JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **27**: 877–906 (2012)
Published online 17 March 2011 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1234

REALIZED GARCH: A JOINT MODEL FOR RETURNS AND REALIZED MEASURES OF VOLATILITY

PETER REINHARD HANSEN,[a,b]* ZHUO HUANG[c] AND HOWARD HOWAN SHEK[a,d]

[a] *Department of Economics, Stanford University, Stanford, CA, USA*
[b] *CREATES, Aarhus, Denmark*
[c] *China Center for Economic Research, National School of Development, Peking University, Beijing, China*
[d] *iCME, Stanford University, Stanford, CA, USA*

SUMMARY

We introduce a new framework, Realized GARCH, for the joint modeling of returns and realized measures of volatility. A key feature is a *measurement equation* that relates the realized measure to the conditional variance of returns. The measurement equation facilitates a simple modeling of the dependence between returns and future volatility. Realized GARCH models with a linear or log-linear specification have many attractive features. They are parsimonious, simple to estimate, and imply an ARMA structure for the conditional variance and the realized measure. An empirical application with Dow Jones Industrial Average stocks and an exchange traded index fund shows that a simple Realized GARCH structure leads to substantial improvements in the empirical fit over standard GARCH models that only use daily returns.

Supporting information may be found in the online version of this article.

1. INTRODUCTION

The latent volatility process of asset returns is relevant for a wide variety of applications, such as option pricing and risk management, and generalized autoregressive conditional heteroskedasticity (GARCH) models are widely used to model the dynamic features of volatility. This has sparked the development of a large number of ARCH and GARCH models since the seminal paper by Engle (1982). Within the GARCH framework, the key element is the specification for conditional variance. Standard GARCH models utilize daily returns (typically squared returns) to extract information about the current level of volatility, and this information is used to form expectations about the next period's volatility. A single return only offers a weak signal about the current level of volatility. The implication is that GARCH models are poorly suited for situations where volatility changes rapidly to a new level. The reason is that a GARCH model is slow at 'catching up' and it will take many periods for the conditional variance (implied by the GARCH model) to reach its new level, as discussed in Andersen *et al*. (2003).

High-frequency financial data are now widely available and the literature has recently introduced a number of realized measures of volatility, including realized variance, bipower variation, the realized kernel, and many related quantities (see Andersen and Bollerslev, 1998; Andersen *et al*., 2001; Barndorff-Nielsen and Shephard, 2002, 2004; Andersen *et al*., 2008; Barndorff-Nielsen *et al*., 2008; Hansen and Horel, 2009; and references therein). Any of these measures is far more informative about the current level of volatility than is the squared return. This makes realized measures very useful for modeling and forecasting future volatility. Estimating a GARCH model

* Correspondence to: Peter Reinhard Hansen, Department of Economics, Stanford University, 579 Serra all, Stanford, CA 94305-6072, USA. E-mail: peter.hansen@stanford.edu

that includes a realized measure in the GARCH equation (known as a GARCH-X model) provides a good illustration of this point. Such models were estimated by Engle (2002), who used the realized variance (see also Forsberg and Bollerslev, 2002). Within the GARCH-X framework no effort is paid to explain the variation in the realized measures, so these GARCH-X models are partial (incomplete) models that have nothing to say about returns and volatility beyond a single period into the future.

Engle and Gallo (2006) introduced the first 'complete' model in this context. Their model specifies a GARCH structure for each of the realized measures, so that an additional latent volatility process is introduced for each realized measure in the model. The model by Engle and Gallo (2006) is known as the multiplicative error model (MEM), because it builds on the MEM structure proposed by Engle (2002). Another complete model is the HEAVY model by Shephard and Sheppard (2010), which, in terms of its mathematical structure, is nested in the MEM framework. Unlike the traditional GARCH models, these models operate with multiple latent volatility processes. For instance, the MEM by Engle and Gallo (2006) has a total of three latent volatility processes and the HEAVY model by Shephard and Sheppard (2010) has two (or more) latent volatility processes. Within the context of stochastic volatility models, Takahashi *et al*. (2009) proposed a joint model for returns and a realized measure of volatility. Importantly, the economic and statistical gains from incorporating realized measures in volatility models are typically found to be large (see, for example, Christoffersen *et al*., 2010; Dobrev and Szerszen, 2010).

In this paper we introduce a new framework that combines a GARCH structure for returns with an integrated model for realized measures of volatility. Models within our framework are called Realized GARCH models, a name that transpires both the objective of these models (similar to GARCH) and the means by which these models operate (using realized measures).

To illustrate our framework and fix ideas, consider a canonical version of the Realized GARCH model that will be referred to as the RealGARCH(1,1) model with a linear specification. This model is given by the following three equations:

$$
\begin{aligned}
r_t &= \sqrt{h_t} z_t, \\
h_t &= \omega + \beta h_{t-1} + \gamma x_{t-1}, \\
x_t &= \xi + \varphi h_t + \tau(z_t) + u_t
\end{aligned}
$$

where r_t is the return, x_t a realized measure of volatility, $z_t \sim$ i.i.d.$(0, 1)$, $u_t \sim$ i.i.d.$(0, \sigma_u^2)$, and $h_t = \text{var}(r_t|\mathcal{F}_{t-1})$ with $\mathcal{F}_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, \ldots)$. The last equation relates the observed realized measure to the latent volatility and is therefore called the measurement equation. This equation is natural when x_t is a consistent estimator of the integrated variance, because the integrated variance may be viewed as the conditional variance plus a random innovation. The latter is, in our model, absorbed by $\tau(z_t) + u_t$. It is easy to verify that h_t is an autoregressive process of order one, $h_t = \mu + \pi h_{t-1} + w_{t-1}$, where $\mu = \omega + \gamma\xi$, $\pi = \beta + \varphi\gamma$, and $w_t = \gamma\tau(z_t) + \gamma u_t$. Thus it is natural to adopt the nomenclature of GARCH models. The inclusion of the realized measure in the model and the fact that x_t has an autoregressive moving average (ARMA) representation motivate the name Realized GARCH. A simple yet potent specification of the leverage function is $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$, which can generate an asymmetric response in volatility to return shocks. The simple structure of the model makes the model easy to estimate and interpret, and leads to a tractable analysis of the quasi maximum likelihood estimator. The framework allows us to use a realized measure that is computed from a shorter period (e.g. 6.5 hours) than the interval that the conditional variance refers to (e.g. 24 hours). In such instances we should expect $\varphi < 1$.

We apply the Realized GARCH framework to Dow Jones Industrial Average (DJIA) stocks and an exchange traded index fund, SPY. We find, in all cases, substantial improvements in the log-likelihood function (both in-sample and out-of-sample) when benchmarked to a standard GARCH model. This is not too surprising, because the standard GARCH model is based on a more limited information set that only includes daily returns. The empirical evidence strongly favors inclusion of the leverage function, and the parameter estimates are remarkably similar across stocks.

The paper is organized as follows. Section 2 introduces the Realized GARCH framework as a natural extension to GARCH. We focus on linear and log-linear specification and show that squared returns, the conditional variance, and realized measures have ARMA representations in this class of Realized GARCH models. Our Realized GARCH framework is compared to MEM and related models in Section 3. Likelihood-based inference is analyzed in Section 4, where we derive the asymptotic properties of the quasi-maximum likelihood estimator (QMLE). Our empirical analysis is given in Section 5. We estimate a range of Realized GARCH models using time series for 28 stocks and an exchange-traded index fund. In Section 6 we derive results related to forecasting and the skewness and kurtosis of returns over one or more periods. The latter shows that the Realized GARCH is capable of generating substantial skewness and kurtosis. Concluding remarks are given in Section 7, and the Appendix presents all proofs.

2. REALIZED GARCH

In this section we introduce the Realized GARCH model. The key variable of interest is the conditional variance, $h_t = \mathrm{var}(r_t|\mathcal{F}_{t-1})$, where $\{r_t\}$ is a time series of returns. In the GARCH(1,1) model the conditional variance, h_t, is a function of h_{t-1} and r_{t-1}^2. In the present framework, h_t will also depend on x_{t-1}, which represents a realized measure of volatility, such as the realized variance. More generally, x_t will denote a vector of realized measures, such as the realized variance, bipower variation, intraday range, and squared return. A measurement equation, which ties the realized measure to the latent volatility, 'completes' the model. Thus the Realized GARCH model fully specifies the dynamic properties of both returns and the realized measure.

To simplify the exposition we will assume $\mathrm{E}(r_t|\mathcal{F}_{t-1}) = 0$. A more general specifications for the conditional mean, such as a constant or the GARCH-in-mean by Engle *et al*. (1987), is accommodated by reinterpreting r_t as the return less its conditional mean. The general framework for the Realized GARCH model is presented next.

2.1. The General Formulation

The general structure of the RealGARCH(p, q) model is given by

$$r_t = \sqrt{h_t} z_t \tag{1}$$

$$h_t = v(h_{t-1}, \dots, h_{t-p}, x_{t-1}, \dots, x_{t-q}) \tag{2}$$

$$x_t = m(h_t, z_t, u_t) \tag{3}$$

where $z_t \sim$ i.i.d.(0, 1) and $u_t \sim$ i.i.d.$(0, \sigma_u^2)$, with z_t and u_t being mutually independent.

We refer to the first two equations as the *return equation* and the *GARCH equation*, and these define a class of GARCH-X models, including those that were estimated by Engle (2002), Barndorff-Nielsen and Shephard (2007), and Visser (2011). The GARCH-X acronym refers to the the fact that x_t is treated as an exogenous variable. The HYBRID GARCH framework by Chen *et al*. (2009) includes variants of the GARCH-X models and some related models.

DOI: 10.1002/jae

We shall refer to (3) as the *measurement equation*, because the realized measure, x_t, can often be interpreted as a measurement of h_t. The simplest example of a measurement equation is $x_t = h_t + u_t$. The measurement equation is an important component because it 'completes' the model. Moreover, the measurement equation provides a simple way to model the joint dependence between r_t and x_t, which is known to be empirically important. This dependence is modeled though the presence of z_t in the measurement equation, which we find to be highly significant in our empirical analysis.

It is worth noting that most (if not all) variants of ARCH and GARCH models are nested in the Realized GARCH framework. See Bollerslev (2009) for a comprehensive list of such models. The nesting can be achieved by setting $x_t = r_t$ or $x_t = r_t^2$, and the measurement equation is redundant for such models, because it is reduced to a simple identity. Naturally, the interesting case is when x_t is a high-frequency-based realized measure, or a vector containing several realized measures. Next we consider some particular variants of the Realized GARCH model.

2.2. Realized GARCH with a Log-Linear Specification

The Realized GARCH model with a simple log-linear specification is characterized by the following GARCH and measurement equations:

$$\log h_t = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + \sum_{j=1}^{q} \gamma_j \log x_{t-j} \tag{4}$$

$$\log x_t = \xi + \varphi \log h_t + \tau(z_t) + u_t \tag{5}$$

where $z_t = r_t/\sqrt{h_t} \sim$ i.i.d.$(0, 1)$, $u_t \sim$ i.i.d.$(0, \sigma_u^2)$, and $\tau(z)$ is called the *leverage function*. Without loss of generality we assume $\mathrm{E}\tau(z_t) = 0$.

A logarithmic specification for the measurement equation seems natural in this context. The reason is that (1) implies that

$$\log r_t^2 = \log h_t + \log z_t^2 \tag{6}$$

and a realized measure is in many ways similar to the squared return, r_t^2, albeit a more accurate measure of h_t. It is therefore natural to explore specifications where $\log x_t$ is expressed as a function of $\log h_t$ and z_t, such as (5). A logarithmic form for the measurement equation makes it convenient to specify the GARCH equation with a logarithmic form, because this induces a convenient ARMA structure.

In our empirical application we adopt a quadratic specification for the leverage function, $\tau(z_t)$. The conditional variance, h_t, is adapted to $\mathcal{F}_{t-1}$. Therefore $\mathcal{F}_t$ must be such that $x_t \in \mathcal{F}_t$ (unless $\gamma = 0$). This requirement is satisfied by $\mathcal{F}_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, \ldots)$, but $\mathcal{F}_t$ could in principle be an even richer σ-field. Also, note that the measurement equation does not require x_t to be an unbiased measure of h_t. For instance, x_t could be a realized measure that is computed with 6.5 hours of high-frequency data, while the return is a close-to-close return that spans 24 hours.

An attractive feature of the log-linear Realized GARCH model is that it preserves the ARMA structure that characterizes some of the standard GARCH models. This shows that the 'ARCH' nomenclature is appropriate for the Realized GARCH model. For the sake of generality we derive the result for the case where the GARCH equation includes lagged squared returns. Thus consider the following GARCH equation:

$$\log\ h_t = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + \sum_{j=1}^{q} \gamma_j \log x_{t-j} + \sum_{j=1}^{q} \alpha_j \log r_{t-j}^2 \tag{7}$$

where $q = \max_i\{(\alpha_i, \gamma_i) \neq (0, 0)\}$.

DOI: 10.1002/jae

Proposition 1. Define $w_t = \tau(z_t) + u_t$ and $v_t = \log z_t^2 - \kappa$, where $\kappa = \mathrm{E} \log z_t^2$. The Realized GARCH model defined by (5) and (7) implies

$$\log\ h_t = \mu_h + \sum_{i=1}^{p\vee q}(\alpha_i + \beta_i + \varphi\gamma_i)\log h_{t-i} + \sum_{j=1}^{q}(\gamma_j w_{t-j} + \alpha_j v_{t-j}),$$

$$\log\ x_t = \mu_x + \sum_{i=1}^{p\vee q}(\alpha_i + \beta_i + \varphi\gamma_i)\log x_{t-i} + w_t + \sum_{j=1}^{p\vee q}\{-(\alpha_j + \beta_j)w_{t-j} + \varphi\alpha_j v_{t-j}\},$$

$$\log\ r_t^2 = \mu_r + \sum_{i=1}^{p\vee q}(\alpha_i + \beta_i + \varphi\gamma_i)\log r_{t-i}^2 + v_t + \sum_{j=1}^{p\vee q}\{\gamma_i(w_{t-j} - \varphi v_{t-j}) - \beta_j v_{t-j}\}$$

where $\mu_h = \omega + \gamma_\bullet\xi + \alpha_\bullet\kappa$, $\mu_x = \varphi(\omega + \alpha_\bullet\kappa) + (1 - \alpha_\bullet - \beta_\bullet)\xi$, and $\mu_r = \omega + \gamma_\bullet\xi + (1 - \beta_\bullet - \varphi\gamma_\bullet)\kappa$, with $\alpha_\bullet = \sum_{j=1}^{q}\alpha_j$, $\beta_\bullet = \sum_{i=1}^{p}\beta_i$, and $\gamma_\bullet = \sum_{j=1}^{q}\gamma_j$, and the conventions $\beta_i = \gamma_j = \alpha_j = 0$ for $i > p$ and $j > q$.

