^*
	-0.5000	1.272E-04

GARCH(1,1)	$\tilde{\mu}^*$	ω^*	β^*	α^*	γ^*
	-0.5000	7.133E-18	0.7018	1.780E-06	404.08

GARCH(C)	$\tilde{\mu}^*$	ω_q^*	ρ_1^*	α_h^*	γ_h^*	ρ_2^*	α_q^*	γ_q^*
	-0.5000	3.348E-07	0.9813	1.450E-06	575.56	0.9993	1.644E-06	87.55

Inverse Gaussian Models

Homoskedastic	μ^*	w^*	η^*
	30.4	1.593E-04	-3.241E-02

IG-GARCH(1,1)	μ^*	w^*	b^*	a^*	c^*	η^*
	405.1	2.446E-16	-0.1902	4.911E+04	5.373E-06	-2.465E-03

IG-GARCH(C)	μ^*	w_q^*	ρ_1^*	a_h^*	c_h^*	ρ_2^*	a_q^*	c_q^*	η^*
	1075.6	3.082E-07	0.9854	2.206E+06	3.153E-06	0.9996	2.573E+06	2.139E-06	-9.293E-04

Panel B: Model Properties and Likelihoods

	Return Mean	Annualized Volatility	Volatility Persistence	Uncond. Skewness	Uncond. Kurtosis	Option Log-Likelihood
Gaussian Models						
Homoskedastic	1.69%	17.90%		0.000	3.000	32,632
GARCH(1,1)	0.32%	24.40%	0.9925	-0.017	5.113	52,590
GARCH(C)	-3.01%	35.51%	0.9999	-0.018	5.485	58,647
Inverse Gaussian Models						
Homoskedastic	1.35%	20.04%		-0.101	3.699	36,946
IG-GARCH(1,1)	0.33%	24.40%	0.9923	-0.592	5.379	53,597
IG-GARCH(C)	-6.81%	44.97%	0.9999	-0.147	5.732	60,201

Parameter values are obtained from ML estimation on options from 1996 to 2012. For each model we report parameter estimates, the maximum log-likelihood values and various model properties. We estimate six models using only options data. Each model has constant or time-varying volatility (GARCH(1,1) or component GARCH), and Normal or IG innovations.

Table 4. Sequential Maximum Likelihood Estimation of ξ

Panel A: Risk-Neutral Parameters Using Table 2 Estimates and ξ Estimated from Options

<u>Gaussian Models</u>

	$\tilde{\mu}^*$	ω^*	β^*	α^*	γ^*				ξ	s_h
GARCH(1,1)	-0.5000	6.976E-18	0.8362	5.821E-06	150.61				24796.2	1.2340

	$\tilde{\mu}^*$	ω_q^*	ρ_1^*	α_h^*	γ_h^*	ρ_2^*	α_q^*	γ_q^*	ξ	s_h
GARCH(C)	-0.5000	2.446E-06	0.7082	1.369E-06	725.86	0.9883	4.082E-06	102.01	21131.6	1.1931

<u>Inverse Gaussian Models</u>

	μ^*	w^*	b^*	a^*	c^*				η^*	ξ	s_h	s_y
IG-GARCH(1,1)	1331.1	3.068E-15	-19.7747	1.634E+07	6.504E-06				-7.510E-04	31933.3	1.2661	0.7906

	μ^*	w_q^*	ρ_1^*	a_h^*	c_h^*	ρ_2^*	a_q^*	c_q^*	η^*	ξ	s_h	s_y
IG-GARCH(C)	1851.4	2.416E-06	0.745	1.033E+06	9.685E-07	0.989	4.909E+07	4.661E-06	-5.400E-04	29287.8	1.2094	0.8275

Panel B: Model Properties and Likelihoods

	Return Mean	Annualized Volatility	Volatility Persistence	Uncond. Skewness	Uncond. Kurtosis	Option Log-Likelihood	LL Increase when ξ > 0
<u>Gaussian Models</u>							
GARCH(1,1)	0.99%	21.48%	0.9682	-0.013	4.745	44,628	6,822
GARCH(C)	0.66%	22.94%	0.9966	-0.017	5.190	51,018	9,049
<u>Inverse Gaussian Models</u>							
IG-GARCH(1,1)	0.89%	21.95%	0.9728	-0.202	4.799	46,030	8,459
IG-GARCH(C)	0.55%	23.44%	0.9972	-0.154	5.238	52,252	11,169

The parameter estimate for ξ is obtained from ML estimation on options from 1996 to 2012. The physical parameter values are fixed at their values in Table 2. For each model we report risk-neutral parameter estimates, the maximum log-likelihood values and various model properties. Each model has constant or time-varying volatility (GARCH(1,1) or component GARCH), and Normal or IG innovations. The final column of Panel B shows the difference in log-likelihood from using the ML estimate of ξ versus ξ=0.

