ARE THE HIGH-ORDER MOMENTS OF THE ASSETS RETURNS DISTRIBUTION FORECASTABLE?

Trino Manuel Ñíguez1, Javier Perote2 and Antonio Rubia
1University of Westminster
2Rey Juan Carlos University
3University of Alicante

Abstract

This paper analyzes the out-of-sample ability of different parametric and semi-parametric GARCH-type models to forecast the conditional variance and the conditional and unconditional kurtosis of three types of financial assets (stock index, exchange rate and Treasury Note). For this purpose, we consider the Gaussian and Student-t GARCH models by Bollerslev (1986, 1987), and two different time-varying conditional kurtosis GARCH models based on the Student-t and a transformed Gram-Charlier density.

Key words: Gram-Charlier densities; Financial data; High-order moments; Out-of-sample forecasting.

JEL classification: C16, G1.

1. Introduction

The literature related to financial econometrics and asset pricing has shown that the conditional distribution of high-frequency returns exhibits stylized features that include excess of kurtosis, negative skewness, and temporal persistence in conditional moments. Remarkably, time dependency may be a characteristic that not only is present in the dynamics of the expected return and the conditional variance, but also in higher-order moments $E_t (r_{t+1}^s)$, $s \geq 3$; see, Nelson (1996). Among these, the conditional skewness and kurtosis (related to the third- and fourth-order conditional central moments, respectively) are particularly relevant for their implications in risk management, asset pricing, and optimal portfolio selection, as pointed out by Chunhachinda, Dandapani, Hamid and Prakash (1997), Harvey and Siddique (2000), Christie-David and Chaudhry (2001) and Schmidt (2002). For instance, rational investors concerned with the non-Gaussian properties of returns would be averse to negative skewness and high kurtosis. As a result, the composition of their optimal portfolio would change (everything else being equal)

1Department of Economics and Quantitative Methods, Westminster Business School, University of Westminster, 35 Marylebone Road, NW1 5LS London, UK. Phone: +44 (0) 22 7911 5000. Fax: +44 (0) 22 7911 5839, Email address: T.M.Niguez@wmin.ac.uk
whenever they expect changes in any of those characteristics. In this regard, Fang and Lai (1997) have reported empirical evidence of positive risk premiums for conditional skewness and conditional kurtosis in the US market. All this has given rise to a string of recent articles which have gone beyond the traditional modelling and forecasting of the conditional volatility to also focus on the time-varying properties of higher-order moments. The models proposed in this literature include both parametric (see, among others, Hansen 1994, Dueker 1997, Harvey and Siddique 1999, and Brooks, Burke, Heravi and Persand 2005) and semi-parametric approaches, such as entropy distributions (Rockinger and Jondeau 2002), and Gram-Charlier densities (León, Rubio and Serna 2005).

The econometric modelling of high-order moments attempts to exploit time dependency to improve the forecasts which are typically needed in financial applications. In this paper, we analyze the out-of-sample ability of different parametric and semi-parametric GARCH-type models to forecast the conditional variance as well as the conditional and unconditional kurtosis of several classes of financial assets. We do not focus on asymmetric distributions (i.e., we do not consider skewness in this paper) so that we can specifically isolate the gains from modelling kurtosis, which is widely considered as the most representative non-Gaussian stylized feature of financial data. We compare four different approaches in our study with increasing degree of complexity, going from the standard GARCH model to more sophisticated specifications. The starting point in our analysis is the simple Gaussian GARCH model by Bollerslev (1986), which implies the same degree of constant conditional kurtosis as the Normal distribution. Second, we consider the straightforward generalization of this model suggested by Bollerslev (1987) which, on the basis of the conditional Student-\(t\) distribution, is able to capture the underlying conditional kurtosis in the data, still assuming constant kurtosis. Next, we consider a further generalization, the so-called Student-\(t\) GARCHK model, suggested in Brooks et al. (2005). This is intended to fit the time-varying dynamics of the conditional variance and the kurtosis separately via a Student-\(t\) distribution with a degrees of freedom parameter that is allowed to vary over time. Finally, we consider a restricted version of the semi-parametric GARCH model with time-varying conditional kurtosis proposed in León et al. (2005) as an alternative to the Student-\(t\) GARCHK model. The semi-parametric approach relies upon a Gram-Charlier type polynomial expansion so that the resulting probability density function is flexible enough to approximate any unknown density, without imposing any assumption on the underlying conditional distribution.

The main questions we try to solve refer to \(i\) whether conditional kurtosis models are able to yield better out-of-sample forecasts, and \(ii\) which conditional kurtosis approach (parametric or semi-parametric) is more appropriate for applied purposes. These are ultimately empirical questions that we shall address statistically in this paper by means of an out-of-sample forecasting analysis. In particular, we compute one-day ahead forecasts of the conditional variances and the conditional and unconditional kurtosis implied by the different models, and compare their forecasting ability in terms of the Mean Square Error (MSE) loss-function. Patton (2006) has recently argued that only few of the volatility loss functions are not affected by the choice of the proxy used, showing that the MSE loss-function is robust. The same arguments may hold when analyzing higher-order conditional moments for which the true values are not directly observable and must be proxied with sampling error. We use the procedure in Diebold and Mariano (1995) to test statistically
Are the High-Order Moments of the Assets Returns Distribution Forecastable?

The remainder of this paper is organized as follows. Section 2 describes the conditional models that we use to fit and forecast conditional variance and kurtosis. Section 3 discusses the main features of the empirical analysis. Finally, Section 4 summarizes and concludes.

2. Modelling Forecasting Conditional Variance and Higher-Order Moments

Let us first introduce the basic data generating process and the general notation used throughout the paper. We observe a sample of (daily) asset prices from which we compute the series of returns

\[ r_t = 100 \log \left( \frac{P_t}{P_{t-1}} \right), \quad t = 1, \ldots, T. \]

We assume that the returns follow the dynamics,

\[ r_t = E_{t-1}(r_t) + \varepsilon_t, \]  
\[ \varepsilon_t = h_t^{1/2} \eta_t; \]  

where the conditional expectation \( E_{t-1}(\cdot) = E(\cdot | I_{t-1}) \) is taken on the observable set of information available up to time \( t - 1 \), denoted as \( I_{t-1} \). The set of random innovations \( \eta_t \) are conditionally distributed according to certain density function \( f(\eta_t | I_{t-1}) \) that satisfies \( E_{t-1}(\eta_t) = 0 \) and \( E_{t-1}(\eta_t^2) = 1 \), with \( E(\eta_t^s) < \infty \) for some \( s > 2 \).