Thus the log-linear Realized GARCH model implies that $\log h_t$ is ARMA($p \vee q, q-1$), whereas $\log r_t^2$ and $\log x_t$ are ARMA($p \vee q, p \vee q$). If $\alpha_1 = \ldots = \alpha_q = 0$, then $\log x_t$ is ARMA($p \vee q, p$).

From Proposition 1 we see that the persistence of volatility is summarized by a *persistence parameter*:

$$\pi = \sum_{i=1}^{p\vee q}(\alpha_i + \beta_i + \varphi\gamma_i) = \alpha_\bullet + \beta_\bullet + \varphi\gamma_\bullet$$

Example 1. For the case $p = q = 1$ we have $\log h_t = \omega + \beta\log h_{t-1} + \gamma\log x_{t-1}$ and $\log x_t = \xi + \varphi\log h_t + \tau(z_t) + u_t$, so that $\log h_t \sim$ AR(1) and $\log x_t \sim$ ARMA(1,1). Specifically $\log h_t = \mu_h + \pi\log h_{t-1} + \gamma w_{t-1}$ and $\log x_t = \mu_x + \pi\log x_{t-1} + w_t - \beta w_{t-1}$, where $\pi = \beta + \varphi\gamma$.

The measurement equation induces a GARCH structure that is similar to an EGARCH with a stochastic volatility component. Take the case in Example 1 where $\log h_t = \mu_h + \pi\log h_{t-1} + \gamma\tau(z_{t-1}) + \gamma u_{t-1}$. Note that $\gamma\tau(z_{t-1})$ captures the leverage effects, whereas γu_{t-1} adds an additional stochastic component that resembles that of stochastic volatility models. Thus the Realized GARCH model can induce a flexible stochastic volatility structure, similar to that in Yu (2008), but does in fact have a GARCH structure because u_{t-1} is $\mathcal{F}_{t-1}$-measurable. Interestingly, for the purpose of forecasting (beyond one-step-ahead predictions), the Realized GARCH is much like a stochastic volatility model since future values of u_t are unknown. This analogy does not apply to one-step-ahead predictions because the lagged values, $\tau(z_{t-1})$ and u_{t-1}, are known at time $t-1$.

An obvious advantage of using a logarithmic specification is that it automatically ensures a positive variance. Here it should be noted that the GARCH model with a logarithmic specification, known as LGARCH (see Geweke, 1986; Pantula, 1986; Milhø j, 1987), has some practical drawbacks. These drawbacks may explain that the LGARCH is less popular in applied work than the conventional GARCH model that uses a specification for the level of volatility (see Teräsvirta, 2009). One drawback is that zero returns are occasionally observed, and will cause havoc for the log-specification unless we impose some ad hoc censoring. Within the Realized GARCH framework, zero returns are not problematic, because $\log r_{t-1}^2$ does not appear in its GARCH equation.

DOI: 10.1002/jae

2.2.1 The Leverage Function

The function $\tau(z)$ is called the leverage function because it captures the dependence between returns and future volatility, a phenomenon that is referred to as the *leverage effect*. We normalize such functions by $\mathrm{E}\tau(z_t) = 0$, and we focus on those that have the form

$$\tau(z_t) = \tau_1 a_1(z_t) + \ldots + \tau_k a_k(z_t), \qquad \text{where} \quad \mathrm{E}a_k(z_t) = 0, \quad \text{for all } k$$

so that the function is linear in the unknown parameters. We shall see that the leverage function induces an EGARCH type structure in the GARCH equation, and we note that the functional form used in Nelson (1991), $\tau(z_t) = \tau_1 z + \tau_+(|z_t| - \mathrm{E}|z_t|)$, is within this class of leverage functions. In this paper we focus on leverage functions that are constructed from Hermite polynomials, i.e.

$$\tau(z) = \tau_1 z + \tau_2(z^2 - 1) + \tau_3(z^3 - 3z) + \tau_4(z^4 - 6z^2 + 3) + \ldots$$

and our baseline choice for the leverage function is a simple quadratic form: $\tau(z_t) = \tau_1 z_t + \tau_2(z_t^2 - 1)$. This choice is convenient because it ensures that $\mathrm{E}\tau(z_t) = 0$, for any distribution with $\mathrm{E}z_t = 0$ and $\mathrm{var}(z_t) = 1$. The polynomial form is also convenient in our quasi-likelihood analysis, and in our derivations of the kurtosis of returns generated by this model.

The leverage function $\tau(z)$ is closely related to the *news impact curve* (see Engle and Ng, 1993), which maps out how positive and negative shocks to the price affect future volatility. We can define the news impact curve by $\nu(z) = \mathrm{E}(\log h_{t+1}|z_t = z) - \mathrm{E}(\log h_{t+1})$, so that $100\nu(z)$ measures the percentage impact on volatility as a function of the Studentized return. From the ARMA representation in Proposition 1 it follows that $\nu(z) = \gamma_1\tau(z)$.

2.3. Realized GARCH with a Linear Specification

In this section we adopt a linear structure that is more similar to the original GARCH model by Bollerslev (1986). One advantage of this formulation is that the measurement equation is simple to interpret in this model. For instance, if x_t is computed from intermittent high-frequency data (i.e. over 6.5 hours) whereas r_t is a close-to-close return that spans 24 hours, then we would expect φ to reflect how much of the daily volatility occurs during trading hours. The linear Realized GARCH model is defined by

$$x_t h_t = \omega + \sum\nolimits_{i=1}^{p} \beta_i h_{t-i} + \sum\nolimits_{j=1}^{q} \gamma_j x_{t-j}, \quad \text{and} \quad = \xi + \varphi h_t + \tau(z_t) + u_t$$

As is the case for the GARCH(1,1) model the RealGARCH(1,1) model with the linear specification implies that h_t has an AR(1) representation $h_t = (\omega + \gamma\xi) + (\beta + \gamma\varphi)h_{t-1} + \gamma w_{t-1}$, where $w_t = u_t + \tau(z_t)$ is an i.i.d. process, and that the realized measure, x_t, is ARMA(1,1), which is consistent with the time series properties of realized measures in this context (see Meddahi, 2003).

3. COMPARISON TO RELATED MODELS

In this section we relate the Realized GARCH model to the multiplicative error model (MEM) by Engle and Gallo (2006) and the HEAVY model by Shephard and Sheppard (2010),[1] and some related approaches.

[1] The Realized GARCH model was conceptualized and developed concurrently and independently of Shephard and Sheppard (2010). However, in our current presentation of the model we have adopted some terminology from Shephard and Sheppard (2010).

DOI: 10.1002/jae

The MEM by Engle and Gallo (2006) utilizes two realized measures in addition to the squared returns. These are the intraday range (high minus low) and the realized variance, whereas the HEAVY model by Shephard and Sheppard (2010) uses the realized kernel (RK) by Barndorff-Nielsen *et al*. (2008). These models introduce an additional latent volatility process for each of the realized measures. Thus the MEM and the HEAVY digress from the traditional GARCH models that only have a single latent volatility factor. Key model features are given in Table I. We have included the level specification of the Realized GARCH model because it is most similar to the GARCH, MEM, and HEAVY models. Based on our empirical analysis in Section 5 we recommend the log-linear specification in practice.

Brownless and Gallo (2010) estimates a restricted MEM model that is closely related to the Realized GARCH with the linear specification. They utilize a single realized measure, which leads to two latent volatility processes in their model: $h_t = \mathrm{E}(r_t^2|\mathcal{F}_{t-1})$ and $\mu_t = \mathrm{E}(x_t|\mathcal{F}_{t-1})$. However, their model is effectively reduced to a single-factor model as they introduce the constraint $h_t = c + d\mu_t$.

The usual MEM formulation is based on a vector of non-negative random innovations, η_t, that are required to have mean $\mathrm{E}(\eta_t) = (1, \ldots, 1)'$. The literature has explored distributions with this property such as certain multivariate Gamma distributions, and Cipollini *et al*. (2009) use copula methods that entail a very flexible class of distributions with the required structure. (A perhaps simpler way to achieve this structure is by setting $\eta_t = Z_t \odot Z_t$, and working with the vector of mean-zero unit-variance random variables, Z_t, instead.) The estimates in Engle and Gallo (2006) and Shephard and Sheppard (2010) are based on a likelihood where the elements of η_t are independent χ^2-distributed random variables with one degree of freedom, which maps into $Z_t \sim N(0, I)$. We have used the alternative formulation in Table I so that $(z_t^2, z_{R,t}^2, z_{RV,t}^2)'$ corresponds to η_t in the MEM by Engle and Gallo (2006).

3.1. Multi-factor Realized GARCH Models

The Realized GARCH framework can be extended to a multi-factor structure. For instance, with m realized measures (including the squared return) we could specify a model with $k \leq m$ latent volatility factors. The Realized GARCH model introduced in this paper has $k = 1$, whereas the MEM has $m = k$. This hybrid framework with $1 \leq k \leq m$, provides a way to bridge the Realized GARCH models with the MEM framework. All these models can be viewed as extensions of standard GARCH models, where the extensions are achieved by incorporating realized measures into the model in various ways.[2]

4. QUASI-MAXIMUM LIKELIHOOD ANALYSIS

In this section we discuss the asymptotic properties of the quasi-maximum likelihood estimator within the Realized GARCH (p, q) model. The structure of the QMLE analysis is very similar to that of the standard GARCH model (see Bollerslev and Wooldridge, 1992; Lee and Hansen, 1994; Lumsdaine, 1996; Jensen and Rahbek, 2004a,b; Straumann and Mikosch, 2006). Both Engle and Gallo (2006) and Shephard and Sheppard (2010) justify the standard errors they report, by referencing existing QMLE results for GARCH models. This argument hinges on the fact that the joint log-likelihood in the MEM is decomposed into a sum of univariate GARCH-X models, whose likelihood can be maximized separately. The factorization of the likelihood is achieved

[2] A realized measure simply refers to a statistic that is constructed from high-frequency data. Well-known examples include realized variance, realized kernel, intraday range, number of transactions, and trading volume.

DOI: 10.1002/jae

Table I. Key model features at a glance: the realized measures, R_t, RV_t, and x_t, denote the intraday range, realized variance, and realized kernel, respectively. In the Realized GARCH model, the dependence between returns and innovations to the volatility (leverage effect) is modeled with $\tau(z_t)$, such as $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$, so that $\mathrm{E}\tau(z_t) = 0$, when $z_t \sim (0, 1)$

	Latent variables[a]	Observables	Distribution[b]
GARCH(1,1) (Bollerslev, 1986)	$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$	$r_t = \sqrt{h_t} z_t$	$z_t \sim$ i.i.d.$N(0, 1)$
MEM (Engle and Gallo, 2006)	$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \delta r_{t-1} + \varphi R_{t-1}^2$		
	$h_{R,t} = \omega_R + \alpha_R R_{t-1}^2 + \beta_R h_{R,t-1} + \delta_R r_{t-1}$	$r_t^2 = h_t z_t^2$	
	$h_{RV,t} = \omega_{RV} + \alpha_{RV} RV_{t-1} + \beta_{RV} h_{RV,t-1}$	$R_t^2 = h_{R,t} z_{R,t}^2$	$\begin{pmatrix} z_t \\ z_{R,t} \\ z_{RV,t} \end{pmatrix} \sim$ i.i.d.$N(0, I)$
	$+\delta_{RV} r_{t-1} + \vartheta_{RV} RV_{t-1} 1_{(r_{t-1}<0)} + \varphi_{RV} r_{t-1}^2$	$RV_t = h_{RV,t} z_{RV,t}^2$	
HEAVY (Shephard and Sheppard, 2009)	$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1} + \gamma x_{t-1}$	$r_t = \sqrt{h_t} z_t$	$\begin{pmatrix} z_t \\ z_{RK,t} \end{pmatrix} \sim$ i.i.d.$N(0, I)$
	$\mu_t = \omega_R + \alpha_R x_{t-1} + \beta_R \mu_{t-1}$	$x_t = \mu_t z_{RK,t}^2$	
Realized GARCH (linear specification)	$h_t = \omega + \beta h_{t-1} + \gamma x_{t-1}$	$r_t = \sqrt{h_t} z_t$	$\begin{pmatrix} z_t \\ \frac{u_t}{\sigma_u} \end{pmatrix} \sim$ i.i.d.$N(0, I)$
		$x_t = \xi + \varphi h_t + \tau(z_t) + u_t$	
Realized GARCH (log-linear specification)	$h_t = \exp\{\omega + \beta \log h_{t-1} + \gamma \log x_{t-1}\}$	$r_t = \sqrt{h_t} z_t$	$\begin{pmatrix} z_t \\ \frac{u_t}{\sigma_u} \end{pmatrix} \sim$ i.i.d.$N(0, I)$
		$\log x_t = \xi + \varphi \log h_t + \tau(z_t) + u_t$	
Realized EGARCH (Section 2)	$h_t = \exp\{\omega + \beta \log h_{t-1} + \tau(z_{t-1}) + \delta \varepsilon_{t-1}\}$	$r_t = \sqrt{h_t} z_t$	$\begin{pmatrix} z_t \\ \frac{\varepsilon_t}{\sigma_\varepsilon} \end{pmatrix} \sim$ i.i.d.$N(0, I)$
		$\log x_t = \xi + \log h_{t+1} + \varepsilon_t$	

[a] The MEM specification listed here is that selected by Engle and Gallo (2006) using BIC (see their Table IV). The MEM framework permits more complex specifications.

[b] The distributional assumptions listed here are those used to specify the quasi log-likelihood function. (Gaussian innovations are not essential for any of the models.) The Realized EGARCH is introduced in Section 2.

DOI: 10.1002/jae

by two facets of these models. One is that all observables (i.e. squared return and each of the realized measures) are being tied to their individual latent volatility process. The other is that the primitive innovations in these models are taken to be independent in the formulation of the likelihood function. The latter inhibits a direct modeling of the leverage effect with a function such as $\tau(z_t)$, which is one of the traits of the Realized GARCH model. However, in the MEM framework one can generate a leverage type dependence by including suitable realized measures in various GARCH equations, such as the realized semivariance (see Barndorff-Nielsen *et al.*, 2009b), or by introducing suitable indicator functions as in Engle and Gallo (2006).

In this section we will derive the underlying QMLE structure for the log-linear Realized GARCH model. The structure of the linear Realized GARCH model is similar. We provide closed-form expressions for the first and second derivatives of the log-likelihood function. These expressions facilitate direct computation of robust standard errors, and provide insight about regularity conditions that would justify QMLE inference. For instance, the first derivative will unearth regularity conditions that enable a central limit theorem be applied to the score function.