Table 5. Implied Volatility RMSE and Bias by Moneyness

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

Model	F/X ≤ 0.96	0.96 < F/X ≤ 0.98	0.98 < F/X ≤ 1.02	1.02 < F/X ≤ 1.04	1.04 < F/X ≤ 1.06	F/X>1.06	All
GARCH(1,1)	4.7538 (2.6256)	7.9637 (5.6904)	5.7226 (3.1309)	6.3972 (3.8681)	6.9295 (5.0128)	7.0103 (5.3526)	6.0435 (3.8773)
IG-GARCH(1,1)	4.6809 (2.8804)	7.8708 (5.7294)	5.6561 (3.2534)	6.3455 (3.9669)	6.8627 (5.0478)	6.9141 (5.3183)	5.9629 (3.9754)
GARCH(C)	3.6092 (2.0670)	5.6891 (4.2006)	4.4718 (2.7153)	5.3334 (3.5339)	5.7224 (4.3948)	6.2743 (5.0975)	5.0723 (3.4780)
IG-GARCH(C)	3.5671 (2.2756)	5.6160 (4.2268)	4.4539 (2.8352)	5.3230 (3.6373)	5.6844 (4.4457)	6.2005 (5.0653)	5.0228 (3.5631)

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

Model	F/X ≤ 0.96	0.96 < F/X ≤ 0.98	0.98 < F/X ≤ 1.02	1.02 < F/X ≤ 1.04	1.04 < F/X ≤ 1.06	F/X>1.06	All
GARCH(1,1)	3.6148 (0.3382)	3.4675 (0.1287)	3.6557 (-0.3241)	3.9518 (0.1918)	3.5736 (0.4219)	4.3980 (1.8215)	3.9507 (0.7058)
IG-GARCH(1,1)	3.4197 (0.5311)	3.4498 (0.1525)	3.5582 (-0.4591)	3.8847 (0.1026)	3.4715 (0.3597)	4.2708 (1.5223)	3.8163 (0.6013)
GARCH(C)	3.0174 (0.5015)	2.4337 (-0.3638)	2.8402 (-0.4170)	2.7141 (-0.1988)	2.4888 (0.1124)	3.6174 (1.3297)	3.2068 (0.5254)
IG-GARCH(C)	2.9337 (0.5423)	2.3214 (-0.5815)	2.5956 (-0.2646)	2.4870 (-0.3216)	2.2640 (-0.0174)	3.4440 (1.3124)	3.0391 (0.5750)

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

Model	F/X ≤ 0.96	0.96 < F/X ≤ 0.98	0.98 < F/X ≤ 1.02	1.02 < F/X ≤ 1.04	1.04 < F/X ≤ 1.06	F/X>1.06	All
GARCH(1,1)	5.1818 (-0.1170)	5.9489 (2.1833)	4.6646 (0.0838)	5.0944 (0.3428)	5.0569 (1.5103)	5.5915 (2.7539)	5.1966 (1.0823)
IG-GARCH(1,1)	4.8429 (0.0344)	5.6658 (1.8871)	4.4779 (-0.0512)	4.9322 (0.1102)	4.8392 (1.2101)	5.3685 (2.4476)	4.9522 (0.9596)
GARCH(C)	4.0257 (-0.6757)	3.8112 (0.6845)	3.4206 (-0.0641)	4.0290 (0.0412)	3.8131 (0.9205)	4.7858 (2.7631)	4.1696 (0.8694)
IG-GARCH(C)	3.7658 (-0.5471)	3.6324 (0.4635)	3.3035 (-0.1076)	3.9351 (-0.0954)	3.6592 (0.7326)	4.6272 (2.5711)	3.9962 (0.8147)

We report implied volatility (IV) RMSE (values before parentheses) and bias (values inside parentheses) in percent by moneyness using the option data from Table 1. 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

Model	DTM ≤ 30	30 < DTM ≤ 60	60 < DTM ≤ 90	90 < DTM ≤ 120	120 < DTM ≤ 180	DTM>180	All
GARCH(1,1)	4.9878 (3.2580)	5.7450 (3.7821)	5.9856 (3.8047)	6.6233 (4.5267)	6.1736 (3.9695)	6.3244 (3.9333)	6.0435 (3.8773)
IG-GARCH(1,1)	4.9501 (3.3390)	5.6622 (3.8676)	5.8679 (3.9021)	6.4871 (4.6084)	6.0932 (4.0735)	6.2699 (4.0482)	5.9629 (3.9754)
GARCH(C)	4.5287 (3.0240)	5.0267 (3.4424)	4.9615 (3.3925)	5.3320 (3.8838)	5.0544 (3.5511)	5.2454 (3.5313)	5.0723 (3.4780)
IG-GARCH(C)	4.5225 (3.0903)	4.9713 (3.5120)	4.8926 (3.4775)	5.2497 (3.9643)	5.0139 (3.6418)	5.2035 (3.6318)	5.0228 (3.5631)