The expected return in the model is given by \( E_{t-1}(r_t) \). We shall use a simple AR(1) model, \( E_{t-1}(r_t) = c + \rho r_{t-1} \), to filter out any predictable component in the conditional mean of the series. The conditional variance of the process is given by \( h_t = E_{t-1}(\varepsilon_t^2) \), which is the main object of interest in many papers that specifically focus on volatility modelling. In this regard, one of the most widely used models is the GARCH(1,1) process of Bollerslev (1986), which assumes a linear functional form,

\[ h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \]  

with the parameter restrictions \( \omega > 0, \alpha, \beta \geq 0 \) ensuring almost sure positiveness in the conditional variance process. The additional restriction \( \alpha + \beta < 1 \) is sufficient and necessary for \( E(\varepsilon_t^2) < \infty \), whereas the existence of higher-order moments imply further restrictions on the driving parameters \( (\alpha, \beta) \) as well as the existence of suitable moments of \( \eta_t \) (e.g., the unconditional fourth-order moment is well defined when \( \kappa_4 \alpha^2 + \beta^2 + 2\alpha \beta < 1 \), with \( \kappa_4 \) denoting the kurtosis of \( \eta_t \)).

The enormous success of the GARCH(1,1) model strives in its appealing interpretation, large degree of statistical parsimony, and computational tractability. The main GARCH equation describes the conditional variance forecast \( h_t \) as a weighted average of a constant term (long-run variance), \( \omega \), the previous variance forecast, \( h_{t-1} \), and a proxy for the conditional variance given the information which was not available when the previous forecast was made (related to new information arrivals), \( \varepsilon_{t-1}^2 \). As a result, the model is able to capture the main stylized feature in the conditional variance (namely, clustering and persistence) by resorting to a small number of parameters in a fairly simple
representation, from which one-step and multi-step forecasts can easily be obtained. Further generalizations that conform the broad GARCH family arise by readily extending this basic structure (for instance, towards including leverage and other non-linear effects), and/or by considering different assumptions on the conditional distribution \( f(\eta_t | I_{t-1}) \). We shall discuss in more detail the basic model and several of its extensions intended to capture excess of kurtosis and time dependence in higher-order moments in the following subsections.

2.1. GARCH Modelling

2.1.1. Gaussian GARCH

The simplest approach in GARCH-type modelling is the Gaussian GARCH(1,1) model of Bollerslev (1986). In addition, to the basic data generating process (1)-(2), it is assumed that the conditional shocks \( \{ \eta_t \} \) are independent and identically normally distributed with mean 0 and variance 1, \( i.e., it is imposed the particularly strong restriction

\[ \eta_t \sim iid \mathcal{N}(0, 1), \]  

(3)

or \( f(\eta_t | I_{t-1}) = f(\eta_t) = (2\pi)^{-1/2} \exp(-\eta_t^2 / 2). \) The model is fully specified with this assumption, and the relevant parameters \( \xi_0 := (c, \rho, \omega, \alpha, \beta)' \) can then be estimated from the sample by Maximum Likelihood [ML henceforth]. Under conventional assumptions on the pre-sample observations which do not play any relevant role when the sample is large enough, the log-likelihood function of the \( t \)-th observation, after dropping a constant term, can be written as:

\[ l_t(\xi_0) = -\frac{1}{2} \log h_t - \frac{\varepsilon_t^2}{2h_t}; \quad t = 1, \ldots, T, \]

(4)

Since the information matrix related to the two sets of parameters involved (conditional mean and conditional variance) is block-diagonal, the respective parameter vectors, say \( \xi_{0m} := (c, \rho)' \) and \( \xi_{0v} := (\omega, \alpha, \beta)', \) can be estimated separately. We shall proceed in this way, computing first the demeaned series \( \hat{\varepsilon}_t = r_t - \hat{c}_T - \hat{\rho}_T r_{t-1}, \) and then estimating the remaining parameters given \( \{ \hat{\varepsilon}_t \}. \)

Conditional normality is a fairly restrictive assumption which is widely accepted not to hold in the majority of applications involving real financial data. This fact is observed even when the data are sampled on a relatively low frequency basis which implies a high degree of aggregation. Fortunately, it is also widely accepted that the Normal assumption does not play a critical role when the main purpose is model fitting and/or volatility forecasting and, in fact, there are both computational and statistical reasons that have supported the wide use of the Gaussian GARCH in applied settings. First, the Gaussian log-likelihood function, \( L(\xi_0) = \sum_{t=1}^{T} l_t(\xi_0), \) is very tractable and typically does not pose any computational problems in order to be optimized numerically – hence, the GARCH

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]

\[ \text{ML} \]
model is directly implemented in most statistical packages, and can be estimated even with a spreadsheet. Second, and more importantly, the resultant estimation, namely $\hat{\xi}_T = \arg \max_{\xi_0} L(\xi_0)$, is known to be $\sqrt{T}$-consistent and asymptotically normally distributed under certain regularity conditions even if $\eta$ are not really Gaussian distributed (in this case $\hat{\xi}_T$ is referred to as the quasi maximum likelihood estimator [QML]); see Weiss (1986), Bollerslev and Wooldridge (1992), Lee and Hansen (1994), and Newey and Steigerwald (1997). These properties ensure tractability and accuracy for many applications in which the main aim is to obtain consistent estimates of $\xi_0$, and/or forecasts of the conditional variance process, which are simply determined as $\hat{h}_{T+s} = \hat{\omega}_T + \hat{\alpha}_T \hat{\varepsilon}_T^2 + \hat{\beta}_T \hat{h}_{T-1}$ for $s = 1$, and $\hat{h}_{T+s} = \hat{\omega}_T + \left( \hat{\alpha}_T + \hat{\beta}_T \right) \hat{h}_{T+s-1}$ for $s > 1$.

### 2.1.2. Student-\(t\) GARCH

The Normal assumption may be convenient, but it turns out to be too restrictive for applications on risk management and asset pricing, because these require the conditional density of $\eta$, and not just volatility estimations.\(^3\) The failure of the Normal assumption is mainly due to the large degree of kurtosis that is typically observable in real data, which in turn is related to the magnitude and the frequency of extreme values that characterize almost any financial time series. Although the unconditional distribution implied by the Gaussian GARCH(1,1) model is leptokurtic (Bollerslev, 1986), often this model cannot generate large enough values to match the range which is observed in practice owing to limitations in its statistical properties; see Carnero, Peña and Ruiz (2004) for a discussion on this topic. Furthermore, the empirical distribution of the estimates $\hat{\eta}$, given the ML estimates $\hat{\xi}_T$, also suggest an excess of conditional kurtosis over the theoretical level which is implied by the Normal distribution.\(^4\) Overall, the empirical evidence largely supports the existence of strong leptokurtosis in both the unconditional and conditional distributions of returns, thereby suggesting model misspecification in the Gaussian GARCH approach.