For the purpose of estimation, we adopt a Gaussian specification, so that the log-likelihood function is given by

$$\ell(r, x; \theta) = -\frac{1}{2}\sum_{t=1}^{n}[\log(h_t) + r_t^2/h_t + \log(\sigma_u^2) + u_t^2/\sigma_u^2]$$

We write the leverage function as $\tau' a_t = \tau_1 a_1(z_t) + \ldots + \tau_k a_k(z_t)$, and denote the parameters in the model by

$$\theta = (\lambda', \psi', \sigma_u^2)', \quad \text{where} \quad \lambda = (\omega, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q)', \quad \psi = (\xi, \varphi, \tau')'$$

To simplify the notation we write $\tilde{h}_t = \log h_t$ and $\tilde{x}_t = \log x_t$, and define $g_t = (1, \tilde{h}_{t-1}, \ldots, \tilde{h}_{t-p}, \tilde{x}_{t-1}, \ldots, \tilde{x}_{t-q})'$ and $m_t = (1, \tilde{h}_t, a_t')'$. Thus the GARCH and measurement equations can be expressed as $\tilde{h}_t = \lambda' g_t$ and $\tilde{x}_t = \psi' m_t + u_t$, respectively. The dynamics that underlie the score and Hessian are driven by h_t and its derivatives with respect to λ. The properties of these derivatives are stated next.

Lemma 1. Define $\dot{h}_t = \frac{\partial \tilde{h}_t}{\partial \lambda}$ and $\ddot{h}_t = \frac{\partial^2 \tilde{h}_t}{\partial \lambda \partial \lambda'}$. Then $\dot{h}_s = 0$ and $\ddot{h}_s = 0$ for $s \leq 0$, and

$$\dot{h}_t = \sum_{i=1}^{p} \beta_i \dot{h}_{t-i} + g_t \qquad \text{and} \qquad \ddot{h}_t = \sum_{i=1}^{p} \beta_i \ddot{h}_{t-i} + (\dot{H}_{t-1} + \dot{H}'_{t-1})$$

where $\dot{H}_{t-1} = (0_{1+p+q\times 1}, \dot{h}_{t-1}, \ldots, \dot{h}_{t-p}, 0_{1+p+q\times q})$ is a $p+q+1 \times p+q+1$ matrix.

Proposition 2.

(i) The score, $\frac{\partial \ell}{\partial \theta} = \sum_{t=1}^{n} \frac{\partial \ell_t}{\partial \theta}$, is given by

$$\frac{\partial \ell_t}{\partial \theta} = -\frac{1}{2}\left\{\left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)\dot{h}_t, -\frac{2u_t}{\sigma_u^2} m_t, \frac{\sigma_u^2 - u_t^2}{\sigma_u^4}\right\}'$$

where $\dot{u}_t = \partial u_t/\partial \log h_t = -\varphi + \frac{1}{2}z_t\tau'\dot{a}_t$ with $\dot{a}_t = \partial a(z_t)/\partial z_t$.

(ii) The second derivative, $\frac{\partial^2 \ell}{\partial\theta\partial\theta'} = \sum_{t=1}^{n} \frac{\partial^2 \ell_t}{\partial\theta\partial\theta'}$, is given by

$$\frac{\partial^2 \ell_t}{\partial\theta\partial\theta'} = \begin{pmatrix} -\frac{1}{2}\left\{z_t^2 + \frac{2(\dot{u}_t^2 + u_t\ddot{u}_t)}{\sigma_u^2}\right\}\dot{h}_t\dot{h}_t' - \frac{1}{2}\left\{1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right\}\ddot{h}_t & \bullet & \bullet \\ \frac{\dot{u}_t}{\sigma_u^2}m_t\dot{h}_t' + \frac{u_t}{\sigma_u^2}b_t\dot{h}_t' & -\frac{1}{\sigma_u^2}m_t m_t' & \bullet \\ \frac{u_t\dot{u}_t}{\sigma_u^4}\dot{h}_t' & \frac{u_t}{\sigma_u^4}m_t' & \frac{1}{2}\frac{\sigma_u^2 - 2u_t^2}{\sigma_u^6} \end{pmatrix}$$

where $b_t = \left(0, 1, -\frac{1}{2}z_t\dot{a}_t'\right)'$ and $\ddot{u}_t = -\frac{1}{4}\tau'\{z_t\dot{a}_t + z_t^2\ddot{a}_t\}$ with $\ddot{a}_t = \partial^2 a(z_t)/\partial z_t^2$.

An advantage of our framework is that we can draw upon results for generalized hidden Markov models. Consider the case $p = q = 1$. From Carrasco and Chen (2002, Proposition 2) it follows that $\tilde{h}_t$ has a stationary representation provided that $\pi = \beta + \varphi\gamma \in (-1, 1)$. If we assign $\tilde{h}_0$ its invariant distribution, then $\tilde{h}_t$ is strictly stationary and β-mixing with exponential decay, and $\mathrm{E}|\tilde{h}_t|^s < \infty$ if $\mathrm{E}|\tau(z_t) + u_t|^s < \infty$. Moreover, $\{(r_t, x_t), t \geq 0\}$ is a generalized hidden Markov model, with hidden chain $\{\tilde{h}_t, t \geq 0\}$, and so by Carrasco and Chen (2002, proposition 4) it follows also that $\{(r_t, x_t)\}$ is stationary β-mixing with an exponential decay rate.[3]

The robustness of the QMLE as defined by the Gaussian likelihood is, in part, reflected by the weak assumptions that make the score a martingale difference sequence. These are stated in the following proposition.

Proposition 3.

(i) Suppose that $\mathrm{E}(u_t|z_t, \mathcal{F}_{t-1}) = 0$, $\mathrm{E}(z_t^2|\mathcal{F}_{t-1}) = 1$, and $\mathrm{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2$. Then $s_t(\theta) = \frac{\partial \ell_t(\theta)}{\partial\theta}$ is a martingale difference sequence.

(ii) Suppose, in addition, that $\{(r_t, x_t, \tilde{h}_t)\}$ is stationary and ergodic. Then

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \ell_t}{\partial\theta} \xrightarrow{d} N(0, \mathcal{J}_\theta) \quad \text{and} \quad -\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \ell_t}{\partial\theta\partial\theta'} \xrightarrow{p} \mathcal{I}_\theta$$

provided that

$$\mathcal{J}_\theta = \begin{pmatrix} \frac{1}{4}E\left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)^2 \mathrm{E}(\dot{h}_t\dot{h}_t') & \bullet & \bullet \\ -\frac{1}{\sigma_u^2}\mathrm{E}(\dot{u}_t m_t \dot{h}_t') & \frac{1}{\sigma_u^2}\mathrm{E}(m_t m_t') & \bullet \\ \frac{-E(u_t^3)E(\dot{u}_t)}{2\sigma_u^6}E(\dot{h}_t') & \frac{E(u_t^3)}{2\sigma_u^6}E(m_t') & \frac{E(u_t^2/\sigma_u^2 - 1)^2}{4\sigma_u^4} \end{pmatrix}$$

[3] See also Straumann and Mikosch (2006), who adopt a stochastic recurrence approach to analyze the QMLE properties for a broad class of GARCH models.

and

$$\mathcal{I}_\theta = \begin{pmatrix} \left\{ \dfrac{1}{2} + \dfrac{\mathrm{E}(\dot{u}_t^2)}{\sigma_u^2} \right\} \mathrm{E}(\dot{h}_t \dot{h}_t') & \bullet & 0 \\ -\dfrac{1}{\sigma_u^2} \mathrm{E}\{(\dot{u}_t m_t + u_t b_t)\dot{h}_t'\} & \dfrac{1}{\sigma_u^2}\mathrm{E}(m_t m_t') & 0 \\ 0 & 0 & \dfrac{1}{2\sigma_u^4} \end{pmatrix}$$

are finite.

Note that in the stationary case we have $\mathcal{J}_\theta = \mathrm{E}\left(\frac{\partial \ell_t}{\partial \theta}\frac{\partial \ell_t}{\partial \theta'}\right)$, so a necessary condition for $|\mathcal{J}_\theta| < \infty$ is that z_t and u_t have finite fourth moments. Additional moments may be required for z_t, depending on the complexity of the leverage function $\tau(z)$, because $\dot{u}_t$ depends on $\tau(z_t)$.

The mathematical structure of the Gaussian quasi log-likelihood function for the Realized GARCH model is quite similar to the structure analyzed in Straumann and Mikosch (2006). Therefore we conjecture that Straumann and Mikosch (2006, Theorem 7.1) can be adapted to the present framework, so that

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, \mathcal{I}_\theta^{-1}\mathcal{J}_\theta\mathcal{I}_\theta^{-1})$$

To make this result rigorous we would need to adapt and verify conditions N.1–N.4 in Straumann and Mikosch (2006). This is not straightforward and would take up much space, so we leave this for future research. Moreover, the results in Straumann and Mikosch (2006) only apply to the stationary case, $\pi < 1$, so the non-stationary case would have to be analyzed separately using methods similar to those in Jensen and Rahbek (2004a,b). For simple ARCH and GARCH models, Jensen and Rahbek (2004a,b) have shown that the QMLE estimator is consistent with a Gaussian limit distribution regardless of the process being stationary or non-stationary. Therefore, unlike the case for autoregressive processes, we need not have a discontinuity of the limit distribution at the knife-edge in the parameter space that separates stationary and non-stationary processes. A similar result may apply to the Realized GARCH model with the linear specification. However, the log-linear specification generates a score function with a structure that may result in convergence to a stochastic integral in the unit root case. We leave this important inference problem for future research.

5. EMPIRICAL ANALYSIS

In this section we present empirical results using returns and realized measures for 28 stocks and and exchange-traded index fund, SPY, that tracks the S&P 500 index. Detailed results are presented for SPY, whereas our results for the other 28 time series are presented with fewer details, to conserve space. We adopt the realized kernel, introduced by Barndorff-Nielsen *et al*. (2008), as the realized measure, x_t. We estimate the realized GARCH models using both open-to-close returns and close-to-close returns. High-frequency prices are only available between ‘open’ and ‘close’, so the population quantity that is estimated by the realized kernel is directly related to the volatility of open-to-close returns, but only captures a fraction of the volatility of close-to-close returns.

We compare the linear and log-linear specifications and argue that the latter is better suited for the problem at hand. So we will mainly present empirical results based on the log-linear specification. We report empirical results for all 29 assets in Table III and find the point estimates to be remarkably similar across the many time series. In-sample and out-of-sample likelihood ratio statistics are computed in Table IV. These results strongly favor the inclusion of the leverage function and show that the realized GARCH framework is superior to standard

DOI: 10.1002/jae

GARCH models, because the partial log-likelihood of any Realized GARCH model is substantially better than that of a standard GARCH(1,1). This is found to be the case in-sample, as well as out-of-sample.

5.1. Data Description

Our sample spans the period from 1 January 2002 to 31 August 2008, which we divide into an in-sample period: 1 January 2002 to 31 December 2007; leaving the eight months 2 January 2008 to 31 August 2008 for out-of-sample analysis. We adopt the realized kernel as the realized measure, x_t, using the Parzen kernel function. This estimator is similar to the well-known realized variance, but is robust to market microstructure noise and is a more accurate estimator of the quadratic variation. Our implementation of the realized kernel follows Barndorff-Nielsen *et al.* (2010) that guarantees a positive estimate, which is important for our log-linear specification. The exact computation is explained in great detail in Barndorff-Nielsen *et al.* (2009a). To avoid outliers that would result from half trading days, we removed days where high-frequency data spanned less than 90% of the official 6.5 hours between 9 : 30 a.m. and 4 : 00 p.m. This removes about three daily observations per year, such as the day after Thanksgiving and days around Christmas. When we estimate a model that involves $\log r_t^2$, we deal with zero returns by the truncation $\max(r_t^2, \varepsilon)$ with $\varepsilon = 10^{-20}$. Summary statistics are available as supporting information in a separate Web Appendix.

5.2. Some Notation Related to the Likelihood and Leverage Effect

The log-likelihood function is (conditionally on $\mathcal{F}_0 = \sigma(\{r_t, x_t, h_t\}, t \leq 0)$) given by

$$\log L(\{r_t, x_t\}_{t=1}^n; \theta) = \sum_{t=1}^n \log f(r_t, x_t | \mathcal{F}_{t-1})$$

Standard GARCH models do not model x_t, so the log-likelihood we obtain for these models cannot be compared to those of the Realized GARCH model. However, we can factorize the joint conditional density for (r_t, x_t) by

$$f(r_t, x_t | \mathcal{F}_{t-1}) = f(r_t | \mathcal{F}_{t-1}) f(x_t | r_t, \mathcal{F}_{t-1})$$

and compare the partial log-likelihood, $\ell(r) := \sum_{t=1}^n \log f(r_t | \mathcal{F}_{t-1})$, with that of a standard GARCH model. Specifically for the Gaussian specification for z_t and u_t, we split the joint likelihood into the sum

$$\ell(r, x) = \underbrace{-\frac{1}{2}\sum_{t=1}^n [\log(2\pi) + \log(h_t) + r_t^2/h_t]}_{=\ell(r)} + \underbrace{-\frac{1}{2}\sum_{t=1}^n [\log(2\pi) + \log(\sigma_u^2) + u_t^2/\sigma_u^2]}_{=\ell(x|r)}$$

Asymmetries in the leverage function are summarized by the two statistics $\rho^- = \text{corr}\{\tau(z_t) + u_t, z_t | z_t < 0\}$ and $\rho^+ = \text{corr}\{\tau(z_t) + u_t, z_t | z_t > 0\}$, which capture the average slope of the news impact curve for negative and positive returns.

DOI: 10.1002/jae

5.3. Empirical Results for the Linear Realized GARCH Model

Detailed empirical results for the linear specification are available as supporting information in a separate Web Appendix. We wish to emphasize one empirical observation that concerns the difference between open-to-close returns and close-to-close returns. With open-to-close SPY returns our estimates for the RealGARCH(1,1) models are

$$h_t = \underset{(0.05)}{0.09} + \underset{(0.16)}{0.29}\, h_{t-1} + \underset{(0.18)}{0.63}\, x_{t-1},$$
$$x_t = -\underset{(0.09)}{0.05} + \underset{(0.19)}{1.01}\, h_t - \underset{(0.02)}{0.02}\, z_t + \underset{(0.01)}{0.06}\, (z_t^2 - 1) + u_t$$

where the numbers in parentheses are standard errors. Note that the empirical estimates of φ and ξ in the measurement equation are consistent with the belief that the realized kernel is roughly an unbiased measurement of (open-to-close) h_t. With close-to-close returns we obtain the following estimates:

$$h_t = \underset{(0.04)}{0.07} + \underset{(0.15)}{0.29}\, h_{t-1} + \underset{(0.25)}{0.87}\, x_{t-1},$$
$$x_t = \underset{(0.08)}{0.00} + \underset{(0.14)}{0.74}\, h_t - \underset{(0.02)}{0.07}\, z_t + \underset{(0.01)}{0.03}\, (z_t^2 - 1) + u_t$$

and it is not surprising that the estimate of φ is less than one in this case. The point estimate of φ suggests that volatility during the 'open period' amounts to about 75% of daily volatility.

5.4. Empirical Results for the Log-Linear Realized GARCH Model

In this section we present detailed results for Realized GARCH models with a log-linear specification of the GARCH and measurement equations. We strongly favor the log-linear specification over the linear specification for reasons that will be evident in Section 5.5, where we compare empirical aspects of the two specifications.

