UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 09-36
Economic Series (20)
July 2009

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Automated Financial Multi-Path GETS Modelling*

Genaro Sucarrat† and Álvaro Escribano‡

13 July 2009

Abstract

General-to-Specific (GETS) modelling has witnessed major advances over the last decade thanks to the automation of multi-path GETS specification search. However, several scholars have argued that the estimation complexity associated with financial models constitutes an obstacle to multi-path GETS modelling in finance. We provide a result with associated methods that overcome many of the problems, and develop a simple but general and flexible algorithm that automates financial multi-path GETS modelling. Starting from a general model where the mean specification can contain autoregressive (AR) terms and explanatory variables, and where the exponential variance specification can include log-ARCH terms, log-GARCH terms, asymmetry terms, Bernoulli jumps and other explanatory variables, the algorithm we propose returns parsimonious mean and variance specifications, and a fat-tailed distribution of the standardised error if normality is rejected. The finite sample properties of the methods and of the algorithm are studied by means of extensive Monte Carlo simulations, and two empirical applications suggest the methods and algorithm are very useful in practice.

JEL Classification: C32, C51, C52, C53, E44, E47, G17
Keywords: General-to-specific modelling, finance, volatility, value-at-risk

* We are grateful to Luc Bauwens, André Alves Portela Santos and seminar participants at BI (Norwegian School of Management) for helpful comments and suggestions. This research was supported by a Marie Curie Intra-European Fellowship within the 6th. European Community Framework Programme, and by Banco de España.
† Corresponding author. Department of Economics, Universidad Carlos III de Madrid (Spain). Email: gsucarra@eco.uc3m.es. Webpage: http://www.eco.uc3m.es/sucarrat/index.html.
‡ Department of Economics, Universidad Carlos III de Madrid (Spain).

1 Introduction

General-to-Specific (GETS) modelling has witnessed major advances over the last decade thanks to the automation of multi-path GETS specification search, see amongst others Hoover and Perez (1999), Hendry and Krolzig (2001, 2005), Krolzig (2003) and Doornik (2009).[1] Key to this success is that estimation is essentially by means of ordinary least squares (OLS) procedures, something which renders automation of GETS multi-path specification search feasible. By contrast, most financial models are highly non-linear and require complex optimisation algorithms and inference strategies in empirical application. Examples of some of the estimation issues that may need careful attention include multiple optima, numerical approximation, convergence issues, negative variance, initial values, parameter constraints, finite sample approximations, and so on. For models with few parameters this may not pose unsurmountable problems. But as the number of parameters increase the resources and effort needed for reliable estimation and model validation multiply. Indeed, this may even become an obstacle to financial multi-path GETS modelling, as for example argued by Granger and Timmermann (1999), and McAleer (2005) regarding automated GETS modelling of financial volatility.

We provide an analytical result that enables consistent least squares estimation and inference of power log-GARCH models under very general assumptions on the standardised error.[2] This enables flexible methods that overcome many of the estimation and inference complexity issues typically associated with financial models. Next, building on the work on automated GETS modelling by Hoover and Perez (1999), Hendry and Krolzig (2005) and Doornik (2009), and on the study by Bauwens and Sucarrat (2008) on GETS modelling of financial volatility, we propose a general and flexible model framework and develop associated algorithms that automate multi-path GETS modelling. Starting from a general model where the mean specification can include autoregressive (AR) terms and explanatory variables, where the exponential variance specification can include log-ARCH and log-GARCH terms, asymmetry terms, Bernoulli jumps and other explanatory variables, our algorithm returns parsimonious mean and variance specifications, and a generalised error distribution (GED) of the standardised error. (Our algorithm can readily be extended to include other densities in the search space like, say, the standardised Student's t and skewed versions, this is simply a programming issue.) The parameters of the variance specifications in the model we propose are consistently estimated, and inference regarding the parameters is performed by means of standard least squares theory. The Bernoulli jumps in the variance specification can be stochastic in the sense that they are not conditioned upon, so we label the model a stochastic expo-

[1] GETS specification search is closely related to, but not the same as, the GETS methodology, see Campos et al. (2005) for a comprehensive overview of the GETS methodology, and Mizon (1995) for a concise overview.