This observation motivated further extensions aiming to capture extreme movements through heavy-tailed distributions. A very simple, yet useful extension, was early suggested by Bollerslev (1987), who proposed a transformed Student \(t\)-distribution with \(v\) degrees of freedom to accommodate the excess of kurtosis, \emph{i.e.},

$$\eta_t \sim iid \ t_{\nu}(0, 1),$$

where the degrees of freedom parameter, \(v\), is directly characterized by the shape of the underlying distribution, and can be estimated by ML from the available data (subject to the restriction \(v > 2\) so that the variance process is well defined). Apart from a constant term, the relevant log-likelihood function is given by,

$$l_t(\xi_t) = \log \left( \frac{\Gamma((v+1)/2)}{\Gamma(v/2)} \right) - \frac{1}{2} \log((v-2) h_t) - \frac{v+1}{2} \log \left[ 1 + \frac{\varepsilon_t^2}{(v-2) h_t} \right],$$

\(^3\)A leading example is the Value at Risk methodology. The percentiles of $\eta_t$, together with the forecasts of the future conditional variance, jointly determine the maximum expected loss of an asset at a certain significance level.

\(^4\)For instance, the standardized residuals of the Gaussian GARCH model studied in Section 3 below have a kurtosis of nearly 6 in the in-sample period considered for the S&P index. The Jarque-Bera test for normality (JB=444.56) rejects the hypothesis of normality. Similar results are obtained for the remaining time-series.
with \( \xi_1 = (v, \xi_0)' \), and \( \Gamma(\cdot) \) denoting the Gamma function. When \( 1/\hat{v}_T \to 0 \), the conditional distribution approaches a Normal distribution and the Gaussian restriction may be acceptable. However, for small values such that \( 1/\hat{v}_T > 0 \), the empirical distribution has fatter tails than the corresponding Normal distribution. For many empirical applications related to risk-management, such as Value at Risk, the Student-t GARCH model tends to provide a superior performance over the Gaussian GARCH model; see, for instance, Alexander (1998).

### 2.2. Further Approaches: Modelling Higher-Order Conditional Moments

#### 2.2.1. The Student-t GARCHK model

The Student-t GARCH model provides further flexibility to capture constant unconditional leptokurtosis. Obviously, there is no prior reason to believe that higher-order conditional moments should remain unchanged, other than for model simplicity and computational tractability. Consequently, Brooks et al., (2005) proposed a further extension of this model, the so-called Student-t GARCHK, by allowing the possibility of heterogeneity in the conditional distribution \( f(\eta_t'|I_{t-1}) \) due to time-varying kurtosis.

Considering the basic GARCH model, the key assumption now is that \( \{\eta_t\} \) are conditionally distributed according to a Student-t distribution with a time-varying number of degrees of freedom, say \( v_t \), which evolves independently of the dynamics followed by the conditional variance. In particular, the characteristic restriction is given by

\[
f(\eta_t'|I_{t-1}) \sim t(v_t, 0, 1);
\]

where \( k_t \) is the conditional kurtosis of the process at time \( t \). In the same spirit of the structural GARCH modelling, an autoregressive moving average process is used to capture the dynamics of the conditional kurtosis:

\[
k_t = \kappa + \delta \left( \frac{\varepsilon_{t-1}^2}{h_{t-1}^2} \right) + \theta k_{t-1}.
\]

As in equation (2), the parameter restrictions \( \kappa > 0, \delta, \theta \geq 0 \), are sufficient for ensuring positiveness in the resultant process. Note that \( k_t \) arises as a weighted combination of a long-run constant value, the previous kurtosis forecast, and a term with updated information of the conditional kurtosis as proxied by \( \left( \frac{\varepsilon_{t-1}^2}{h_{t-1}} \right)^2 \). Under the restriction \( \delta = \theta = 0 \), the model reduces to the constant kurtosis model studied in the previous subsection, which suggests an easy way to statistically test for the suitability of the time-varying specification.

It is important to remark that the conditional variance process, \( h_t \), and the conditional kurtosis process, \( k_t \), are not contemporaneously functionally related, so they may be parameterized individually as desired by using different specifications than those discussed above, for instance, introducing nonlinearities or dependence upon other variables. The log-likelihood of the model, apart from a constant term, is given by,

\[
l_t(\xi_2) = \log \left( \frac{\Gamma[(v_t + 1)/2]}{\Gamma(v_t/2)} \right) - \frac{1}{2} \log \left( (v_t - 2)h_t \right) - \frac{v_t + 1}{2} \log \left[ 1 + \frac{\varepsilon_t^2}{h_t(v_t - 2)} \right],
\]
Are the High-Order Moments of the Assets Returns Distribution Forecastable?

with \( \xi_2 = (\kappa, \delta, \theta, \xi_0')' \), and \( \nu > 4 \) to ensure the existence of the first fourth-order moments. The similitudes between (9) are (6) are obvious, since the time-varying kurtosis generalizes the constant kurtosis model by simply allowing time variability in the degrees of freedom parameter.

The empirical in-sample evidence discussed in Brooks et al., (2005, Section 3) for several US and UK equities and bonds supports the hypothesis of heterogeneity in the conditional kurtosis, largely outperforming the specification with constant kurtosis.

2.2.2. The Gram-Charlier GARCHK model

Let us start this section by recalling the dynamics of the conditional variance-kurtosis models which have been discussed thus far:

\[
\begin{align*}
    r_t &= E_{t-1}(r_t) + \varepsilon_t; \quad \varepsilon_t = h_t^{1/2} \eta_t, \\
    h_t &= \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \\
    k_t &= \kappa + \delta \left( \frac{\varepsilon_{t-1}^4}{h_{t-1}^2} \right) + \theta k_{t-1},
\end{align*}
\]

(10)

given the set of unknown parameter \( \xi_2 = (\kappa, \delta, \theta, \xi_0')' \). Instead of imposing a particular assumption on the conditional distribution of \( \eta_t \) as we did in the previous sections (Normal, Student-\( t \), or Student-\( t \) with time-varying degrees of freedom), we may use a Gram-Charlier type of expansion to fit semi-parametrically the unknown density function \( f(\eta_t|I_{t-1}) \). This is the central point discussed in the model proposed in León et al. (2005), which we summarize below.