5.4.1 Log-Linear Models for SPY (Table II)

Table II presents out empirical results for the log-linear specification using six variants of the Realized GARCH model. For the sake of comparison we use the logarithmic GARCH(1,1) model as the conventional benchmark when comparing the empirical fit in terms of the partial likelihood function (for returns). The left-hand panel has the empirical results for open-to-close SPY returns and the right-hand panel has the corresponding results for close-to-close SPY returns.

From Table II we see that the extended model RG(2,2)*, which includes the squared return in the GARCH equation, results in very marginal improvements over the standard model RG(2,2), and the ARCH parameter, α, is clearly insignificant. Comparing the RG(2,2)† with the standard model shows that the leverage function is highly significant. The improvement in the log-likelihood function is almost 100 units.

The robust standard errors suggest that β_2 is significant, when it is actually not the case. This is simply a manifestation of a common problem with standard errors and t-statistics in the context with collinearity. In this case, $\log h_{t-1}$ and $\log h_{t-2}$ are highly collinear, which causes the likelihood surface to be almost flat along lines where $\beta_1 + \beta_2$ is constant, while there is sufficient curvature along the axis to make the standard errors small.

The estimates of φ are close to unity, $\hat{\varphi} \simeq 1$, for both open-to-close and close-to-close returns. This suggests that the realized measure, x_t, is roughly proportional to the conditional variance for

Table II. Results for the log-linear specification

Model	Open-to-close returns							Close-to-close returns						
	G (1,1)	RG (1,1)	RG (1,2)	RG (2,1)	RG (2,2)	RG (2,2)†	RG (2,2)*	G (1,1)	RG (1,1)	RG (1,2)	RG (2,1)	RG (2,2)	RG (2,2)†	RG (2,2)*
Panel A: Point estimates and log-likelihood														
ω	0.04 (0.01)	0.06 (0.02)	0.04 (0.02)	0.06 (0.02)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.05 (0.00)	0.18 (0.03)	0.11 (0.02)	0.19 (0.03)	0.01 (0.01)	0.01 (0.01)	0.04 (0.05)
α	0.03 (0.01)						0.00 (0.00)	0.03 (0.00)						0.00 (0.00)
β_1	0.96 (0.01)	0.55 (0.03)	0.70 (0.05)	0.40 (0.05)	1.43 (0.04)	1.42 (0.09)	1.45 (0.05)	0.96 (0.01)	0.54 (0.03)	0.72 (0.05)	0.37 (0.05)	1.40 (0.07)	1.40 (0.10)	1.35 (0.42)
β_2				0.13 (0.05)	−0.44 (0.04)	−0.44 (0.07)	−0.46 (0.04)				0.15 (0.05)	−0.42 (0.06)	−0.43 (0.08)	−0.39 (0.30)
γ_1		0.41 (0.03)	0.45 (0.04)	0.43 (0.04)	0.46 (0.04)	0.40 (0.05)	0.42 (0.04)		0.43 (0.05)	0.48 (0.06)	0.46 (0.05)	0.45 (0.06)	0.42 (0.05)	0.46 (0.07)
γ_2			−0.18 (0.06)		−0.44 (0.04)	−0.38 (0.04)	−0.41 (0.04)			−0.21 (0.07)		−0.43 (0.05)	−0.40 (0.04)	−0.42 (0.08)
ξ		−0.18 (0.05)	−0.18 (0.05)	−0.18 (0.05)	−0.23 (0.05)	−0.16 (0.05)	−0.18 (0.05)		−0.42 (0.06)	−0.42 (0.06)	−0.42 (0.06)	−0.42 (0.05)	−0.41 (0.04)	−0.42 (0.04)
φ		1.04 (0.06)	1.04 (0.07)	1.04 (0.07)	0.96 (0.08)	1.07 (0.08)	1.03 (0.07)		0.99 (0.10)	1.00 (0.10)	0.99 (0.10)	0.99 (0.10)	1.03 (0.08)	0.99 (0.08)
σ_u		0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.41 (0.08)	0.38 (0.08)		0.39 (0.08)	0.38 (0.08)	0.39 (0.08)	0.38 (0.08)	0.41 (0.08)	0.38 (0.08)
τ_1		−0.07 (0.01)	−0.07 (0.01)	−0.07 (0.01)	−0.07 (0.01)		−0.07 (0.01)		−0.11 (0.01)	−0.11 (0.01)	−0.11 (0.01)	−0.11 (0.01)		−0.11 (0.01)
τ_2		0.07 (0.01)	0.07 (0.01)	0.07 (0.01)	0.07 (0.01)		0.07 (0.01)		0.04 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.01)		0.04 (0.01)
$\ell(r,x)$		−2395.6	−2388.8	−2391.9	−2385.1	−2495.7	−2382.9		−2576.9	−2567.2	−2571.7	−2563.9	−2661.7	−2563.5
Panel B: Auxiliary statistics														
π	0.988	0.975	0.986	0.976	0.999	0.999	0.999	0.988	0.974	0.987	0.975	0.999	0.999	0.999
ρ		−0.18	−0.18	−0.16	−0.19		−0.16		−0.27	−0.25	−0.25	−0.25		−0.28
ρ^-		−0.33	−0.32	−0.32	−0.35		−0.35		−0.31	−0.29	−0.28	−0.28		−0.31
ρ^+		0.12	0.12	0.13	0.13		0.14		−0.01	−0.03	−0.03	0.03		−0.02
$\ell(r)$	−1752.7	−1712.0	−1710.3	−1711.4	−1712.3	−1708.9	−1709.6	−1938.2	−1876.5	−1875.5	−1876.1	−1875.7	−1874.9	−1876.1

Note: G(1,1) denotes the LGARCH(1,1) model, which does not utilize a realized measure of volatility. RG(2,2)† denotes the Realized GARCH(2,2) model without the $\tau(z)$ function, which captures the dependence between returns and innovations in volatility. RG(2,2)* is the RG(2,2) extended to include the ARCH-term $\alpha \log r_{t-1}^2$. The latter is insignificant. Standard errors (in parentheses) are robust standard errors based on the sandwich estimator $\mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}$.

both open-to-close returns and close-to-close returns. The fact that ξ is estimated to be smaller (more negative) for close-to-close returns than for open-to-close returns simply reflects that the realized measure is computed over an interval that spans a shorter period than close-to-close returns.

In terms of partial log-likelihood function, $\ell(r)$, the Realized GARCH models clearly dominate the conventional logarithmic GARCH(1,1). The corresponding results for the Realized GARCH models based on a linear specification are reported as supporting information in a separate Web Appendix. These show that the log-linear specification dominates the linear specification. In the RG(2,2)*, which includes $\log r_{t-1}^2$ in the GARCH equation, we replace about 10 squared zero close-to-close returns with the truncation parameter ε. The standard errors in this model are rather sensitive to the value of the truncation parameter. The problem disappears if we use a smaller truncation parameter, but the smaller truncation parameter also causes the performance of the LGARCH to deteriorate substantially.

5.4.2 *Log-Linear RealGARCH(1,2) for All Stocks (Table III)*

Table III shows the parameter estimates for the log-linear Realized GARCH(1,2) model for all 29 assets. The empirical results are based on open-to-close returns. We observe that the estimates are remarkably similar across the stocks that span different sectors and have varying market dynamics. The conditional correlations, ρ^- and ρ^+, reveal a strong asymmetry for the index fund, SPY, since $\hat{\rho}^- = -0.32$ and $\hat{\rho}^+ = 0.13$. For the individual stocks the asymmetry is less pronounced, which is consistent with the existing literature (see, for example, Yu, 2008, and reference therein). However, two stocks, CVX and XOM, have strong asymmetries of the same magnitude as the index fund, SPY.

5.4.3 *News Impact Curve (Figure 1)*

The leverage function, $\tau(z)$, is closely related to the *news impact curve* that was introduced by Engle and Ng (1993). High-frequency data facilitate a more detailed study of the news impact curve than is possible with daily returns. Chen and Ghysels (2011) study the news impact curve in this context, but their approach is very different from ours. However, the shape of the news impact curve they estimate is very similar to ours. The news impact curve shows how volatility is impacted by a shock to the price, and our Hermite specification for the leverage function presents a very flexible framework for estimating this effect. In the log-linear specification we define the news impact curve by $\nu(z) = \mathrm{E}(\log h_{t+1}|z_t = z) - \mathrm{E}(\log h_{t+1})$, so that $100\nu(z)$ measures the percentage impact on volatility as a function of a return shock that is measured in units of its standard deviation. Here we have $\nu(z) = \gamma_1\tau(z)$ (see Section 2). We have estimated the log-linear RealGARCH(1,2) model for both IBM and SPY using a flexible leverage function based on the first four Hermite polynomials. The point estimates are $(\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\tau}_4) = (-0.036, 0.090, 0.001, -0.003)$ for IBM and $(\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3, \hat{\tau}_4) = (-0.068, 0.081, 0.014, 0.002)$ for SPY. Note that the Hermite polynomials of orders three and four add little beyond the first two polynomials. The news impact curves implied by these estimates are presented in Figure 1.

The estimated news impact curve for IBM is more symmetric about zero than that of SPY, and this empirical result is fully consistent with the existing literature. The most common approach to model the news impact curve is to adopt a specification with a 'kink' at zero, such as that used in the EGARCH model by Nelson (1991), $\tau(z) = \tau_1 z + \tau_+(|z| - \mathrm{E}|z|)$. We also estimated the leverage functions with the piecewise linear function, which leads to similar empirical results. Specifically, the implied news impact curves have the most pronounced asymmetry for the index fund, SPY, and the two oil-related stocks: CVX and XOM. However, the likelihood function tends to be larger with the polynomial leverage function, $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$, and the polynomial specification simplifies the likelihood analysis.

DOI: 10.1002/jae

Table III. Estimates for the log-linear Realized GARCH(1, 2) model

	ω	β	γ_1	γ_2	ξ	φ	σ_u	τ_1	τ_2	$\ell(r)$	$\ell(r,x)$	π	ρ	ρ^-	ρ^+
AA	0.03	0.77	0.33	−0.14	−0.07	1.15	0.40	−0.04	0.09	−2776.4	−3519.9	0.98	−0.08	−0.32	0.24
AIG	0.02	0.74	0.45	−0.21	−0.06	1.02	0.45	−0.02	0.04	−2403.1	−3317.2	0.98	−0.06	−0.17	0.08
AXP	0.05	0.70	0.38	−0.12	−0.16	1.08	0.43	−0.02	0.10	−2371.1	−3217.9	0.99	−0.05	−0.30	0.25
BA	0.02	0.82	0.31	−0.17	−0.13	1.22	0.39	−0.03	0.09	−2536.0	−3260.0	0.99	−0.09	−0.36	0.26
BAC	0.00	0.78	0.51	−0.29	0.00	0.99	0.42	−0.04	0.08	−2016.9	−2823.4	0.99	−0.09	−0.31	0.21
C	−0.02	0.74	0.45	−0.19	0.09	0.99	0.39	−0.03	0.09	−2260.5	−2974.0	0.99	−0.07	−0.31	0.24
CAT	0.03	0.82	0.37	−0.22	−0.14	1.07	0.38	−0.03	0.09	−2621.1	−3279.4	0.99	−0.08	−0.32	0.27
CVX	0.03	0.71	0.33	−0.14	−0.09	1.32	0.39	−0.08	0.08	−2319.1	−3021.8	0.97	−0.19	−0.35	0.14
DD	−0.01	0.77	0.37	−0.17	0.08	1.08	0.40	−0.05	0.08	−2301.2	−3067.3	0.98	−0.13	−0.35	0.20
DIS	0.01	0.85	0.39	−0.25	−0.05	1.10	0.41	−0.04	0.09	−2518.5	−3289.6	1.00	−0.09	−0.35	0.22
GE	0.00	0.81	0.38	−0.19	0.01	0.98	0.41	−0.01	0.08	−2197.8	−2988.7	0.99	−0.02	−0.26	0.25
GM	0.06	0.84	0.39	−0.24	−0.32	1.02	0.47	−0.01	0.12	−2987.9	−3967.3	0.99	−0.01	−0.33	0.31
HD	0.01	0.79	0.39	−0.20	0.00	1.01	0.41	−0.05	0.09	−2538.4	−3318.4	0.99	−0.13	−0.37	0.20
IBM	0.00	0.74	0.41	−0.15	0.01	0.94	0.39	−0.04	0.08	−2192.6	−2896.7	0.98	−0.09	−0.32	0.24
INTC	0.02	0.87	0.46	−0.33	−0.11	1.03	0.36	−0.02	0.07	−2869.1	−3481.1	1.00	−0.05	−0.24	0.22
JNJ	−0.03	0.80	0.38	−0.19	0.13	1.04	0.44	0.02	0.10	−1874.8	−2777.3	0.99	0.04	−0.25	0.30
JPM	0.01	0.81	0.49	−0.30	−0.02	0.98	0.42	−0.04	0.09	−2463.0	−3276.8	0.99	−0.10	−0.30	0.22
KO	−0.05	0.76	0.45	−0.21	0.19	0.93	0.38	−0.02	0.07	−1886.7	−2573.6	0.99	−0.06	−0.28	0.19
MCD	0.00	0.88	0.37	−0.25	−0.01	0.98	0.45	−0.05	0.11	−2461.8	−3371.9	0.99	−0.09	−0.35	0.26
MMM	0.00	0.77	0.43	−0.23	0.02	0.98	0.41	−0.02	0.07	−2140.3	−2944.8	0.97	−0.04	−0.23	0.21
MRK	0.03	0.84	0.33	−0.21	−0.19	1.23	0.47	0.01	0.07	−2479.2	−3478.5	0.98	0.04	−0.13	0.18
MSFT	−0.01	0.79	0.44	−0.22	0.08	0.92	0.38	−0.03	0.08	−2330.7	−3021.1	0.99	−0.08	−0.31	0.24
PG	−0.04	0.78	0.43	−0.25	0.18	1.04	0.41	−0.05	0.08	−1850.7	−2646.6	0.98	−0.14	−0.32	0.14
T	0.00	0.86	0.53	−0.38	0.01	0.86	0.46	−0.03	0.10	−2560.4	−3512.7	0.99	−0.07	−0.32	0.25
UTX	−0.01	0.80	0.45	−0.24	0.06	0.88	0.40	−0.01	0.10	−2302.4	−3059.2	0.99	−0.05	−0.34	0.29
VZ	−0.01	0.79	0.40	−0.20	0.07	1.01	0.43	−0.03	0.09	−2343.4	−3196.7	0.99	−0.08	−0.31	0.23
WMT	−0.02	0.80	0.37	−0.19	0.12	1.04	0.39	−0.01	0.09	−2164.9	−2893.6	0.99	−0.02	−0.29	0.30
XOM	0.03	0.71	0.34	−0.12	−0.10	1.26	0.38	−0.08	0.08	−2334.7	−2994.1	0.98	−0.20	−0.37	0.15
SPY	0.04	0.70	0.45	−0.18	−0.18	1.04	0.38	−0.07	0.07	−1710.3	−2388.8	0.99	−0.17	−0.32	0.13
Average	0.01	0.79	0.41	−0.21	−0.02	1.04	0.41	−0.03	0.09			0.99	−0.08	−0.30	0.22

5.4.4 In-Sample and Out-of-Sample Log-Likelihood Results

Table IV presents likelihood ratio statistics using open-to-close returns. The inference we draw from these statistics is that RealGARCH(1,2) is generally a good model. Moreover, the leverage function is highly significant, whereas α is insignificant.