[2] "GARCH" is short for generalised autoregressive conditional heteroscedasticity, and the origins of the acronym are Engle (1982) and Bollerslev (1986).

nential ARCH (SEARCH) model. The acronym also connotates our main motivation for the model, namely that it facilitates specification search.

The finite sample properties of the methods and algorithms we propose and develop are evaluated by means of extensive Monte Carlo simulation, and by means of two empirical applications. Under the null of financial returns being IID, our multi-path GETS algorithm performs substantially better than the multi-path GETS algorithms of Hoover and Perez (1999), PcGets (Hendry and Krolzig 2005) and Autometrics (Doornik 2009). In particular, our algorithm is less likely to retain irrelevant variables. This comes to some extent at the price of our algorithm's power to detect relevant variables, but we believe a conservative algorithm is more suitable for financial economics and business finance. Financial markets are notoriously difficult to predict *ex ante*, and falsely suggesting predictive or explanatory power by retaining irrelevant variables may potentially lead to substantial losses and damage if the model is used to guide investment decisions by business, or if the model is used to guide policy decisions.

In addition to being a contribution to the GETS specification search literature, the estimation and inference methods we propose is also a contribution to the theoretical literature on exponential volatility models. To our knowledge, there currently exists no general estimation procedure that provides consistent estimates of general classes of exponential variance specifications with ARCH, see Linton (2008), and Mikosch and Straumann (2006).[3] The least squares estimation methods we propose for the power log-GARCH(P, Q) model consistently estimates the parameters of interest under very general assumptions on the standardised error. Also, our simulations suggest the least squares methods compares favourably to Quasi Maximum Likelihood (QML) methods when the standardised error is non-normal.

The rest of the paper is organised into four sections and one appendix. The next section, section 2, outlines the overall statistical framework and the SEARCH model. Section 3 studies the properties of our automated GETS algorithm through extensive Monte Carlo simulation. Section 4 contains two empirical applications of the methods and algorithms. Section 5 concludes, whereas the final part of the paper is an appendix with various supporting information.

[3]The only exception we know off is Dahl and Iglesias (2008), which proves consistency of Gaussian Quasi Maximum Likelihood (QML) for an exponential GARCH(1,1) structure that nests the (second power) log-GARCH(1,1), but not the EGARCH(1,1). Their result is limited, however, since it assumes no mean specification, and since it does not apply to higher order GARCH models, say, log-GARCH(2,1).

2 Statistical framework

2.1 The power log-GARCH model

The log-GARCH model can be viewed as a dynamic version of Harvey's (1976) multiplicative heteroscedasticity model, and the log-GARCH model was proposed by Pantula (1986), Geweke (1986) and Milhøj (1987). One of the main motivations for the log-GARCH is that it ensures non-negative variances. However, it does so at the cost of possibly applying the log-operator on zero-values. If the residuals are rarely equal to zero, then this is not a serious shortcoming in practice. Nevertheless, this problem is not present in the EGARCH model of Nelson (1991), which might explain why so little work has been explicitly devoted to the log-GARCH model since the 1980s.[4]

The power log-GARCH model is given by

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0,1), \quad Prob(z_t = 0) = 0, \quad \sigma_t > 0 \tag{1}$$

$$\log \sigma_t^\delta = \alpha_0 + \sum_{p=1}^{P} \alpha_p \log |\epsilon_{t-p}|^\delta + \sum_{q=1}^{Q} \beta_q \log \sigma_{t-q}^\delta, \quad \delta > 0, \tag{2}$$

where P is the ARCH order, Q is the GARCH order and δ is the power. If $\{\epsilon_t\}$ are residuals resulting from an estimation procedure of a mean equation, then the estimation methods we propose further below are generally valid as long as each residual is a consistent estimate of the corresponding error with probability 1.