Under certain regularity conditions, any probability density function (pdf henceforth) can be expanded in an infinite series of derivatives of the standard Normal density, \( \phi(\eta_t) \), as follows,

\[
f(\eta_t|I_{t-1}) = \phi(\eta_t) \sum_{s=0}^{\infty} d_s H_s(\eta_t),
\]

(11)

where \( H_s(\eta_t) \) is the \( s^{th} \) order Hermite polynomial defined in terms of the \( s^{th} \) order derivative of the Gaussian pdf:

\[
\frac{d^s \phi(\eta_t)}{d\eta_t^s} = (-1)^s \phi(\eta_t) H_s(\eta_t).
\]

(12)

For applied purposes, the infinite expansion is not operative and has to be truncated. Thus, considering the finite expansion (approximation) of \( f(\eta_t|I_{t-1}) \) in (11) with a truncation factor up to the fourth-order moment, we obtain:

\[
f(\eta_t|I_{t-1}) \simeq \phi(\eta_t) \left[ 1 + \frac{s_4}{3!} (\eta_t^3 - 3\eta_t) + \frac{k_t - 3}{4!} (\eta_t^4 - 6\eta_t^2 + 3) \right] = \phi(\eta_t) \psi(\eta_t),
\]

(13)

where the polynomial \( \psi(\eta_t) \) is defined implicitly, and the terms \( s_4 \) and \( k_t \) correspond to the conditional skewness and kurtosis, respectively. Note that the resulting approximation, \( \phi(\eta_t) \psi(\eta_t) \), is characterized by the underlying dynamics of the conditional moments up to the fourth-order moment and the set of unknown parameters \( \xi_2 \). We do not overload
unnecessarily the notation by remarking the latter feature as this is completely clear at this point.

Since the approximation based on a finite polynomial expansion of \( f(\eta_t|I_{t-1}) \) implies certain amount of truncation error, the right-hand side of (13) cannot be seen as a proper density function. The main reason is that \( \phi(\eta_t)\psi(\eta_t) \) is not ensured to be almost surely positive uniformly on the parameter space of \( \xi_2 \). This unappealing feature does not only suppose a major shortcoming from a theoretical viewpoint, but also may cause the failure of the ML estimation in empirical settings. León et al. (2005) propose a solution building on the same methodology as Gallant and Nychka (1987), Gallant and Tauchen (1989) and Gallant Nychka and Fenton (1996). In essence, they achieved a well-defined pdf by first using a simple positive transformation of \( \psi(\eta_t) \) that ensures almost-surely positiveness (namely, squaring \( \psi(\eta_t) \), although other transformation in similar spirit are possible as well), and then re-normalizing the resulting function by a suitable scaling factor such that the resulting function integrates up to one. More specifically, given the normalizing factor,

\[
\Delta_t = \int_{-\infty}^{\infty} \phi(\eta_t)\psi^2(\eta_t) \, d\eta_t = 1 + \frac{s_t^2}{3!} + \frac{(k_t - 3)^2}{4!},
\]

the transformed Gram-Charlier probability density function, denoted \( f^*(\eta_t|I_{t-1}) \), can readily be written as

\[
f^*(\eta_t|I_{t-1}) = \left( \frac{1}{\Delta_t} \right) \phi(\eta_t)\psi^2(\eta_t).
\]

Note that the Hermite polynomial that characterize \( \psi(\eta_t) \) convey information about the empirical degree of conditional moments, and so does \( f^*(\eta_t|I_{t-1}) \), from which \( \xi_2 \) can be identified from the observable data. However, the terms \( s_t \) and \( k_t \) no longer admit the interpretation of conditional moments, and further adjustments to forecast the conditional moments given \( f^*(\eta_t|I_{t-1}) \) are necessary; see Section 3.2.2 for further details. Since we are restricting ourselves to symmetric conditional distributions, we set \( s_t = 0 \) in (13) for all \( t \), and denote as \( \bar{\psi}(\eta_t) \) the restricted version of the model. Hence, the normalizing factor reduces accordingly to \( \bar{\Delta}_t = 1 + (k_t - 3)^2/4! \), and the corresponding log-likelihood function, apart from a constant term, is given by

\[
l_t(\bar{\xi}_2) = -\frac{1}{2} \ln h_t - \frac{\varepsilon_t^2}{2h_t} - \ln \bar{\Delta}_t + \ln [\varphi^2(\eta_t)].
\]

It is worth remarking at this point the similitudes between this function and the Gaussian log-likelihood function (4) used in the basic GARCH model: The Gram-Charlier log-likelihood function simply adds two adjustment terms to the latter in order to capture the non-Gaussian features of the data, in our case, conditional kurtosis dynamics. In fact, the Gaussian likelihood (4) is nested as a particular case by setting constant kurtosis (\( \delta = \theta = 0 \)), equal to that of the Normal distribution (i.e., \( k_t = \kappa = 3 \)). As in Brooks et al. (2005), the empirical results in León et al. (2005) indicate a significant presence of time-variability in the higher-order moments which support the suitability of conditional kurtosis.
3. Empirical Analysis

3.1. The Data

The data used in this study are daily returns (scaled by a factor of 100) of the S&P500 index (SP), the GBP(£)/US Dollar($) exchange rate (FX), and the 10 years Treasury Notes (TN). The series are sampled over the period June 9, 1993 to June 8, 2008 for a total of \( T = 3,912 \) observations obtained from Datastream. Table 1 displays some descriptive information for the total sample. As expected, stock returns are much more volatile (as measured by the unconditional volatility) than the other series. The unconditional distribution of any of these series shows clearly non-Gaussian features, such as a (mild) skewness in the case of the SP and FX series, and a remarked excess of kurtosis over the Normal distribution due to outliers in the three time series considered. The Jarque-Bera tests for normality are easily rejected, particularly in the case of the stock index time-series. The analysis of dependence through the Ljung-Box portmanteau test statistics shows some form of weak dependence in the level of the returns, and a strong, persistent correlation in higher-order moments.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>SP</th>
<th>FX</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>9/06/1993 - 8/06/2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3913</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.0561</td>
<td>-0.0067</td>
<td>-0.0093</td>
</tr>
<tr>
<td>Median</td>
<td>0.1186</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Maximum</td>
<td>16.107</td>
<td>3.4233</td>
<td>6.1278</td>
</tr>
<tr>
<td>Minimum</td>
<td>-15.419</td>
<td>-4.2211</td>
<td>-5.1238</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>1.8624</td>
<td>0.5112</td>
<td>1.1807</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.1033</td>
<td>-0.0131</td>
<td>0.3568</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>11.01</td>
<td>5.8469</td>
<td>5.4564</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>10470 (0.00)</td>
<td>1324 (0.00)</td>
<td>1066.8 (0.00)</td>
</tr>
<tr>
<td>Ljung-Box Q(1)-( r_1 )</td>
<td>17.29 (0.00)</td>
<td>16.65 (0.00)</td>
<td>4.0521 (0.04)</td>
</tr>
<tr>
<td>Ljung-Box Q(20)-( r_2^2 )</td>
<td>1047.1 (0.00)</td>
<td>229.4 (0.00)</td>
<td>1322.2 (0.00)</td>
</tr>
<tr>
<td>Ljung-Box Q(20)-( r_3^3 )</td>
<td>93.50 (0.00)</td>
<td>399.7 (0.00)</td>
<td>73.28 (0.00)</td>
</tr>
<tr>
<td>Ljung-Box Q(20)-( r_4^4 )</td>
<td>120.7 (0.00)</td>
<td>459.7 (0.00)</td>
<td>297.1 (0.00)</td>
</tr>
<tr>
<td>LR: (( \kappa, \delta, \theta )) = 0</td>
<td>146.9 (0.00)</td>
<td>154.4 (0.00)</td>
<td>153.2 (0.00)</td>
</tr>
<tr>
<td>LR: (( \delta, \theta )) = 0</td>
<td>77.70 (0.00)</td>
<td>9.43 (0.00)</td>
<td>14.1 (0.00)</td>
</tr>
</tbody>
</table>