The statistics in panel A are conventional likelihood ratio statistics:

$$\mathrm{LR}_i = 2\{\ell_{RG(2,2)^*}(r,x) - \ell_i(r,x)\}, \quad i = 1,\ldots,5$$

where each of the five smaller models are benchmarked against the largest model. The largest model is the log-linear RealGARCH(2,2) model, which includes the squared returns in the GARCH equation (in addition to the realized measure). In the QMLE framework the limit distribution of likelihood ratio statistic, LR_i, is usually given as a weighted sum of χ^2-distributed random variables. Thus comparing the LR_i with the usual critical value of a χ^2-distribution is only indicative of significance. The LR statistics in the column labeled '(2,2)' are small in most cases, so there is little evidence that α is significant. This is consistent with the existing literature, since squared returns are typically found to be insignificant once a realized measure is included in the GARCH equation. The LR statistics in column '(2,2)†' are well over 100 in all cases. This shows that the leverage function, $\tau(z_t)$, is highly significant. Similarly, the LR statistics show that the hypothesis, $\beta_2 = \gamma_2 = 0$, is rejected in most cases. Therefore the empirical evidence does not support a simplification of the model to the RealGARCH(1,1). The results for the two sub-hypotheses $\beta_2 = 0$ and $\gamma_2 = 0$ are less conclusive. The likelihood ratio statistics for the hypothesis

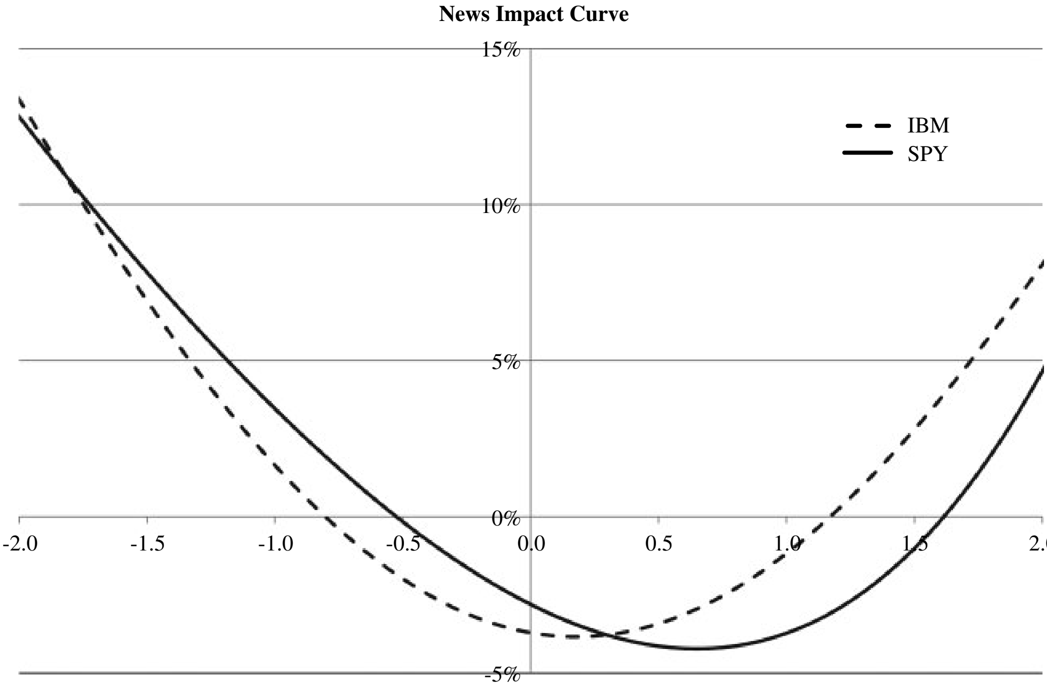

Figure 1. News impact curve for IBM and SPY

$\beta_2 = 0$ (which is given as the difference between the statistics in columns '(1,2)' and '(2,2)') are on average $5.7 = 9.6 - 3.9$. In a correctly specified model, this would be borderline significant. The LR statistics for the hypothesis $\gamma_2 = 0$ tend to be larger with an average value of $16.6 = 20.5 - 3.9$. Thus the empirical evidence favors the RealGARCH(1,2) model over the RealGARCH(2,1) model.

The statistics in panel B are out-of-sample likelihood ratio statistics, defined by $\sqrt{\frac{n}{m}}\{\ell_{\mathrm{RG}(2,2)}(r,x) - \ell_j(r,x)\}$, where n and m denote the sample sizes, in-sample and out-of-sample, respectively. The in-sample estimates are simply plugged into the out-of-sample log-likelihood function. The asymptotic distribution of the out-of-sample LR statistic is non-standard. However, in a nested comparison where the larger model has k additional parameters (which are all zero in population), it can be shown that

$$\sqrt{\frac{n}{m}}\{\ell_i(r,x) - \ell_j(r,x)\} \overset{d}{\to} Z_1'\Lambda Z_2, \quad \text{as} \quad m, n \to \infty \quad \text{with} \quad m/n \to 0$$

where Z_1 and Z_2 are independent $Z_i \sim N_k(0, I)$, and Λ is a diagonal matrix with eigenvalues from $\mathcal{I}^{-1}\mathcal{J}$ (see Hansen, 2009). For correctly specified models ($\Lambda = I$) the (two-sided) critical values can be inferred from the distribution of $|Z_1'Z_2|$. For $k = 1$ the 5% and 1% critical values are 2.25 and 3.67, respectively, and for two degrees of freedom ($k = 2$), these are 3.05 and 4.83, respectively. Although our model is likely to be misspecified to some degree, we will later argue that the log-linear Gaussian specification is not very misspecified. Therefore we will take these critical values as reasonable approximations. Compared to these critical values we find, on average, significant evidence in favor of a model with more lags than RealGARCH(1,1). The statistical evidence in favor of a leverage function is again very strong. Adding the ARCH parameter, α, will (on average) result in a worse out-of-sample log-likelihood. On average, we have a tie between the three models: RealGARCH(1,2), RealGARCH(2,1), and RealGARCH(2,2).

DOI: 10.1002/jae

Table IV. In-sample and out-of-sample likelihood ratio statistics

	Panel A: In-sample likelihood ratio						Panel B: Out-of-sample likelihood ratio						Panel C: Out-of-sample partial likelihood ratio						
	(1,1)	(1,2)	(2,1)	(2,2)	$(2,2)^{\dagger}$	$(2,2)^{*}$	(1,1)	(1,2)	(2,1)	(2,2)	$(2,2)^{\dagger}$	$(2,2)^{*}$	G11	(1,1)	(1,2)	(2,1)	(2,2)	$(2,2)^{\dagger}$	$(2,2)^{*}$
AA	24.0	8.3	11.7	3.9	187.4	0	6.9	4.5	6.4	0	21.9	0.1	4.6	3.3	1.4	2.6	0.0	0.6	0.7
AIG	43.8	15.2	20.6	14.2	201.2	0	15.7	−0.2	6.0	0	25.5	12.5	56.1	9.3	5.9	7.4	6.0	0.0	7.2
AXP	30.5	25.0	26.0	0.0	197.0	0	1.2	2.7	1.5	0	13.3	0.2	24.0	0.0	0.3	0.1	1.3	1.7	1.4
BA	34.8	6.0	18.8	0.0	197.1	0	−1.4	0.4	−2.6	0	24.5	0.0	1.1	0.6	1.8	1.7	0.0	0.3	0.0
BAC	46.3	5.9	20.5	5.1	198.8	0	8.0	−0.7	−0.3	0	56.5	−2.8	147.9	4.1	0.6	0.0	1.5	19.7	0.9
C	26.8	9.5	16.0	6.3	228.4	0	1.3	−2.7	−2.5	0	−0.1	−7.2	26.9	0.3	0.9	0.5	1.1	0.0	0.4
CAT	39.8	2.3	13.8	1.6	227.3	0	1.7	−1.0	−4.2	0	9.0	2.2	47.3	0.1	0.8	0.0	1.0	1.6	1.5
CVX	17.7	0.1	3.5	0.0	194.6	0	5.3	0.0	2.2	0	20.8	−0.1	30.1	0.6	0.3	0.2	0.3	0.0	0.3
DD	31.1	10.8	15.9	6.1	167.0	0	4.6	4.1	3.0	0	−7.1	6.9	19.2	1.2	1.5	1.5	1.2	1.4	0.0
DIS	57.7	12.6	33.1	0.0	213.7	0	4.8	4.8	3.7	0	24.3	0.0	35.4	0.0	2.0	1.4	0.8	1.1	0.8
GE	37.7	12.6	20.2	12.2	200.3	0	14.1	0.6	5.9	0	9.3	0.6	41.6	0.9	0.0	0.6	0.0	0.3	0.7
GM	62.7	18.8	39.9	18.4	319.9	0	17.0	−0.2	6.8	0	7.2	14.6	57.5	1.4	0.0	0.2	0.0	0.3	0.9
HD	29.8	4.0	14.4	2.0	201.3	0	2.2	−1.0	−1.0	0	28.9	0.5	45.5	0.9	0.0	0.1	0.2	2.1	0.4
IBM	28.5	16.2	20.2	1.0	176.7	0	3.3	−0.1	1.0	0	25.0	−0.2	12.0	0.4	0.5	0.7	0.1	0.3	0.0
INTC	75.9	13.9	47.8	9.7	130.8	0	0.9	0.1	−3.4	0	36.8	19.1	133.1	1.7	0.5	2.0	0.0	0.1	1.8
JNJ	34.6	1.3	4.8	0.1	235.4	0	7.1	−0.8	2.0	0	30.7	0.2	21.8	0.4	0.0	0.1	0.1	0.1	0.0
JPM	50.7	3.5	23.6	1.9	213.5	0	0.3	−2.5	−5.4	0	−3.6	−0.6	34.9	0.0	1.3	0.1	1.6	2.0	1.6
KO	39.7	22.0	28.3	0.0	186.1	0	2.1	−2.8	−1.0	0	22.6	0.0	5.6	0.4	0.6	0.0	0.5	0.9	0.5
MCD	54.8	2.9	26.1	0.8	278.6	0	−13.0	−0.6	−9.3	0	47.4	1.4	6.8	0.0	1.5	0.2	1.8	2.0	1.7
MMM	32.0	6.7	20.4	0.1	183.5	0	−2.4	−0.2	−2.8	0	24.5	−0.1	14.5	0.0	1.9	0.9	1.4	1.0	1.4
MRK	64.6	12.0	10.6	2.9	309.0	0	−3.4	−2.5	−1.2	0	32.4	0.4	11.8	0.0	1.6	1.7	1.9	2.4	1.8
MSFT	37.3	11.5	20.5	10.2	186.1	0	8.3	4.2	7.1	0	17.0	4.1	37.1	0.1	0.5	0.4	0.0	0.2	0.1
PG	36.5	3.0	12.3	2.9	160.4	0	−1.3	0.3	−0.4	0	35.0	1.0	24.8	0.0	1.4	0.6	1.3	1.5	1.7
T	69.8	4.8	39.0	0.0	198.3	0	19.7	−0.3	13.5	0	35.1	−0.2	19.2	2.8	0.2	2.6	0.0	1.7	0.0
UTX	39.6	21.6	29.0	0.0	223.7	0	5.8	−2.0	1.2	0	33.1	−0.4	12.3	1.4	0.2	0.9	0.2	0.0	0.1
VZ	31.5	4.3	13.2	0.5	188.7	0	15.0	3.2	11.1	0	16.6	−1.3	8.7	3.5	0.7	2.8	0.3	0.0	0.1
WMT	36.2	12.0	23.0	8.3	190.0	0	−7.2	−3.8	−7.9	0	28.7	9.4	30.5	0.0	0.6	0.0	1.3	1.4	2.6
XOM	14.7	1.1	4.3	0.9	234.0	0	5.7	−0.1	1.0	0	27.9	0.6	21.6	0.0	0.5	0.3	0.5	0.5	0.5
SPY	25.3	11.6	17.9	4.2	225.6	0	6.3	−1.2	1.9	0	24.4	1.4	40.8	0.8	0.6	0.7	0.0	2.5	1.3
Average	39.8	9.6	20.5	3.9	208.8	0	4.4	0.1	1.1	0	23.0	2.1	33.5	1.2	1.0	1.0	0.8	1.6	1.0

DOI: 10.1002/jae

In panel C, we report partial out-of-sample likelihood ratio statistics that are defined by $2\{\max_i \ell_i(r|x) - \ell_j(r|x)\}$. These LR statistics are based on the partial likelihood for returns, which enables us to compare the empirical fit to the conventional LGARCH. Again we see that the Realized GARCH models strongly dominate the LGARCH(1,1) model. This is impressive because the Realized GARCH models are not seeking to maximize the partial likelihood, as is the objective for the LGARCH model.[4]

5.5. A Comparison of the Linear and Log-Linear Specifications

One of the reasons we prefer the log-linear Gaussian specification over the linear Gaussian specification is that the former is much less at odds with the data. The log-linear specification results in far less heteroskedasticity, as is evident from Figure 2. The left-hand panel is a scatter plot of x_t against $\hat{h}_t$, (for the linear RealGARCH(1,2) model) and the right-hand panel is a scatter plot of $\log x_t$ against $\log \hat{h}_t$ (for the log-linear RealGARCH(1,2) model). The two models produce very similar value for h_t; however, there is obviously a very pronounced degree of heteroskedasticity in the linear models. The linear model may be improved by modifying the leverage function, but our point is that the simple measurement equation does a good job within the log-linear specification. Homoskedastic errors are not essential for the QMLEs but heteroskedasticity causes the QMLE to be inefficient. Moreover, misspecification causes the likelihood ratio statistic to have an asymptotic distribution that is a weighted sum of $\chi^2_{(1)}$-distributed random variables, rather than a pure sum of such. Comparing likelihood ratio statistics with critical values to a standard χ^2-distribution, as an approximation, becomes dubious when the model is highly misspecified.