The error ϵ_t can be written as $\sigma_t z_t = \sigma_t^* z_t^*$, where

$$\sigma_t^* = \sigma_t E(|z_t|^\delta)^{1/\delta}, \quad z_t^* = \frac{z_t}{E(|z_t|^\delta)^{1/\delta}}, \quad E(|z_t^*|^\delta) = 1. \tag{3}$$

This decomposition is useful because it enables an ARMA representation of the power log-GARCH specification that is readily estimable by means of ordinary estimation methods. For example, the power log-ARCH(1) specification is given by $\log \sigma_t^\delta = \alpha_0 + \alpha_1 \log |\epsilon_{t-1}|^\delta$. Adding $\log E|z_t|^\delta + \log |z_t^*|^\delta$ to each side and then adding $E(\log |z_t|^\delta) - E(\log |z_t|^\delta)$ to the right-hand side, yields the AR(1) representation $\log |\epsilon_t|^\delta = \alpha_0^* + \alpha_1 \log |\epsilon_{t-1}|^\delta + u_t^*$, where $\alpha_0^* = \alpha_0 + \log E|z_t|^\delta + E(\log |z_t^*|^\delta)$, and where $u_t^* = \log |z_t^*|^\delta - E(\log |z_t^*|^\delta)$ is a zero-mean IID process. For known power δ, the parameters α_0^* and α_1 can thus be estimated consistently by means of ordinary estimation methods (least squares, QML, etc.) subject to usual assumptions. In order to recover α_0 we need estimates of $\log E|z_t|^\delta$ and $E(\log |z_t^*|^\delta)$, and the proposition we state below provides very simple formulas for consistent estimation of $\log E|z_t|^\delta$

[4]Some of the theoretical results in later work, see for example Nelson (1991) and Dahl and Iglesias (2008), apply to structures that nest the log-GARCH model. But these works do not have the log-GARCH model as their main focus.

and $E(\log|z_t^*|^\delta)$ under very general assumptions. More generally, the power log-GARCH(P,Q) model with $P \geq Q$ admits the ARMA(P,Q) representation

$$\log|\epsilon_t|^\delta = \alpha_0^* + \sum_{p=1}^{P} \alpha_p^* \log|\epsilon_{t-p}|^\delta + \sum_{q=1}^{Q} \beta_q^* u_{t-q}^* + u_t^*, \tag{4}$$

where

$$\begin{aligned}
\alpha_0^* &= \alpha_0 + (1 - \sum_{q=1}^{Q} \beta_q) \cdot \left[\log E|z_t|^\delta + E(\log|z_t^*|^\delta)\right] \\
\alpha_1^* &= \alpha_1 + \beta_1 \\
&\vdots \\
\alpha_P^* &= \alpha_P + \beta_P \\
\beta_1^* &= -\beta_1 \\
&\vdots \\
\beta_Q^* &= -\beta_Q,
\end{aligned}$$

and where $u_t^* = \log|z_t^*|^\delta - E(\log|z_t^*|^\delta)$ as earlier. When $P > Q$, then $\beta_{Q+1} = \cdots = \beta_P = 0$ by assumption. Also, it should be noted that the equations are not affected by the inclusion of explanatory variables in the log-variance specification (2). The consequence of all this is that consistent estimates of all the ARMA parameters—and hence all the log-GARCH parameters except α_0—can readily be obtained by means of common estimation procedures (least squares, QML, etc.) subject to usual assumptions,[5] as long as the power δ is known, and as long as $P \geq Q$. If $P < Q$, then non-standard, restricted estimation procedures are needed. To see this consider for example a power log-GARCH(0,1) model whose ARMA(1,1) representation is $\log|\epsilon_t|^\delta = \alpha_0^* + \beta_1 \log|\epsilon_{t-1}|^\delta - \beta_1 u_{t-1}^* + u_t^*$ (the AR and MA parts have common roots). It is also worth noting the ease with which certain non-stationary power log-GARCH specifications can be formulated and estimated. For example, an integrated power log-GARCH(1,1) with specification $\log\sigma_t^\delta = \alpha_0 + (1-\beta_1)\log|\epsilon_{t-1}|^\delta + \beta_1 \log\sigma_{t-1}^\delta$ can be written as the MA(1) representation $\Delta\log|\epsilon_t|^\delta = \alpha_0^* + \beta_1^* u_{t-1}^* + u_t^*$. More generally, if $\log\sigma_t^\delta$ is $I(1)$, then the estimates of the stationary $AR(P)$ representation $\Delta\log|\epsilon_t|^\delta = \alpha_0^* + \sum_{p=1}^{P} \alpha_p \Delta\log|\epsilon_{t-p}|^\delta + u_t^*$ can in many cases be used to obtain estimates of the non-stationary representation, or at least as a reasonable approximation.