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

Model	DTM ≤ 30	30 < DTM ≤ 60	60 < DTM ≤ 90	90 < DTM ≤ 120	120 < DTM ≤ 180	DTM>180	All
GARCH(1,1)	4.4182 (1.2492)	4.3651 (1.2116)	3.9403 (0.7438)	3.6902 (1.0911)	3.6926 (0.5963)	3.6796 (0.0952)	3.9507 (0.7058)
IG-GARCH(1,1)	4.2098 (1.1282)	4.1509 (1.0663)	3.7766 (0.6254)	3.5372 (1.0074)	3.6041 (0.4783)	3.6370 (0.0312)	3.8163 (0.6013)
GARCH(C)	4.2251 (1.1280)	3.8247 (1.1506)	3.2521 (0.6782)	3.0523 (0.6340)	2.6903 (0.4332)	2.5343 (-0.1688)	3.2068 (0.5254)
IG-GARCH(C)	4.2921 (1.2239)	3.6488 (1.2718)	3.0162 (0.8016)	2.7520 (0.7240)	2.5094 (0.4483)	2.3327 (-0.2183)	3.0391 (0.5750)

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

Model	DTM ≤ 30	30 < DTM ≤ 60	60 < DTM ≤ 90	90 < DTM ≤ 120	120 < DTM ≤ 180	DTM>180	All
GARCH(1,1)	4.4707 (1.7785)	5.0268 (1.6895)	5.2312 (1.1290)	5.5357 (1.6661)	5.2482 (0.8060)	5.3722 (0.3635)	5.1966 (1.0823)
IG-GARCH(1,1)	4.2917 (1.7483)	4.7541 (1.6391)	4.9206 (1.0436)	5.1700 (1.5374)	5.0181 (0.6620)	5.1922 (0.1537)	4.9522 (0.9596)
GARCH(C)	4.0309 (1.7748)	4.2877 (1.6852)	4.0785 (1.0964)	4.1947 (1.2993)	3.9767 (0.5857)	4.2531 (-0.1026)	4.1696 (0.8694)
IG-GARCH(C)	3.9603 (1.7970)	4.1118 (1.6924)	3.8615 (1.0839)	3.9193 (1.2566)	3.8005 (0.5159)	4.1037 (-0.2431)	3.9962 (0.8147)

We report implied volatility (IV) RMSE (values before parentheses) and bias (values inside parentheses) in percent by maturity using the option data from Table 1. 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.

Research Papers
2013

CREATES
Center for Research in Econometric
Analysis of Time Series

2015-37:	Markku Lanne and Jani Luoto: Estimation of DSGE Models under Diffuse Priors and Data-Driven Identification Constraints
2015-38:	Lorenzo Boldrini and Eric Hillebrand: Supervision in Factor Models Using a Large Number of Predictors
2015-39:	Lorenzo Boldrini and Eric Hillebrand: The Forecasting Power of the Yield Curve, a Supervised Factor Model Approach
2015-40:	Lorenzo Boldrini: Forecasting the Global Mean Sea Level, a Continuous-Time State-Space Approach
2015:41:	Yunus Emre Ergemen and Abderrahim Taamouti: Parametric Portfolio Policies with Common Volatility Dynamics
2015-42:	Mikkel Bennedsen: Rough electricity: a new fractal multi-factor model of electricity spot prices
2015-43:	Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Hybrid scheme for Brownian semistationary processes
2015-44:	Jonas Nygaard Eriksen: Expected Business Conditions and Bond Risk Premia
2015-45:	Kim Christensen, Mark Podolskij, Nopporn Thamrongrat and Bezirgen Veliyev: Inference from high-frequency data: A subsampling approach
2015-46:	Asger Lunde, Anne Floor Brix and Wei Wei: A Generalized Schwartz Model for Energy Spot Prices - Estimation using a Particle MCMC Method
2015-47:	Annastiina Silvennoinen and Timo Teräsvirta: Testing constancy of unconditional variance in volatility models by misspecification and specification tests
2015-48:	Harri Pönkä: The Role of Credit in Predicting US Recessions
2015-49:	Palle Sørensen: Credit policies before and during the financial crisis
2015-50:	Shin Kanaya: Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach
2015-51:	Tommaso Proietti: Exponential Smoothing, Long Memory and Volatility Prediction
2015-52:	Mark Podolskij, Christian Schmidt and Mathias Vetter: On U- and V-statistics for discontinuous Itô semimartingale
2015-53:	Mark Podolskij and Nopporn Thamrongrat: A weak limit theorem for numerical approximation of Brownian semi-stationary processes
2015-54:	Peter Christoffersen, Mathieu Fournier, Kris Jacobs and Mehdi Karoui: Option-Based Estimation of the Price of Co-Skewness and Co-Kurtosis Risk
2015-55	Kadir G. Babaglou, Peter Christoffersen, Steven L. Heston and Kris Jacobs: Option Valuation with Volatility Components, Fat Tails, and Nonlinear Pricing Kernels