Figure 3 presents additional evidence in favor of log-linear specification, and shows that the leverage function is critical for the validity of the assumed independence between z_t and u_t. The figure has four scatter plots of the residuals, $\{\hat{z}_t, \hat{u}_t\}_{t=1}^n$, where each scatter plot corresponds to a particular specification. The upper panels in Figure 3 are for the linear specification and the two lower panels are for the log-linear specification. Left panels are based on residuals obtained without a leverage function (i.e. $\tau(z) = 0$), and those on the right are the residuals obtained with a quadratic specification for $\tau(z)$. Ideally the scatter plot would look like one of two independent Gaussian distributed random variables. The upper panels are clearly at odds with this, which

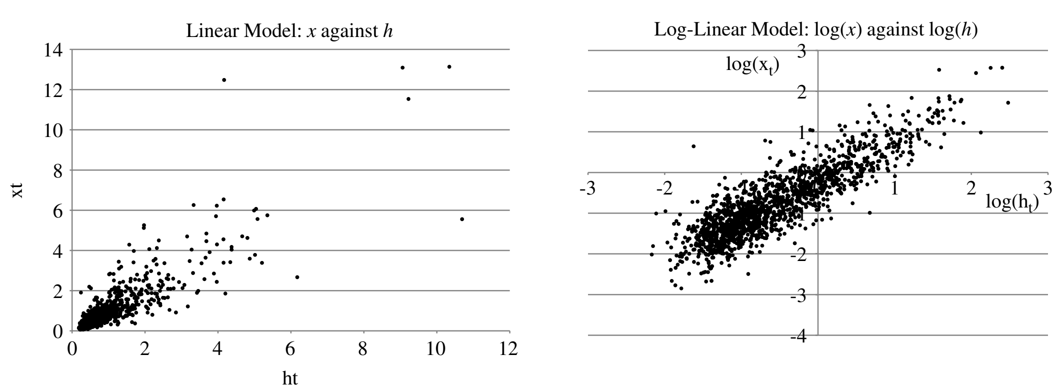

Figure 2. Heteroskedasticity in measurement equation

[4] The asymptotic distribution of these statistics is very non-standard (and generally unknown) because we are comparing a model that maximizes the partial likelihood (the LGARCH(1,1) model) with models that maximize a joint likelihood (the Realized GARCH models).

DOI: 10.1002/jae

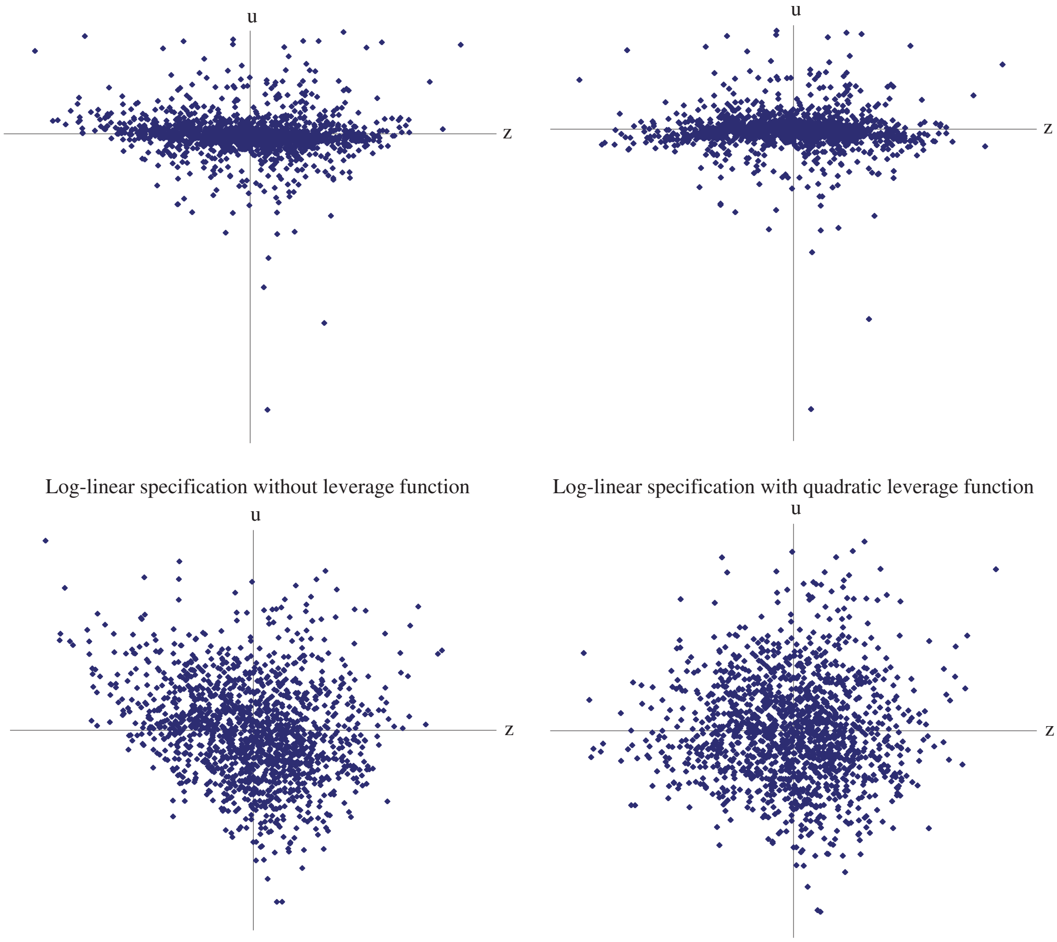

Figure 3. Scatter plots of the residuals, $(\hat{z}_t, \hat{u}_t)$, obtained with four different RealGARCH(1,2) models. The upper panels are for the linear specification and the lower panels are for the log-linear specification. The left-hand panels are for models without a leverage function and the right-hand panels are with a quadratic leverage function. The log-linear specification with the leverage function is clearly best suited for the Gaussian structure of the quasi log-likelihood function. This figure is available in color online at wileyonlinelibrary.com/journal/jae

confirms that the linear specification is highly misspecified. The lower-left panel is the log-linear specification without a leverage function, and it clearly reveals unmodeled dependence between z_t and u_t. The log-linear model with quadratic leverage function (lower-right panel) offers a much better agreement with the underlying assumptions.

The fact that the log-linear model is far less misspecified than the linear model can also be illustrated by comparing robust and non-robust standard errors. In Table V we have computed standard errors using those of the two information matrices (the diagonal elements of $\mathcal{I}^{-1}$ and J^{-1}) and the robust standard errors computed from the diagonal of $\mathcal{I}^{-1}J\mathcal{I}^{-1}$. In a correctly specified model these standard errors should be in agreement. This is obviously not the case for the linear specification, whereas there is better agreement with the log-linear specification.

DOI: 10.1002/jae

Table V. Conventional and robust standard errors computed for Realized GARCH(1,2) model with a quadratic leverage function. The data are open–close SPY returns

	Standard errors for RealGARCH(1,2) Model					
	Linear model			Log-linear model		
	$\mathcal{I}^{-1}$	$\mathcal{J}^{-1}$	$\mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}$	$\mathcal{I}^{-1}$	$\mathcal{J}^{-1}$	$\mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}$
ω	0.007	0.004	0.019	0.015	0.015	0.016
β	0.034	0.017	0.125	0.040	0.031	0.053
γ_1	0.053	0.040	0.133	0.030	0.025	0.040
γ_2	0.054	0.032	0.177	0.046	0.036	0.062
ξ	0.038	0.037	0.096	0.044	0.042	0.051
φ	0.080	0.064	0.212	0.044	0.033	0.069
σ_u	0.009	0.002	0.054	0.005	0.005	0.006
τ_1	0.013	0.014	0.016	0.010	0.011	0.011
τ_2	0.008	0.013	0.011	0.006	0.008	0.006

6. MOMENTS, FORECASTING, AND INSIGHT ABOUT THE REALIZED MEASURE

In this section we discuss the skewness and kurtosis (for cumulative returns) that the Realized GARCH model can generate for realistic values of the parameters and we discuss issues related to multi-period forecasting in the Realized GARCH context.

6.1. Properties of Cumulative Returns: Skewness and Kurtosis

We consider the skewness and kurtosis for returns generated by a Realized GARCH model. Some analytical results in closed form can be derived for the linear Realized GARCH model. These results are given as supporting information in a separate Web Appendix. Here we will focus on the log-linear Realized GARCH model. We have the following results for the kurtosis of a single period return.

Proposition 1. Consider the log-linear RealGARCH(1,1) model and define $\pi = \beta + \varphi\gamma$ and $\mu = \omega + \varphi\xi$, so that

$$\log h_t = \pi \log h_{t-1} + \mu + \gamma w_{t-1}, \quad \text{where} \quad w_t = \tau_1 z_t + \tau_2(z_t^2 - 1) + u_t$$

with $z_t \sim$ i.i.d.$N(0,1)$ and $u_t \sim$ i.i.d.$N(0, \sigma_u^2)$. The kurtosis of the return $r_t = \sqrt{h_t} z_t$ is given by

$$\frac{\mathrm{E}(r_t^4)}{\mathrm{E}(r_t^2)^2} = 3\left(\prod_{i=0}^{\infty} \frac{1 - 2\pi^i\gamma\tau_2}{\sqrt{1 - 4\pi^i\gamma\tau_2}}\right) \exp\left\{\sum_{i=0}^{\infty} \frac{\pi^{2i}\gamma^2\tau_1^2}{1 - 6\pi^i\gamma\tau_2 + 8\pi^{2i}\gamma^2\tau_2^2}\right\} \exp\left\{\frac{\gamma^2\sigma_u^2}{1 - \pi^2}\right\} \tag{8}$$

There does not appear to be a way to further simplify expression (8); however, when $\gamma\tau_2$ is small, as we found it to be empirically, we have the approximation $\frac{\mathrm{E}(r_t^4)}{\mathrm{E}(r_t^2)^2} \simeq 3\exp\left\{\frac{\gamma^2\tau_2^2}{-\log \pi} + \frac{\gamma^2(\tau_1^2 + \sigma_u^2)}{1 - \pi^2}\right\}$ (see the Web Appendix for details). The skewness for single period returns is non-zero, if and only if the Studentized return, z_t, has non-zero skewness. This follows directly from the identity $r_t = \sqrt{h_t} z_t$, and the assumption that $z_t \perp\!\!\!\perp h_t$, which shows that $\mathrm{E}(r_t^d) = \mathrm{E}(h_t^{d/2} z_t^d) = \mathrm{E}\{\mathrm{E}(h_t^{d/2} z_t^d | \mathcal{F}_{t-1})\} = \mathrm{E}(h_t^{d/2})\mathrm{E}(z_t^d)$, and in particular that $\mathrm{E}(r_t^3) = \mathrm{E}(h_t^{3/2})\mathrm{E}(z_t^3)$. Thus a symmetric

DOI: 10.1002/jae

distribution for z_t implies that r_t has zero skewness, and this is a property that is shared by the standard GARCH model and the Realized GARCH model alike.

For the skewness and kurtosis of cumulative returns, $r_t + \ldots + r_{t+k}$, the situation is very different, because the leverage function induces skewness. For this problem we resort to simulation methods using a design based on our empirical estimates for log-linear Realized GARCH(1,2) model that we obtained for the SPY open-to-close returns. The skewness and kurtosis of cumulative returns are shown in Figure 4, and it is evident that the Realized GARCH model can produce strong and persistent skewness and kurtosis.

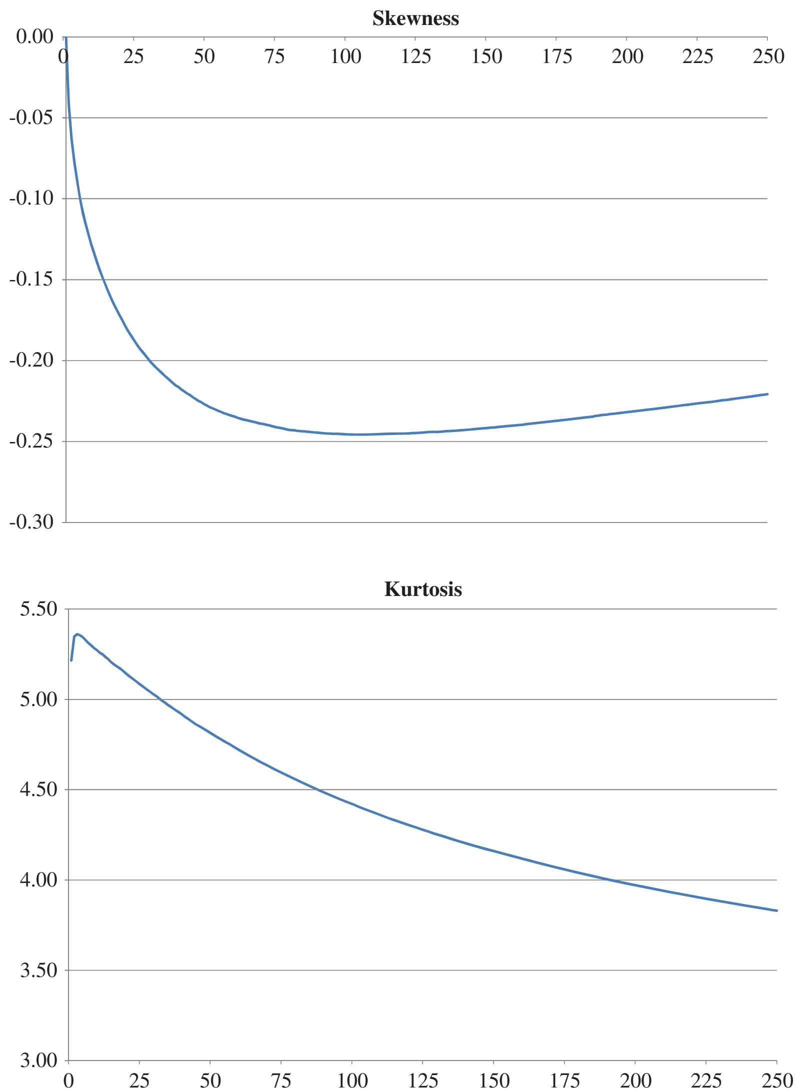

Figure 4. Skewness and kurtosis of cumulative returns from a Realized GARCH model with the log-linear specification. The x-axis gives the number of periods (days) over which returns are accumulated. This figure is available in color online at wileyonlinelibrary.com/journal/jae

DOI: 10.1002/jae

6.2. Multi-period Forecast

The Realized GARCH model can be used to predict both the conditional return variance and the realized measure. The latter has been the subject of a very active literature (see, for example, Andersen *et al.*, 2003, 2004, 2005, 2007, 2011b).

One of the main advantages of having a complete specification, i.e. a model that fully describes the dynamic properties of x_t, is that multi-period-ahead forecasting is feasible. In contrast, the GARCH-X model can only be used to make one-step-ahead predictions. Multi-period-ahead predictions are not possible without a model for x_t, such as the one implied by the measurement equation in the Realized GARCH model.