In order to recover α_0 we need estimates of $\log E(|z_t|^\delta)$ and $E(\log|z_t^*|^\delta)$, and the following proposition gives very general conditions under which they can be estimated consistently after estimation of the ARMA-representation (4).

[5]For example, in the case of estimating the AR(P) representation by means of OLS, the most important assumptions for the current purposes are that the roots of $(1 - \alpha_1 c - \cdots - \alpha_P c^P) = 0$ are outside the unit circle, that $E(u_t^{*2}) < \infty$ and that $E(u_t^{*4}) < \infty$. The conditions $E(u_t^{*2}) < \infty$ and $E(u_t^{*4}) < \infty$ are satisfied when $z_t \sim GED(\tau)$.

Proposition 1. Suppose the power δ is known and that $P \geq Q$. Suppose further that a consistent estimation procedure of the ARMA representation (4) of the power log-GARCH model (1)-(2) exhibits the property $\hat{u}_t^* \xrightarrow{P} u_t^*$ for each t, where $\{\hat{u}_t^*\}$ are estimates of $\{u_t^*\}$. If $0 < E|z_t|^\delta < \infty$ and if $|E(\log|z_t|)| < \infty$ is an event with probability 1, then

a)
$$-\log\left[\frac{1}{T}\sum_{t=1}^{T}\exp(\hat{u}_t^*)\right] \xrightarrow{P} E(\log|z_t^*|^\delta) \tag{5}$$

is an event with probability 1, and

b)
$$-\log\left[\frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^{*2}\right]^{\delta/2} \xrightarrow{P} \log E|z_t|^\delta \tag{6}$$

is an event with probability 1, where $\{\hat{z}_t^*\} = \{\epsilon_t / \sqrt[\delta]{\hat{\sigma}_t^{*\delta}}\}$.

Proof: See appendix.

When the power δ is equal to 2, then $\log E|z_t|^\delta = 0$ and so the second correction b) is not needed. The a) can thus be viewed as a correction due to the application of the logarithm operator, and b) can be viewed as a "power correction". The property $\hat{u}_t^* \xrightarrow{P} u_t^*$ is essentially a consequence of consistent estimation. For the two most common powers, $\delta = 1$ and $\delta = 2$, the proposition holds under very general assumptions. Specifically, the conditions $0 < E|z_t|^\delta < \infty$ and $|E(\log|z_t|)| < \infty$ are satisfied for the most commonly used densities in finance: The Normal, the GED and the Student's t (with an appropriate number of degrees of freedom). It should also be noted that the proposition is likely to hold in many cases if the $\{\epsilon_t\}$ are estimated in a previous step, as long as the estimation procedure exhibits $\hat{\epsilon}_t \xrightarrow{P} \epsilon_t$ for each t. In words, in sufficiently large samples the estimated residuals are distributed as the true errors, and so are the $\{\log|\hat{\epsilon}_t|^\delta\}$ with probability 1 due to continuity. An important example is the case where one fits a power log-ARCH(P) specification to $\log|\hat{\epsilon}_t|^\delta$ by means of OLS. In particular, under well-known and general assumptions the SEARCH model can consistently be estimated by means of OLS when there are no log-GARCH terms in the mean specification. Table 1 contains some simulations that sheds light on the finite sample accuracy of some ordinary estimation methods for selected specifications. Table 2 compares OLS with Gaussian (Q)ML estimation for the log-ARCH(1) case. Although the simulations are admittedly incomplete, they nevertheless suggest that OLS compares favourably to QML when the standardised errors are non-Gaussian. When this is the case, then OLS exhibits smaller finite-sample estimation bias of the parameters, and the estimation variances are comparable to or smaller than those of QML.