The Jarque-Bera normality test is asymptotically distributed as a \( \chi^2(2) \) under the null of normality, the Ljung-Box is asymptotically distributed as a \( \chi^2(\xi) \), \( \xi \) being the autocorrelation order, the Likelihood Ratio test (LR) is asymptotically distributed as a \( \chi^2(q) \) being \( q \) the number of restrictions under the null, (asymptotic p-values in parenthesis). The critical values of \( \chi^2(1), \chi^2(2), \chi^2(3) \) and

\[ JB = T \left[ \frac{s^2}{3} + \frac{(k-3)^2}{4} \right] \]

where \( s \) and \( k \) are the sample skewness and kurtosis, respectively. The test is asymptotically distributed as \( \chi^2(2) \), and is rejected for non-zero values of the sampled skewness and/or excess of kurtosis.
\( \chi^2(20) \) are 3.84, 5.99, 7.81, 31.41, at 5% level, respectively.

### 3.2. Modelling and Out-of-Sample Forecasting

We split the total sample into an in-sample period to estimate the models, and an out-of-sample window to make a total of \( N = 500 \) one-step predictions of the conditional variance and kurtosis by means of a rolling-window procedure. To assess the ability of the different GARCH models involved, we need a time-varying measure of the actual conditional moments. In both cases, the main problem for addressing forecasting ability is that the true conditional variance and kurtosis are not observable and have to be approached by means of statistical proxies which often can only provide a crude measure.

In the context of volatility forecasting, the empirical proxies considered in most papers are based on measurable transformations of the absolute-valued unexpected returns \( |\epsilon_t| \), most frequently \( \epsilon^2_t \). Following this literature, we shall consider \( \epsilon^2_{T+1} \) as a proxy for variance in this paper.\(^6\) Although there exists an agreement (at least empirically) on how conditional variance could be proxied, to the best of our knowledge there is no obvious guidance on how to approach conditional kurtosis. Given that the estimation bias may be more significant when considering higher-order moments, and that the choice of the proxy necessarily conditions the results, we consider different proxies for the conditional kurtosis. In particular, we take the sample kurtosis in the \( m \) days immediately following the last in-sample observation, namely

\[
\bar{k}_{T+1,m} = \left( \frac{1}{m} \sum_{j=1}^{m} (\epsilon_{T+j} - \bar{\epsilon}_m)^4 \right) / \left[ \left( \frac{1}{m} \sum_{j=1}^{m} (\epsilon_{T+j} - \bar{\epsilon}_m)^2 \right)^2 \right]; \quad \bar{\epsilon}_m = \frac{1}{m} \sum_{j=1}^{m} \epsilon_{T+j}, \quad m = \{5, 50, 500\}. \quad (17)
\]

with \( m = \{5, 50, 500\} \).\(^7\) The choice of \( m \) here seeks a compromise between the tautological notion of conditional and the statistical problems related to measurement errors in the relevant statistic when using a few number of observations. As \( m \to 1 \), the proxy is more erratic and extremely noisy, and can severely be influenced by a few large observations, whereas the largest window in our analysis (\( m = 500 \) observations) is related to the out-of-sample unconditional kurtosis.

We consider the MSE as a loss function for any of the conditional variance and kurtosis forecasts series, i.e., we compute the statistic \( N^{-1} \sum_{i=1}^{N} (\hat{\pi}_{T+1,i} - \bar{\pi}_{T+1,i})^2 \) for each model and time-series, where \( \hat{\pi}_{T+1,i} \) is the \( i \)-th prediction (either conditional variance or kurtosis) and \( \bar{\pi}_{T+1,i} \) the proxy for the actual value. The forecasting performance is compared in statistical terms by means of the test proposed by Diebold and Mariano (1995). This test assumes no differences between the loss functions of two alternative models under the null hypothesis. The null is rejected for large values of the statistic \( DM = \bar{x} / \sqrt{2\pi f_\bar{x}(w=0)} / \sqrt{N}, \) where \( \bar{x} \) denotes the sample mean of the differences in the forecasting errors of the two alternative models, \( f_\bar{x}(w=0) \) is the spectral density function of the forecasting error differences evaluated at the zero frequency (long-run variance), and \( N \) is the total number of observations.

\(^6\)The availability of intraday data has motivated the use of a new strand of proxies for volatility which provide a more accurate measure based on realized volatility.

\(^7\)We also used other values for \( m \), noting no qualitative difference with the results reported in the main text.
forecasts. The statistic is asymptotically distributed as a standard Normal random variable under the null.

3.2.1. In-sample analysis

The slight (positive) autocorrelation pattern in the conditional mean of the returns is filtered out by fitting an AR(1) process estimated by Least-Squares. Given the demeaned series, \( \hat{\varepsilon}_t \), all the conditional variance-kurtosis models are then estimated by optimizing the corresponding log-likelihood functions using the Newton-Raphson method, and initializing the conditional variance and kurtosis dynamics with values equal to the corresponding unconditional moments. Convergence in the optimization process to the global extremes is easily obtained in the case of the simplest Gaussian and Student-t GARCH models. Similarly, the estimation of the Student-t and Gram-Charlier GARCHK models is not computationally troublesome providing that the starting values are chosen properly. The main results from the estimation in the in-sample period are displayed in Table 2 below.

Estimation results (robust QML \( t \)-statistics in brackets) for the Gaussian GARCH (GARCH-n), the Student-t GARCH (GARCH-t), the Student-t GARCH (GARCHK-t) and the Gram-Charlier GARCHK models (GARCHK-GC). AIC denotes the Akaike Information Criterion statistic. The row DoF shows the estimated degrees of freedom parameter in the Student-t distribution under the GARCH-t model, \( \hat{\nu}_T \), and the unconditional kurtosis implied by the estimated parameters, \( \kappa/(1 - \delta - \theta) \), in case of the GARCHK-t model.