Multi-period-ahead predictions with the Realized GARCH model is straightforward for both the linear and log-linear Realized GARCH models. Let $\tilde{h}_t$ denote either h_t or $\log h_t$, and consider first the case where $p = q = 1$. By substituting the GARCH equation into the measurement equation we obtain the VARMA(1,1) structure

$$\begin{bmatrix} \tilde{h}_t \\ \tilde{x}_t \end{bmatrix} = \begin{bmatrix} \beta & \gamma \\ \varphi\beta & \varphi\gamma \end{bmatrix} \begin{bmatrix} \tilde{h}_{t-1} \\ \tilde{x}_{t-1} \end{bmatrix} + \begin{bmatrix} \omega \\ \xi + \varphi\omega \end{bmatrix} + \begin{bmatrix} 0 \\ \tau(z_t) + u_t \end{bmatrix}$$

which can be used to generate the predictive distribution of future values of $\tilde{h}_t$, $\tilde{x}_t$, as well as returns r_t, using

$$\begin{bmatrix} \tilde{h}_{t+k} \\ \tilde{x}_{t+k} \end{bmatrix} = \begin{bmatrix} \beta & \gamma \\ \varphi\beta & \varphi\gamma \end{bmatrix}^h \begin{bmatrix} \tilde{h}_t \\ \tilde{x}_t \end{bmatrix} + \sum_{j=0}^{k-1} \begin{bmatrix} \beta & \gamma \\ \varphi\beta & \varphi\gamma \end{bmatrix}^j \left\{ \begin{bmatrix} \omega \\ \xi + \varphi\omega \end{bmatrix} + \begin{bmatrix} 0 \\ \tau(z_{t+h-j}) + u_{t+h-j} \end{bmatrix} \right\}$$

This is easily extended to the general case ($p, q \geq 1$) where we have $Y_t = AY_{t-1} + b + \varepsilon_t$, with the conventions

$$Y_t = \begin{bmatrix} \tilde{h}_t \\ \vdots \\ \tilde{h}_{t-p+1} \\ \tilde{x}_t \\ \vdots \\ \tilde{x}_{t-q+1} \end{bmatrix}, \quad A = \begin{pmatrix} (\beta_1, \ldots, \beta_p) & (\gamma_1, \ldots, \gamma_q) \\ (I_{p-1\times p-1}, 0_{p-1\times 1}) & 0_{p-1\times q} \\ \varphi(\beta_1, \ldots, \beta_p) & \varphi(\gamma_1, \ldots, \gamma_q) \\ 0_{q-1\times p} & (I_{q-1\times q-1}, 0_{q-1\times 1}) \end{pmatrix}, \quad b = \begin{pmatrix} \omega \\ 0_{p-1\times 1} \\ \xi + \varphi\omega \\ 0_{q-1\times 1} \end{pmatrix},$$

$$\varepsilon_t = \begin{bmatrix} 0_{p\times 1} \\ \tau(z_t) + u_t \\ 0_{q\times 1} \end{bmatrix}$$

so that $Y_{t+k} = A^k Y_t + \sum_{j=0}^{k-1} A^j (b + \varepsilon_{t+k-j})$. The predictive distribution for $\tilde{h}_{t+h}$ and/or $\tilde{x}_{t+h}$ is given from the distribution of $\sum_{i=0}^{k-1} A^i \varepsilon_{t+h-i}$, which also enables us to compute a predictive distribution for r_{t+k} and cumulative returns $r_{t+1} + \ldots + r_{t+k}$. An advantage of the linear specification is that it directly produces a point forecast of future volatility, $(h_{t+1}, \ldots, h_{t+k})$. With the log-linear specification one would have to account for distributional aspects of $\log h_{t+k}$ in order to produce an unbiased forecast of h_{t+k}. This should not be a major obstacle because the log-linear Gaussian specification appears to work well with these data. Note that one is not required to generate auxillary future values of the realized measure when the objective is to predict future values of h_t or the distribution of future (cumulative) returns. The reason is that the innovations z_t and u_t are sufficient for generating the volatility path (and returns).

DOI: 10.1002/jae

7. CONCLUSION

In this paper we have proposed a complete model for returns and realized measures of volatility, x_t, where the latter is tied directly to the conditional volatility h_t. We have demonstrated that the model is straightforward to estimate and offers a substantial improvement in the empirical fit, relative to standard GARCH models based on daily returns only. The model is informative about realized measurement, such as its accuracy.

Our empirical analysis can be extended in a number of ways. For instance, inclusion of a jump-robust realized measure would be an interesting extension, because Andersen *et al.* (2007) have shown that the predictability in volatility is largely driven by the continuous component of volatility. Moreover, Bollerslev *et al.* (2009) have found that the leverage effect primarily acts through the continuous volatility component. Another possible extension is to introduce a bivariate model of open-to-close and close-to-open returns as an alternative to modeling close-to-close returns (see Andersen *et al.*, 2011a).

The Realized GARCH framework is naturally extended to a multi-factor structure, e.g. with m realized measures and k latent volatility variables, the Realized GARCH model discussed in this paper corresponds to the case $k = 1$, whereas the MEM framework corresponds to the case $m = k$. Such a hybrid framework would enable us to conduct an inference about the number of latent factors, k. We could, in particular, test the one-factor structure used in this paper against the multi-factor structure implied by MEM.

ACKNOWLEDGEMENTS

We thank Tim Bollerslev, Giampiero Gallo, Asger Lunde, and seminar participants at Humboldt University, the Einaudi Institute, the 2009 Stanford Institute for Theoretical Economics, and the Volatility Institute at New York University Stern School of Business for valuable comments.

REFERENCES

Andersen TG, Bollerslev T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**(4): 885–905.

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2001. The distribution of exchange rate volatility. *Journal of the American Statistical Association* **96**(453): 42–55 (correction published 2003, Vol. 98, p. 501).

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2003. Modeling and forecasting realized volatility. *Econometrica* **71**(2): 579–625.

Andersen TG, Bollerslev T, Meddahi N. 2004. Analytic evaluation of volatility forecasts. *International Economic Review* **45**: 1079–1110.

Andersen TG, Bollerslev T, Meddahi N. 2005. Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica* **73**: 279–296.

Andersen T, Bollerslev T, Diebold FX. 2007. Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* **89**: 701–720.

Andersen T, Dobrev D, Schaumburg E. 2008. Duration-based volatility estimation. Working paper Department of Finance, Kellogg School, Northwestern University.

Andersen TG, Bollerslev T, Huang X. 2011a. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics* **160**: 176–189.

Andersen TG, Bollerslev T, Meddahi N. 2011b. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* **160**: 220–234.

Barndorff-Nielsen OE, Shephard N. 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B* **64**: 253–280.

Barndorff-Nielsen OE, Shephard N. 2004. Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* **2**: 1–48.

DOI: 10.1002/jae

Barndorff-Nielsen OE, Shephard N. 2007. Variation, jumps and high frequency data in financial econometrics. In *Advances in Economics and Econometrics: Theory and Applications*. Ninth World Congress, Econometric Society Monographs. Cambridge University Press: Cambridge, UK; 328–372.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**: 1481–536.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2009a. Realised kernels in practice: trades and quotes. *Econometrics Journal* **12**: 1–33.
Barndorff-Nielsen OE, Kinnebrock S, Shephard N. 2009b. Measuring downside risk: realised semivariance. In *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, Watson MW, Bollerslev T, Russell J (eds). Oxford University Press: Oxford; 117–137.
Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2010. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, DOI: 10.1016/j.jeconom.2010.07.009.
Bollerslev T. 1986. Generalized autoregressive heteroskedasticity. *Journal of Econometrics* **31**: 307–327.
Bollerslev T. 2009. Glossary to ARCH (GARCH). In *Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle*, Bollerslev T, Russell JR, Watson M (eds). Oxford University Press: Oxford, UK; 137–164.
Bollerslev T, Wooldridge JM. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariance. *Econometric Reviews* **11**: 143–172.
Bollerslev T, Kretschmer U, Pigorsch C, Tauchen G. 2009. A discrete-time model for daily S&P500 returns and realized variations: jumps and leverage effects. *Journal of Econometrics* **150**: 151–66.
Brownless CT, Gallo GM. 2010. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* **8**: 29–56.
Carrasco M, Chen X. 2002. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**: 17–39.
Chen X, Ghysels E, Wang F. 2009. The HYBRID GARCH class of models. Working paper University of North Carolina at Chapel Hill.
Christoffersen P, Feunou B, Jacobs K, Meddahi N. 2010. The economic value of realized volatility. Working paper McGill University.
Cipollini F, Engle RF, Gallo GM. 2009. A model for multivariate non-negative valued processes in financial econometrics. Working paper Available at SSRN: http://ssrn.com/abstract=1333869.
Dobrev D, Szerszen P. 2010. The information content of high-frequency data for estimating equity return models and forecasting risk. Working paper Board of Governors of the Federal Reserve System, International Finance Discussion Papers,nr. 1005.
Engle RF. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **45**: 987–1007.
Engle RF. 2002. New frontiers of ARCH models. *Journal of Applied Econometrics* **17**: 425–446.
Engle RF, Gallo G. 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* **131**: 3–27.
Engle RF, Ng V. 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* **48**: 1747–1778.
Engle RF, Lilien DV, Robins RP. 1987. Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* **55**: 391–407.
Forsberg L, Bollerslev T. 2002. Bridging the gap between the distribution of realized (ECU) volatility and ARCH modeling (of the EURO): the GARCH-NIG model. *Journal of Applied Econometrics* **17**(5): 535–548.
Geweke J. 1986. Modelling persistence in conditional variances: a comment. *Econometric Reviews* **5**: 57–61.
Chen X, Ghysels E. 2011. News−-good or bad−-and its impact on volatility predictions over multiple horizons. *Review of Financial Studies* **24**: 46–81.
Hansen PR. 2009. In-sample fit and out-of-sample fit: their joint distribution and its implications for model selection. Working paper Department of Economics, Stanford University.
Hansen PR, Horel G. 2009. Quadratic variation by Markov chains. Working paper. Department of Economics, Stanford University.
Jensen ST, Rahbek A. 2004a. Asymptotic inference for nonstationary GARCH. *Econometric Theory* **20**: 1203–1226.
Jensen ST, Rahbek A. 2004b. Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* **72**: 641–646.
Lee S, Hansen BE. 1994. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* **10**: 29–52.

Lumsdaine RL. 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* **10**: 29–52.
Meddahi N. 2003. ARMA representation of integrated and realized variances. *Econometrics Journal* **6**: 335–356.
Milhøj A. 1987. A multiplicative parameterization of ARCH models. Technical report, Department of Statistics, University of Copenhagen.
Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**(2): 347–370.
Pantula SG. 1986. Modelling persistence in conditional variances: a comment. *Econometric Reviews* **5**: 71–74.
Shephard N, Sheppard K. 2010. Realising the future: forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics* **25**: 197–231.
Straumann D, Mikosch T. 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equation approach. *Annals of Statistics* **34**: 2449–2495.
Takahashi M, Omori Y, Watanabe T. 2009. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics and Data Analysis* **53**: 2404–2406.
Teräsvirta T. 2009. An introduction to univariate GARCH models. In *Handbook of Financial Time Series*, Mikosch T, Kreiss T, Davis J-P, Andersen RA, Gustav T (eds). Springer: Berlin; 17–42.
Visser MP. 2011. Garch parameter estimation using high-frequency data. *Journal of Financial Econometrics* **9**: 162–197.
Yu J. 2011. A semiparametric stochastic volatility model. *Journal of Econometrics* (forthcoming).

APPENDIX: PROOFS

Proof of Proposition 1

The first result followsby substituting $\log x_t = \varphi \log h_t + \xi + w_t$ and $\log r_t^2 = \log h_t + \kappa + v_t$ into the GARCH equation and rearranging. Next, we substitute $\log h_t = (\log x_t - \xi - w_t)/\varphi$, $\log r_t^2 = (\log x_t - \xi - w_t)/\varphi + \kappa + v_t$, and multiply by φ, and find

$$\log x_t - \xi - w_t = \varphi\omega + \sum_{i=1}^{p\vee q}(\beta_i + \alpha_i)(\log x_{t-i} - \xi - w_{t-i}) + \varphi\sum_{j=1}^{q}\gamma_j \log x_{t-j} + \varphi\sum_{j=1}^{q}\alpha_j(\kappa + v_{t-j})$$

so with $\pi_i = \alpha_i + \beta_i + \gamma_i\varphi$ we have

$$\log x_t = \xi(1 - \beta_\bullet - \alpha_\bullet) + \varphi\kappa\alpha_\bullet + \varphi\omega + \sum_{i=1}^{p\vee q}\pi_i \log x_{t-i} + w_t - \sum_{i=1}^{p}(\alpha_i + \beta_i)w_{t-i} + \varphi\sum_{j=1}^{q}\alpha_j v_{t-j}$$

When $\varphi = 0$ the measurement equation shows that $\log x_t$ is an i.i.d. process. □

Proof of Lemma 1

First note that $\frac{\partial g_t'}{\partial\lambda} = (0, \dot{h}_{t-1}, \ldots, \dot{h}_{t-p}, 0_{p+q+1\times q}) =: \dot{H}_{t-1}$. Thus from the GARCH equation $\tilde{h}_t = \lambda' g_t$ we have $\dot{h}_t = \frac{\partial g_t'}{\partial\lambda}\lambda + g_t = \dot{H}_{t-1}\lambda + g_t = \sum_{i=1}^{p}\beta_i\dot{h}_{t-i} + g_t$. Similarly, the second-order derivative is given by

$$\begin{aligned}\ddot{h}_t &= \frac{\partial(g_t + \dot{H}_{t-1}\lambda)}{\partial\lambda'} = \frac{\partial g_t}{\partial\lambda'} + \dot{H}_{t-1} + \frac{H_{t-1}}{\partial\lambda'}\lambda = \dot{H}_{t-1}' + \dot{H}_{t-1} + \sum_{i=1}^{p}\beta_i\frac{\partial\dot{h}_{t-i}}{\partial\lambda'} \\ &= \sum_{i=1}^{p}\beta_i\ddot{h}_{t-i} + \dot{H}_{t-1}' + \dot{H}_{t-1}\end{aligned}$$

For the starting values we observe the following. Regardless of $(h_0, \ldots, h_{p-1})$ being treated as fixed or as a vector of unknown parameters, we have $\dot{h}_s = \ddot{h}_s = 0$. Given the structure of $\ddot{h}_t$ this implies $\ddot{h}_1 = 0$. When $p = q = 1$ it follows immediately that $\dot{h}_t = \sum_{j=0}^{t-1}\beta^j g_{t-j}$. Similarly we have $\ddot{h}_t = \sum_{j=0}^{t-1}\beta^j(\dot{H}_{t-1-j} + \dot{H}_{t-1-j}) = \sum_{j=0}^{t-2}\beta^j(\dot{H}_{t-1-j} + \dot{H}_{t-1-j})$, where $\dot{H}_t = (0_{3\times1}, \dot{h}_t, 0_{3\times1})$ and where the second equality follows by $\dot{H}_0 = 0$. The result now follows from

$$\begin{aligned}\sum_{i=0}^{t-2}\beta^i\dot{h}_{t-1-i} &= \sum_{i=0}^{t-2}\beta^i\sum_{j=0}^{t-1-i-1}\beta^j g_{t-1-i-j} = \sum_{i=0}^{t-2}\beta^i\sum_{k-i-1=0}^{t-i-2}\beta^{k-i-1}g_{t-k} = \sum_{i=0}^{t-2}\sum_{k=i+1}^{t-1}\beta^{k-1}g_{t-k} \\ &= \sum_{k=1}^{t-1}k\beta^{k-1}g_{t-k}\end{aligned}$$