2.2 The SEARCH model

The power log-GARCH model is nested by the SEARCH model, and so it is useful to establish a general and compact notation. We do this in terms of a "generalised SEARCH" (GSEARCH) model, which nests the SEARCH model. Indeed, several of the methods we discuss are applicable not only to the SEARCH model, but also to many more general structures contained in the GSEARCH model. The GSEARCH model is given by

$$r_t = \mu(\psi, \mathbf{v}_t) + \epsilon_t, \tag{7}$$

$$E(r_t|\mathcal{I}_t) = \mu(\psi, \mathbf{v}_t), \tag{8}$$

$$\epsilon_t = \sigma_t z_t, \quad z_t|\mathcal{I}_t \sim IID(0,1), \quad \sigma_t > 0, \tag{9}$$

$$\log \sigma_t^2 = g(\gamma, \mathbf{w}_t). \tag{10}$$

The economic interpretation of r_t we have in mind is financial return, but it can of course be interpreted differently. The $\mathbf{v}_t$ is a $(1 \times K)$ vector of contemporaneous and/or lagged variables in the mean function μ, and the property (8) means the expectation of r_t conditional on the information set $\mathcal{I}_t$ (made precise below; for now suffice to say that $\mathbf{v}_t \subset \mathcal{I}_t$) is equal to the mean specification $\mu(\psi, \mathbf{v}_t)$. The $\mathbf{w}_t$ is a $(1 \times L)$ vector of contemporaneous and/or lagged variables in the logarithmic variance specification $\log \sigma_t^2$. The dimensions of the parameter vectors ψ and γ do not necessarily correspond to the dimensions of $\mathbf{v}_t$ and $\mathbf{w}_t$, respectively, which means that μ or g (or both) can be non-linear. The conditioning information $\mathcal{I}_t$ can include both contemporaneous and/or lagged variables, and $\mathbf{v}_t \cup \mathbf{w}_t$ is not necessarily contained in $\mathcal{I}_t$. In particular, if some (or all) of the variables in $\mathbf{w}_t$ are not contained in $\mathcal{I}_t$, that is, if $\mathbf{w}_t - \mathcal{I}_t \neq \emptyset$ where "$-$" is the set-difference operator, then we may refer to σ_t^2 as a stochastic variance or volatility (SV) specification.

If automated multi-path GETS modelling is an objective, then there are limits[6] to the specification of μ and g. An example of a general and flexible structure that is amenable to automated multi-path GETS modelling is what we label the SEARCH model. It is for this model that we develop automated GETS multi-path algorithms for. In words, the SEARCH model can be described as an AR(M) specification with explanatory variables in the mean, and as a power log-GARCH(P, Q) specification with an asymmetry term, Bernoulli jumps and explanatory variables in the

[6]One should maybe add the qualifier "current", because future developments are likely to broaden the class of models that are amenable to automated GETS modelling.

logarithmic variance specification. Specifically, the SEARCH model is given by

$$r_t = \phi_0 + \sum_{m=1}^{M} \phi_m r_{t-m} + \sum_{n=1}^{N} \eta_n x_{nt} + \epsilon_t \tag{11}$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \overset{IID}{\sim} GED(\tau), \quad \tau > 1 \tag{12}$$

$$\begin{aligned} \log \sigma_t^\delta = {} & \alpha_0 + \sum_{p=1}^{P} \alpha_p \log |\epsilon_{t-p}|^\delta + \sum_{q=1}^{Q} \beta_q \log \sigma_{t-q}^\delta + \lambda (\log |\epsilon_{t-1}|^\delta) I_{\{z_{t-1}<0\}} \\ & + (\log \kappa^\delta) J_t + \omega_0 \log EWMA_{t-1} + \sum_{d=1}^{D} \omega_d y_{dt}, \quad P \geq Q, \quad \delta > 0 \end{aligned} \tag{13}$$

$$J_t \in \{0, 1\}, \quad J_t \sim IID, \quad J_t \perp z_t, \quad \kappa \geq 1 \tag{14}$$