The estimates of the conditional variance for the three series show the usual degree of high persistence and low sensitivity to shocks which is commonly observed in daily asset returns. Persistence is related to the magnitude of the coefficient \( \hat{\alpha}_T + \hat{\beta}_T \), which tends to be slightly smaller than unity, while sensitivity to new information arrivals is measured through \( \hat{\alpha}_T \), which takes small values empirically. Owing to the large degree of unconditional kurtosis in the data, the Student-t GARCH model determines a degrees of freedom parameter around 6 for all the series. This result confirms that extreme observations in real data are much more likely to occur in relation to the Normal distribution. Assuming that the true distribution is a Student-t, higher-order moments larger than 6 would not be well-defined. The models that allow for time-varying kurtosis reject the hypothesis of constant kurtosis, since the restriction \( \delta = \theta = 0 \) is easily rejected by a standard Likelihood Ratio test in the three time series considered. Overall, the empirical evidence we observe perfectly agrees with the results in Brooks et al. (2005) and León et al. (2005), showing that extending GARCH models toward accounting for time-varying kurtosis leads to a better in-sample fitting.

There are two further interesting features that arise when comparing the results observed across the different types of estimation techniques involving time-varying kurtosis, and the different classes of financial assets considered. First, whereas the estimates of the GARCH equation remain virtually unaltered given the different models, the estimates of the driving parameters of the conditional kurtosis, say \( \xi_{2k} = (\kappa, \delta, \theta)' \), reveal very different dynamics depending on whether a Student-t or a transformed Gram-Charlier densities is used. In particular, the dynamics of the conditional kurtosis tend to be much more persistent under

---

8Also, in order to avoid convergence to local extremes, the optimization routine is monitored using a grid of different starting values. Normal convergence is obtained in all the cases.
## Table 2. GARCH in-sample estimation results

<table>
<thead>
<tr>
<th>Mean equation</th>
<th>GARCH-n</th>
<th>GARCH-t</th>
<th>GARCHK-t</th>
<th>GARCHK-GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel 1: SP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>μ</strong></td>
<td>0.056</td>
<td>0.083</td>
<td>0.083</td>
<td>0.083</td>
</tr>
<tr>
<td><strong>φ</strong></td>
<td>0.118 (3.05)</td>
<td>0.107 (2.76)</td>
<td>0.107 (2.96)</td>
<td>0.113 (3.49)</td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>0.159 (5.36)</td>
<td>0.139 (4.94)</td>
<td>0.155 (5.66)</td>
<td>0.155 (6.01)</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.809 (22.7)</td>
<td>0.832 (23.3)</td>
<td>0.819 (24.2)</td>
<td>0.807 (25.9)</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>0.801 (1.49)</td>
<td>1.887 (4.12)</td>
<td>0.024 (1.07)</td>
<td>0.008 (0.64)</td>
</tr>
<tr>
<td><strong>κ</strong></td>
<td>0.842 (8.87)</td>
<td>0.448 (3.39)</td>
<td>6.202 (9.60)</td>
<td>[5.977]</td>
</tr>
<tr>
<td><strong>δ</strong></td>
<td>0.007 (1.96)</td>
<td>0.003 (1.89)</td>
<td>0.004 (1.83)</td>
<td>0.003 (2.22)</td>
</tr>
<tr>
<td><strong>θ</strong></td>
<td>0.036 (3.30)</td>
<td>0.031 (3.96)</td>
<td>0.032 (3.59)</td>
<td>0.038 (5.83)</td>
</tr>
<tr>
<td><strong>ν</strong></td>
<td>3.731</td>
<td>3.704</td>
<td>3.681</td>
<td>3.687</td>
</tr>
<tr>
<td>Panel 2: FX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>μ</strong></td>
<td>-0.007 (-0.89)</td>
<td>-0.065 (-4.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>φ</strong></td>
<td>0.007 (1.96)</td>
<td>0.003 (1.89)</td>
<td>0.004 (1.83)</td>
<td>0.003 (2.22)</td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>0.036 (3.30)</td>
<td>0.031 (3.96)</td>
<td>0.032 (3.59)</td>
<td>0.038 (5.83)</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.937 (44.1)</td>
<td>0.967 (80.4)</td>
<td>0.955 (69.6)</td>
<td>0.944 (93.9)</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>2.883 (1.63)</td>
<td>2.684 (2.66)</td>
<td>0.049 (0.28)</td>
<td>0.008 (5.82)</td>
</tr>
<tr>
<td><strong>κ</strong></td>
<td>0.599 (2.61)</td>
<td>0.223 (0.77)*</td>
<td>5.319 (10.7)</td>
<td>[5.344]</td>
</tr>
<tr>
<td><strong>δ</strong></td>
<td>0.665 (1.38)</td>
<td>0.001 (0.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>θ</strong></td>
<td>0.216 (0.72)</td>
<td>0.806 (15.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ν</strong></td>
<td>1.478</td>
<td>1.424</td>
<td>1.422</td>
<td>1.432</td>
</tr>
<tr>
<td>Panel 3: TN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>μ</strong></td>
<td>-0.0091 (-0.46)</td>
<td>0.0322 (2.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>φ</strong></td>
<td>0.012 (2.45)</td>
<td>0.007 (2.11)</td>
<td>0.007 (2.13)</td>
<td>0.009 (2.49)</td>
</tr>
<tr>
<td><strong>ω</strong></td>
<td>0.039 (5.16)</td>
<td>0.038 (5.71)</td>
<td>0.041 (4.87)</td>
<td>0.037 (5.73)</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.951 (96.8)</td>
<td>0.958 (126)</td>
<td>0.954 (104)</td>
<td>0.953 (117)</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td>4.197 (2.11)</td>
<td>0.671 (3.73)</td>
<td>0.665 (1.38)</td>
<td>0.001 (0.28)</td>
</tr>
<tr>
<td><strong>κ</strong></td>
<td>0.216 (0.72)</td>
<td>0.806 (15.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>δ</strong></td>
<td>5.685 (10.3)</td>
<td>[4.695]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>θ</strong></td>
<td>3.009</td>
<td>2.969</td>
<td>2.956</td>
<td>2.963</td>
</tr>
</tbody>
</table>
the assumption of the Student-\(t\) distribution, whereas the parameter related to the arrivals of new information, \(\delta\), tend to be not significant under the Gram-Charlier fitting. This feature shows that the driving parameters of the kurtosis are particularly sensitive to the model assumptions that must capture the actual tail-behavior of the underlying distribution. Second, the estimated dynamics of the conditional kurtosis of the Treasure Notes time series differ remarkably from those estimated for the SP and the FX series. This evidence is in sharp contrast to the dynamics followed by the variance, which tend to show the same type of pattern across fairly different classes of financial assets.