□

DOI: 10.1002/jae

Proof of Proposition 2

Recall that $u_t = \tilde{x}_t - \psi' m_t$ and $\tilde{h}_t = g_t'\lambda$. Thus the derivatives with respect to $\tilde{h}_t$ are given by

$$\frac{\partial z_t}{\partial \tilde{h}_t} = \frac{\partial r_t \exp\left(-\frac{1}{2}\tilde{h}_t\right)}{\partial \tilde{h}_t} = -\frac{1}{2}z_t \quad \text{so that} \quad \frac{\partial z_t^2}{\partial \tilde{h}_t} = -z_t^2,$$

$$\dot{u}_t = \frac{\partial u_t}{\partial \tilde{h}_t} = -\varphi + \frac{1}{2}z_t\tau'\dot{a}_t, \text{ and } \ddot{u}_t = \frac{\partial \dot{u}_t}{\partial \tilde{h}_t} = \frac{\partial\left(-\varphi + \frac{1}{2}z_t\dot{a}(z_t)'\tau\right)}{\partial \tilde{h}_t} = -\frac{1}{4}\tau'(z_t\dot{a}_t + z_t^2\ddot{a}_t)$$

Thus with $\ell_t = -\frac{1}{2}\{\tilde{h}_t + z_t^2 + \log(\sigma_u^2) + u_t^2/\sigma_u^2\}$ we have

$$\frac{\partial \ell_t}{\partial u_t} = 2\frac{u_t}{\sigma_u^2} \qquad \text{and} \qquad \frac{\partial \ell_t}{\partial \tilde{h}_t} = -\frac{1}{2}\left\{1 + \frac{\partial z_t^2}{\partial \tilde{h}_t} + \frac{\partial u_t^2/\partial \tilde{h}_t}{\sigma_u^2}\right\} = -\frac{1}{2}\left\{1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right\}$$

Derivatives with respect to λ are $\frac{\partial z_t}{\partial \lambda} = \frac{\partial z_t}{\partial \tilde{h}_t}\frac{\partial \tilde{h}_t}{\partial \lambda} = -\frac{1}{2}z_t\dot{h}_t$, $\frac{\partial u_t}{\partial \lambda} = \frac{\partial u_t}{\partial \tilde{h}_t}\frac{\partial \tilde{h}_t}{\partial \lambda} = \dot{u}_t\dot{h}_t$, $\frac{\partial \dot{u}_t}{\partial \lambda'} = \ddot{u}_t\dot{h}_t'$, and $\frac{\partial \ell_t}{\partial \lambda} = \frac{\partial \ell_t}{\partial \tilde{h}_t}\dot{h}_t = -\frac{1}{2}\left\{1 - z_t^2 + \frac{2u_t\dot{u}_t}{\sigma_u^2}\right\}\dot{h}_t$. Derivatives with respect to ψ are $\frac{\partial u_t}{\partial \xi} = -1$, $\frac{\partial \dot{u}_t}{\partial \xi} = 0$, $\frac{\partial \ell_t}{\partial \xi} = \frac{\partial \ell_t}{\partial u_t}\frac{\partial u_t}{\partial \xi} = -2\frac{u_t}{\sigma_u^2}$, $\frac{\partial u_t}{\partial \varphi} = -\tilde{h}_t$, $\frac{\partial \dot{u}_t}{\partial \varphi} = -1$, $\frac{\partial \ell_t}{\partial \varphi} = \frac{\partial \ell_t}{\partial u_t}\frac{\partial u_t}{\partial \varphi} = -2\frac{u_t}{\sigma_u^2}\tilde{h}_t$, $\frac{\partial u_t}{\partial \tau} = -a_t$, $\frac{\partial \dot{u}_t}{\partial \tau} = \frac{1}{2}z_t\dot{a}_t$, $\frac{\partial \ell_t}{\partial \tau} = \frac{\partial \ell_t}{\partial u_t}\frac{\partial u_t}{\partial \tau} = -2\frac{u_t}{\sigma_u^2}a_t$. Similarly, $\frac{\partial \ell_t}{\partial \sigma_u^2} = -\frac{1}{2}(\sigma_u^{-2} - u_t^2\sigma_u^{-4})$. Now we turn to the second-order derivatives:

$$-2\frac{\partial^2 \ell_t}{\partial \lambda \partial \lambda'} = \dot{h}_t\left\{-\frac{\partial z_t^2}{\partial \lambda'} + \frac{2}{\sigma_u^2}\left(\dot{u}_t\frac{\partial u_t}{\partial \lambda'} + u_t\frac{\partial \dot{u}_t}{\partial \lambda'}\right)\right\} + \left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)\frac{\partial \dot{h}_t}{\partial \lambda'}$$
$$= \dot{h}_t\left\{z_t^2 + \frac{2}{\sigma_u^2}(\dot{u}_t^2 + u_t\ddot{u}_t)\dot{h}_t'\right\} + \left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)\ddot{h}_t$$

Similarly, since $\frac{\partial z_t}{\partial \psi} = 0$ we have

$$-2\frac{\partial^2 \ell_t}{\partial \lambda \partial \xi} = \frac{\partial\left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)\dot{h}_t}{\partial \xi} = 2\dot{h}_t\left(\frac{\partial u_t}{\partial \psi'}\frac{\dot{u}_t}{\sigma_u^2} + \frac{u_t}{\sigma_u^2}\frac{\partial \dot{u}_t}{\partial \xi}\right) = 2\dot{h}_t\left(-\frac{\dot{u}_t}{\sigma_u^2} + 0\right)$$

$$-2\frac{\partial^2 \ell_t}{\partial \lambda \partial \varphi} = \frac{\partial\left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)\dot{h}_t}{\partial \varphi} = 2\dot{h}_t\left(\frac{\partial u_t}{\partial \varphi}\frac{\dot{u}_t}{\sigma_u^2} + \frac{u_t}{\sigma_u^2}\frac{\partial \dot{u}_t}{\partial \varphi}\right) = 2\dot{h}_t\left(-\tilde{h}_t\frac{\dot{u}_t}{\sigma_u^2} - \frac{u_t}{\sigma_u^2}\right)$$

$$-2\frac{\partial^2 \ell_t}{\partial \lambda \partial \tau'} = 2\dot{h}_t\left(\frac{\partial u_t}{\partial \tau'}\frac{\dot{u}_t}{\sigma_u^2} + \frac{u_t}{\sigma_u^2}\frac{\partial \dot{u}_t}{\partial \tau'}\right) = 2\dot{h}_t\left(-a_t'\frac{\dot{u}_t}{\sigma_u^2} + \frac{u_t}{\sigma_u^2}\frac{1}{2}z_t\dot{a}_t\right)$$

DOI: 10.1002/jae

so that

$$\frac{\partial^2 \ell_t}{\partial\lambda\partial\psi'} = \frac{\dot{u}_t}{\sigma_u^2}\dot{h}_t m_t' + \frac{u_t}{\sigma_u^2}\dot{h}_t b_t', \quad \text{with} \quad b_t = \left(0, 1, -\frac{1}{2}z_t\dot{a}_t'\right)',$$

$$\frac{\partial^2 \ell_t}{\partial\lambda\partial\sigma_u^2} = -\frac{1}{2}\frac{\partial\left(1 - z_t^2 + \frac{2u_t}{\sigma_u^2}\dot{u}_t\right)\dot{h}_t}{\partial\sigma_u^2} = \frac{u_t\dot{u}_t\dot{h}_t}{\sigma_u^4}, \quad \frac{\partial^2 \ell_t}{\partial\psi\partial\psi'} = -\frac{1}{\sigma_u^2}m_t m_t',$$

$$\frac{\partial^2 \ell_t}{\partial\psi\partial\sigma_u^2} == -\frac{1}{2}\left(-\frac{2u_t}{\sigma_u^4}\right)m_t = \frac{u_t}{\sigma_u^4}m_t, \quad \frac{\partial^2 \ell_t}{\partial\sigma_u^2\partial\sigma_u^2} = -\frac{1}{2}\left(\frac{-1}{\sigma_u^4} + 2\frac{u_t^2}{\sigma_u^6}\right) = \frac{1}{2}\frac{\sigma_u^2 - 2u_t^2}{\sigma_u^6}$$

□

Lemma 2. Let $W = \tau_1 Z + \tau_2(Z^2 - 1) + U$, where $Z \sim N(0, 1)$ and $U \sim N(0, \sigma_u^2)$. Then

$$\mathrm{E}(\exp\{\pi^i\gamma W\}) = \frac{1}{\sqrt{1 - 2\pi^i\gamma\tau_2}}\exp\left\{\frac{\pi^{2i}\gamma^2\tau_1^2}{2(1 - 2\pi^i\gamma\tau_2)} - \pi^i\gamma\tau_2 + \frac{\pi^{2i}\gamma^2\sigma_u^2}{2}\right\}$$

Proof. We have $\mathrm{E}(e^{aZ + \frac{b}{2}(Z^2-1)})$ equals

$$\int_{-\infty}^{\infty} e^{az + \frac{b}{2}(z^2-1)}\frac{1}{\sqrt{2\pi}}e^{\frac{z^2}{2}}\mathrm{d}z = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{a^2}{2(1-b)} - \frac{b}{2} - \frac{1}{2}\frac{(z - \frac{a}{1-b})}{(1-b)^{-1}})}\mathrm{d}z = \frac{1}{\sqrt{1-b}}e^{\frac{a^2}{2(1-b)} - \frac{b}{2}}$$

and from the moment-generating function for the Gaussian distribution we have $\mathrm{E}(e^{cU}) = e^{\frac{c^2\sigma_u^2}{2}}$. Since Z and U are independent, we have $\mathrm{E}(\exp\{\pi^i\gamma W\}) = \mathrm{E}(\exp\{\pi^i\gamma\tau_1 Z + \pi^i\gamma\tau_2(Z^2 - 1)\})\mathrm{E}(\exp\{\pi^i\gamma U\}) = \frac{1}{\sqrt{1 - 2\pi^i\gamma\tau_2}}e^{\frac{\pi^{2i}\gamma^2\tau_1^2}{2(1-2\pi^i\gamma\tau_2)} - \pi^i\gamma\tau_2}e^{\frac{\pi^{2i}\gamma^2\sigma_u^2}{2}}$.

□

Proof of Proposition 4

We note that

$$h_t = \exp\left(\sum_{i=0}^{\infty}\pi^i(\mu + \gamma w_{t-1})\right) = e^{\frac{\mu}{1-\pi}}\prod_{i=0}^{\infty}\mathrm{E}\exp(\gamma\pi^i\tau(z_{t-i}))\mathrm{E}(\exp\{\pi^i\gamma u_{t-i}\}),$$

$$h_t^2 = \exp\left(2\sum_{i=0}^{\infty}\pi^i(\mu + \gamma w_{t-1})\right) = e^{\frac{2\mu}{1-\pi}}\prod_{i=0}^{\infty}\mathrm{E}\exp(2\gamma\pi^i\tau(z_{t-i}))\mathrm{E}(\exp\{2\pi^i\gamma u_{t-i}\})$$

and using results such as Lemma 2 and

$$\mathrm{E}\left(\prod_{i=0}^{\infty}\exp\{\pi^i\gamma u_{t-i}\}\right) = \prod_{i=0}^{\infty}\mathrm{E}(\exp\{\pi^i\gamma u_{t-i}\}) = \prod_{i=0}^{\infty}e^{\frac{\pi^{2i}\gamma^2\sigma_u^2}{2}} = e^{\sum_{i=0}^{\infty}\frac{\pi^{2i}\gamma^2\sigma_u^2}{2}} = e^{\frac{\gamma^2\sigma_u^2/2}{1-\pi^2}}$$

we find that $\frac{\mathrm{E}h_t^2}{(\mathrm{E}h_t)^2} = \frac{e^{\frac{2\mu}{1-\pi}}\prod_{i=0}^{\infty}\mathrm{E}\exp(2\gamma\pi^i w_{t-1})}{e^{\frac{2\mu}{1-\pi}}\prod_{i=0}^{\infty}\{\mathrm{E}\exp(\gamma\pi^i w_{t-1})\}^2}$ equals

$$\left(\prod_{i=0}^{\infty}\frac{1-2\pi^i\gamma\tau_2}{\sqrt{1-4\pi^i\gamma\tau_2}}\right)\frac{e^{\sum_{i=0}^{\infty}\frac{4\pi^{2i}\gamma^2\tau_1^2}{2(1-4\pi^i\gamma\tau_2)}}}{e^{2\sum_{i=0}^{\infty}\frac{\pi^{2i}\gamma^2\tau_1^2}{2(1-2\pi^i\gamma\tau_2)}}}\frac{e^{-\frac{2\gamma\tau_2}{1-\pi}}}{e^{-2\frac{\gamma\tau_2}{1-\pi}}}\frac{e^{\frac{2\gamma^2\sigma_u^2}{1-\pi^2}}}{e^{\frac{\gamma^2\sigma_u^2}{1-\pi^2}}}$$
$$=\left(\prod_{i=0}^{\infty}\frac{1-2\pi^i\gamma\tau_2}{\sqrt{1-4\pi^i\gamma\tau_2}}\right)e^{\sum_{i=0}^{\infty}\frac{2\pi^{2i}\gamma^2\tau_1^2}{(1-4\pi^i\gamma\tau_2)}-\frac{\pi^{2i}\gamma^2\tau_1^2}{(1-2\pi^i\gamma\tau_2)}}e^{\frac{\gamma^2\sigma_u^2}{1-\pi^2}}$$
$$=\left(\prod_{i=0}^{\infty}\frac{1-2\pi^i\gamma\tau_2}{\sqrt{1-4\pi^i\gamma\tau_2}}\right)e^{\sum_{i=0}^{\infty}\frac{\pi^{2i}\gamma^2\tau_1^2}{(1-6\pi^i\gamma\tau_2+8\pi^{2i}\gamma^2\tau_2^2)}}e^{\frac{\gamma^2\sigma_u^2}{1-\pi^2}}$$

where the last equality uses that $\frac{2\pi^{2i}\gamma^2\tau_1^2}{(1-4\pi^i\gamma\tau_2)}-\frac{\pi^{2i}\gamma^2\tau_1^2}{(1-2\pi^i\gamma\tau_2)}$ equals

$$\frac{2\pi^{2i}\gamma^2\tau_1^2(1-2\pi^i\gamma\tau_2)-\pi^{2i}\gamma^2\tau_1^2(1-4\pi^i\gamma\tau_2)}{(1-4\pi^i\gamma\tau_2)(1-2\pi^i\gamma\tau_2)}=\frac{\pi^{2i}\gamma^2\tau_1^2}{(1-4\pi^i\gamma\tau_2)(1-2\pi^i\gamma\tau_2)}$$
$$=\frac{\pi^{2i}\gamma^2\tau_1^2}{(1-6\pi^i\gamma\tau_2+8\pi^{2i}\gamma^2\tau_2^2)}$$

□

DOI: 10.1002/jae