In the mean specification (11) ϕ_0 is the mean intercept, M is the number of autoregressive (AR) terms and N is the number of other conditioning variables that may be contemporaneous and/or lagged. Moving average (MA) terms are not included in the mean specification in order to simplify estimation and specification search. However, proposition 1 is valid subject to general conditions if MA terms are included. One thing the proposition does not admit, though, is GARCH-in-mean terms.[7] The standardised errors $\{z_t\}$ are IID and follows a Generalised Error Distribution (GED) with shape parameter τ. Hence, if $\mathbf{v}_t \subset \mathcal{I}_t$, then $E(z_t|\mathcal{I}_t) = 0$ and $Var(z_t|\mathcal{I}_t) = 1$ for all t. When $\tau = 2$ the GED is equal to the standard normal. When $1 < \tau < 2$ then the GED has thicker tails than the standard normal, whereas when $\tau > 2$ then the GED has thinner tails than the standard normal. In particular, when $\tau \to 1$ then the GED tends to a double exponential distribution, and when $\tau \to \infty$ then the GED tends to a uniform distribution on the interval $[-\sqrt{3}, \sqrt{3}]$. An important motivation for the GED is that, in addition to containing the normal as a special case, it allows for *both* fatter and thinner tails than the normal. The former is a common property of financial returns, whereas the latter is a real possibility in explanatory financial return modelling, since the distributional properties of $\{z_t\}$ depends on the explanatory power of the information in the mean and variance specifications (cf. Bauwens and Sucarrat 2008, and Sucarrat 2009). Another advantage of the GED is that, under certain conditions, it ensures that the SEARCH exhibit finite moments, see Nelson (1991). By contrast, a commonly used alternative distribution for which this is not necessarily the case is the t distribution.[8] In the logarithmic variance specification (13), P is the number of log-ARCH terms, Q is the number of log-GARCH

[7] This is not necessarily a serious drawback, since proxies for financial variability (functions of past squared returns, bid-ask spreads, function so high-low values, etc.) that can be included as regressors in the mean are readily available.

[8] On the other hand, an advantage with the t distribution relative to the GED distribution is that the t distribution allows for fatter tails.

terms and δ is the power. λ is the impact of the logarithmic asymmetry term analogous to that of Glosten et al. (1993),[9] $\kappa \geq 1$ is the jump-size ($\kappa = 1$ means no jump) of the Bernoulli jump process $\{J_t\}$, $EWMA_{t-1}$ is an equally weighted moving average computed as $(1/T^*)\sum_{t^*=1}^{T^*}|\epsilon_{t-t^*}|^\delta$ where T^* is the length of the moving average, and D is the number of other conditioning variables that may be contemporaneous and/or lagged.[10] It should be noted that the term $\log EWMA_{t-1}$ can be viewed as a local approximation to $\log \sigma_{t-1}^\delta$. The advantage with $\log EWMA_{t-1}$ compared with $\log \sigma_t^\delta$, though, is that it is more general and flexible, it is simpler to estimate, and ordinary least squares inference can be used regarding the parameter ω_0. The conditioning information set $\mathcal{I}_t$ contains past values of r_t, contemporaneous and past values of the x_{nt}, contemporaneous and past values of σ_t, past values of z_t, contemporaneous and past values of the jumps J_t, and contemporaneous and past values of the y_{dt}. In other words, $\mathcal{I}_t$ contains the contemporaneous jump J_t. When the conditioning information set does not contain the contemporaneous jumps, a case which we will label $\mathcal{I}_t^{sv}$, then the variance specification σ_t^δ is stochastic. However, the contemporaneous independence with z_t implies nevertheless that the moment structure is similar to that of ARCH models when $\delta = 2$: $E(\epsilon_t|\mathcal{I}_t^{sv}) = E(\sigma_t|\mathcal{I}_t^{sv})E(z_t|\mathcal{I}_t^{sv}) = 0$, $Var(r_t|\mathcal{I}_t^{sv}) = Var(\epsilon_t|\mathcal{I}_t^{sv}) = E(\sigma_t^2|\mathcal{I}_t^{sv})E(z_t^2|\mathcal{I}_t^{sv}) = E(\sigma_t^2|\mathcal{I}_t^{sv})$, and conditional variability $E(r_t^2|\mathcal{I}_t^{sv}) = [E(r_t|\mathcal{I}_t^{sv}]^2 + E(\sigma_t^2|\mathcal{I}_t^{sv})$. The conditional variance $E(\sigma_t^2|\mathcal{I}_t^{sv}) = \sigma_t^2$ is equal to $\exp(\alpha_0 + \sum_{p=1}^P \alpha_p \log \epsilon_{t-p}^2 + \sum_{q=1}^Q \beta_q \log \sigma_{t-q}^2 + \lambda(\log \epsilon_{t-1}^2)I_{\{z_{t-1}<0\}} + \sum_{d=1}^D \omega_d y_{dt}) \cdot E(\kappa^{2J_t})$, where the term $E(\kappa^{2J_t}) = [Prob(J_t = 0) + \kappa^2 \cdot Prob(J_t = 1)]$ is due to the fact that the jumps are not in the conditioning set $\mathcal{I}_t^{sv}$. Note that an equivalent parametrisation is simply to move $E(\kappa^{2J_t})$ inside the exponent by adding $\log E(\kappa^{2J_t})$ to the log-variance intercept α_0. It should also be noted that estimation remains straightforward when the jumps are not in the conditioning set. The only changes are that α_0 must be replaced by $\alpha_0 + \log E(\kappa^{2J_t})$, and that z_t must be replaced by $\kappa^{J_t} z_t / \sqrt{E(\kappa^{2J_t})}$. Without jumps, that is, $\log \kappa^2 = 0$ (or alternatively $\kappa = 1$), then the conditional variance $E(\sigma_t^2|\mathcal{I}_t^{sv})$ is equal to $\exp(\alpha_0 + \sum_{p=1}^P \alpha_p \log \epsilon_{t-p}^2 + \sum_{q=1}^Q \beta_q \log \sigma_{t-q}^2 + \lambda \log \epsilon_{t-1}^2 I_{\{z_{t-1}<0\}} + \sum_{d=1}^D \omega_d y_{dt})$. Finally, if $\lambda = \kappa = \omega_1 = \cdots = \omega_D = 0$, then modulus greater than one for all the roots of the polynomial equation $[1 - (\alpha_1 + \beta_1)c - \cdots - (\alpha_P + \beta_P)c^P] = 0$ is a sufficient condition for stability in the variance of r_t.