Both features seem to suggest that, whereas the dynamics of the conditional variance can always be characterized by ‘stylized features’ (a small estimated \(\alpha\) coefficient, and high persistence as measured by the estimated term \(\alpha + \beta\)), the dynamics of the conditional kurtosis may exhibit a much more idiosyncratic behavior and vary across the class of financial asset and sample period considered. This empirical observation should not be very surprising, since the dynamics of the conditional kurtosis are strongly related to the likelihood and the magnitude of extreme observations (i.e., outliers), which in turn are known to show a large degree of heterogeneity and irregular behavior. The main implication is that, whereas GARCH models tend to yield similar estimation outcomes regardless the financial time-series and the market considered, conditional-kurtosis modelling may yield quite different results depending on the asset considered, and the relevant assumption about the tail-behaviour of the conditional distribution.

### 3.2.2. Out-of-sample analysis

One-step forecasts of conditional variance from the Gaussian GARCH, Student-\(t\) GARCH, and Student-\(t\) GARCHK models are easily obtained as

\[
\hat{h}_{T+1} = \hat{E}_T(h_{T+1}) = \hat{\omega}_T + \hat{\alpha}_T \hat{\varepsilon}_T^2 + \hat{\beta}_T \hat{h}_T. \tag{18}
\]

For the Gram-Charlier GARCHK model, the forecast \(\hat{h}_{T+1}\) is obtained as

\[
\hat{h}_{T+1} = \left(\hat{\omega}_T + \hat{\alpha}_T \hat{\varepsilon}_T^2 + \hat{\beta}_T \hat{h}_T\right) \left[1 + \frac{216\hat{d}_4^2}{1 + 24\hat{d}_4^2}\right] ; \quad \hat{d}_4 = \frac{\hat{k}_{T+1} - 3}{4!}. \tag{19}
\]

For the Gaussian GARCH model, the kurtosis is constant and equals 3, whilst for the Student-\(t\) GARCH model the kurtosis forecast is given by \(\hat{k}_{T+1} = 3(\hat{\nu}_T - 2)/(\hat{\nu}_T - 4)\), being \(\hat{\nu}_T\) the degrees of freedom parameter estimated in the in-sample period. The forecasts of the conditional kurtosis of the Student-\(t\) GARCH models are simply given by

\[
\hat{k}_{T+1} = \hat{E}_T(k_{T+1}) = \hat{\tau}_T + \hat{\delta}_T \left(\frac{\hat{\varepsilon}_T^2}{\hat{h}_T}\right)^2 + \hat{\theta}_T \hat{k}_T \tag{20}
\]

while the Gram-Charlier GARCHK determines a conditional forecast given by:

\[
\hat{k}_{T+1} = \frac{(3 + 2952\hat{d}_4^2 + 12\hat{d}_4^2)(1 + 24\hat{d}_4^2)}{(1 + 216\hat{d}_4^2)^2} \tag{21}
\]
The MSEs for the GARCH models used to forecast conditional variance and kurtosis are presented in Table 3, while Table 4 shows the p-values related to the Diebold-Mariano test (a value smaller than 0.05 implies that the model with smallest MSE in the comparison yields a significant improvement at the 95% confidence level). Some comments follow. We note that the differences in the MSE loss functions for volatility forecasting are not generally significant across the different GARCH models considered in all the series analyzed. This is not surprising, since Gaussian GARCH forecasts are known to be accurate in mean from the property of consistency (discussed in Section 2), and because the dynamics of the conditional kurtosis are modelled independently of the dynamics followed by the conditional variance, we should expect no interaction between them. Therefore, considering further dynamics in the higher-order moments, or allowing for excess of kurtosis in the conditional distribution, would hardly improve empirically the on-average accuracy of the variance forecasts made by a simple Gaussian GARCH model.

In relation to (un)conditional kurtosis, the main results of our analysis are the following. First, the model that yields better out-of-sample forecast of the unconditional kurtosis given the proxy \( \bar{k}_{T+1}^{(500)} \), consistently across the three series considered, is the Gram-Charlier GARCHK model. Owing to its semi-parametric nature, the Gram-Charlier type modelling does not rely upon a specific assumption on the underlying distribution of the data, which provides robustness against potential departures over the parametric models (which, on the other hand, would be consistent and more efficient under correct specification). As we have seen from the empirical results for the conditional variance, robustness turns out to be a precious property when making predictions, and of course this property also applies when considering higher-order moments. Second, and related to the previous consideration, we observe that the Gram-Charlier GARCHK model largely over performs the Student-\( t \) GARCHK model in forecasting conditional kurtosis, as proxied for small values of \( m \) in \( \bar{k}_{T+1}^{(m)} \). Overall, these findings suggest that the assumption of a Student-\( t \) distribution with time-varying degrees of freedom may not be appropriate for applied purposes related to conditional-kurtosis forecasting.

Finally, we can only observe mixed and somewhat inconclusive evidence regarding the empirical importance of modelling time-varying kurtosis, since only in the case of the TN time series there seems to be statistical improvements over the simplest Gaussian GARCH model, and only when using the Gram-Charlier GARCHK specification. There are several reasons that may explain, at least partially, the seeming failure of the conditional-kurtosis GARCH models in the SP and FX time series. First, the presence of measurement errors in the proxy considered \( k_{T+1}^{(m)} \) may end up playing a significant role in the MSE loss-function (particularly as \( m \to 1 \)) given the particularities of the time-series involved, and leading to distorted empirical conclusions. Second, although the conditional-kurtosis GARCH models may provide a better fit in the in-sample period, this does not necessarily imply that these models have to improve the out-of-sample forecast performance. There are two different reasons supporting this statement, both of them being rooted in the high degree of heterogeneity and idiosyncratic behavior of the (conditional) kurtosis. On the one hand, Korkie, Sivakumar and Turtle (2006) have argued that the persistence in the higher-order moments of financial returns may be a statistical artifact related to variance spillovers, so there would not be any gain from forecasting these dynamics. If this pervasive effect exists, it may be more important for some variables than for others, as we have
Are the High-Order Moments of the Assets Returns Distribution Forecastable? 15
documented statistical gains from modelling kurtosis in the case of the TN time series. On
the other hand, even if the conditional kurtosis does really change over time, its dynamics
are necessarily linked to the particularities of the data generating process that drives extreme
observations and irregular outliers. This feature brings up further statistical concerns,
because the time-varying kurtosis process has to be characterized empirically by finite-
sample ML estimates that are strongly conditioned by the assumption on the underlying
distribution (as we have seen in the previous section) and which, furthermore, may suffer
from important biases related to the occurrence and magnitude of outliers in the in-sample
period. A few irregular, large enough outliers are perfectly able to strongly bias the ML
estimates of the conditional kurtosis in the attempt to provide the best possible in-sample fit,
given the underlying assumption that determines the theoretical likelihood and magnitude
of extreme observations, but at the logical cost of poorly forecasting on-average the out-of-
sample dynamics in which such extreme observations do not occur.

Table 3. Out-of-sample volatility and kurtosis MSE forecasting performance

<table>
<thead>
<tr>
<th></th>
<th>A: Gaussian GARCH</th>
<th>C: Student-t GARCHK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{k}_{T+1}^{(m)} )</td>
<td>( \bar{k}_{T+1}^{(m)} )</td>
</tr>
<tr>
<td></td>
<td>( h_{T+1} ) 5 50 500</td>
<td>( h_{T+1} ) 5 50 500</td>
</tr>
<tr>
<td>SP</td>
<td>37.88 5.627 0.221 2.177</td>
<td>37.81 14.08 7.662 1.698</td>
</tr>
<tr>
<td>FX</td>
<td>0.133 22.17 1.266 0.771</td>
<td>0.133 34.74 19.31 10.75</td>
</tr>
<tr>
<td>TN</td>
<td>0.719 5.321 0.717 5.847</td>
<td>0.717 31.06 24.46 13.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B: Student-t GARCH</th>
<th>D: Gram-Charlier GARCHK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{k}_{T+1}^{(m)} )</td>
<td>( \bar{k}_{T+1}^{(m)} )</td>
</tr>
<tr>
<td></td>
<td>( h_{T+1} ) 5 50 500</td>
<td>( h_{T+1} ) 5 50 500</td>
</tr>
<tr>
<td>SP</td>
<td>37.53 14.60 8.201 1.937</td>
<td>37.97 6.305 0.149 0.643</td>
</tr>
<tr>
<td>FX</td>
<td>0.133 34.64 19.07 10.56</td>
<td>0.133 21.90 1.229 0.011</td>
</tr>
<tr>
<td>TN</td>
<td>0.715 17.97 10.52 1.653</td>
<td>0.717 6.184 0.552 3.017</td>
</tr>
</tbody>
</table>

This table shows the Mean Square Error (MSE) for the one-step ahead conditional variance and
kurtosis forecasts from the different GARCH models used in the analysis. The proxies considered
for the kurtosis, \( \bar{k}_{T+1,m} \), are estimated from the sample kurtosis of the first \( m = 5, 50 \) and 500 days
immediately following the last day in the in-sample window.

4. Concluding Remarks

Several papers have argued that the kurtosis of returns may exhibit clusters and time
dependency similar to the characteristic patterns which are observable in the proxies of
conditional variance. The modelling of the conditional third- and fourth-order moments
tend to improve the in-sample goodness of fit over the simplest GARCH models that assume
constant higher-order moments. The main aim of this paper is to provide better insight on
whether accounting for time-varying kurtosis is valuable for out-of-sample forecasting of
Table 4. Diebold and Mariano statistics

<table>
<thead>
<tr>
<th></th>
<th>GARCH-t</th>
<th>GARCHK-t</th>
<th>GARCHK-GC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_{T+1}$</td>
<td>$k_{T+1}^{(m)}$</td>
<td>$h_{T+1}$</td>
</tr>
<tr>
<td></td>
<td>5 50 500</td>
<td></td>
<td>5 50 500</td>
</tr>
<tr>
<td>SP</td>
<td>0.05 0.00 0.00 0.00</td>
<td>0.32 0.00 0.00 0.00</td>
<td>0.22 0.00 0.00 0.00</td>
</tr>
<tr>
<td>GARCH-n</td>
<td>0.00 0.00 0.00 0.00</td>
<td>0.01 0.00 0.00 0.00</td>
<td>0.07 0.00 0.00 0.00</td>
</tr>
<tr>
<td>GARCH-t</td>
<td>0.28 0.00 0.00 0.00</td>
<td>0.21 0.00 0.00 0.00</td>
<td>0.06 0.29 0.00 0.00</td>
</tr>
<tr>
<td>GARCHK-t</td>
<td>0.34 0.13 0.09 0.08</td>
<td>0.49 0.00 0.00 0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.44 0.00 0.00 0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FX</td>
<td>0.03 0.00 0.00 0.00</td>
<td>0.09 0.00 0.00 0.07</td>
<td>0.01 0.00 0.00 0.00</td>
</tr>
<tr>
<td>GARCH-n</td>
<td>0.13 0.02 0.01 0.01</td>
<td>0.09 0.00 0.00 0.00</td>
<td></td>
</tr>
<tr>
<td>GARCH-t</td>
<td>0.35 0.00 0.00 0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCHK-t</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both conditional variance and conditional kurtosis, and which procedure (among several of the parametric and semi-parametric alternatives that have been suggested in the literature) is better suited for empirical purposes.

This table reports the results of the DM test for the difference of the MSE loss function from the GARCH models under analysis (see notation in Table 2). The entries are DM test $p$-values for the predictive ability of the model in the row versus the model in the column.

As in the previous literature, our empirical results on three different classes of financial assets confirm that the semi-parametric Gram-Charlier and the Student-$t$ GARCHK models that allow for time-varying kurtosis provide a better in-sample goodness-of-fit over the constant-kurtosis GARCH models, with the parametric Student-$t$ distribution slightly overperforming the Gram-Charlier type distribution. For forecasting purposes, the best procedure to forecast the unconditional kurtosis seems to be the Gram-Charlier GARCHK model, as the semi-parametric nature of this approach provides robust properties against model misspecification which may ruin the out-of-sample forecasting ability of the model. Similarly, this methodology largely overperforms the Student-$t$ distribution with time-varying degrees of freedom parameter in forecasting conditional kurtosis, which overall suggests that the semi-parametric approximation may be better indicated in practice. Unfortunately, the Gram-Charlier GARCHK model does not always achieve a significant success in beating the forecasts made by the simplest Gaussian GARCH model, at least given the proxies for conditional kurtosis considered in this paper. The lack of conclusive results for some of the time series analyzed may be, at least partially, a statistical artifact due to the sizeable measurement errors in the proxies used in the analysis. However, it is also possible that the empirical success of forecasting higher-order moments strongly depends on the class of asset and the sample period considered, given that the data generating process of the conditional kurtosis does not seem to exhibit the same degree of parameter uniformity as, for instance, the conditional variance does: whereas we always observe the
same sort of stylized features in the GARCH-type estimates of the conditional variance, the conditional kurtosis exhibits a large degree of idiosyncratic behavior. Hence, the econometric modelling allowing for time-varying kurtosis may not generally necessarily enhance the out-of-sample forecasting performance of the models, even if in-sample results seem to suggest the opposite. More research on the empirical role of the dynamics of the conditional kurtosis seems deserved.

Acknowledgement

Financial support from the Spanish Ministry of Education through project SEJ2006-06104/ECON is gratefully acknowledged.

References


REVIEWER’S NAME AND AFFILIATION

Professor Tibor Neugebauer
Faculty of Law, Economics and Finance
University of Luxembourg
Campus Kirchberg 4, rue Albert Borschette (Room K2B2 2.13)
L-1246 Luxembourg
Telephone: +352466644-6285
Fax: +3524666446835
Email: Tibor.Neugebauer@uni.lu