[9] One may readily include more asymmetry terms, or alternatively consider other asymmetry specifications, for example those of Ding et al. (1993). The question of which approach to asymmetry that is most appropriate we leave for future research.

[10] The parametrisation $(\log \kappa^\delta)$ is attractive because of its exponential counterpart. For example, consider the specification $\log \sigma_t^\delta = \alpha_0 + (\log \kappa^\delta)J_t$. Most of the time, in the absence of jumps (that is, when $J_t = 0$), the conditional standard deviation is $\sigma_t = \exp(\alpha_0/\delta)$. When jumps occur then $\sigma_t = \exp(\alpha_0/\delta)\kappa$. In words, the standard deviation σ_t is a κ multiple of the "normal" standard deviation $\exp(\alpha_0/\delta)$.

2.3 Least squares inference in the SEARCH model

In many practical finance applications the mean is either equal to zero or adequately treated as if equal to zero. Or, alternatively, the residuals from the mean specification are treated *as if* observable. When this is the case, and when the logarithmic variance specification does not contain log-GARCH terms, then inference regarding the parameters γ—apart from the first element α_0—can be undertaken by means of the usual ordinary least squares theory. When log-GARCH terms enter the log-variance specification, then a different inferential approach is needed for the log-ARCH and log-GARCH terms, but not for α_0 and the vector ω.

Suppose no log-GARCH terms enter the log-variance. In this case, if $\mathbf{W}$ is the matrix of observations on the regressors of the logarithmic variance specification (13) fitted to $\log|\epsilon_t|^\delta$, that is, the first column consists of ones and each row of $\mathbf{W}$ is equal to the vector $\mathbf{w}_t$, then the usual test statistic

$$\frac{\hat{\gamma}_k^*}{se(\hat{\gamma}_k^*)} \tag{15}$$

is approximately $N(0,1)$ in large samples for $k = 2, \ldots, K$ under the null of $\gamma_k = 0$. The $\hat{\gamma}_k$ is the OLS estimate of the kth. coefficient, and $se(\hat{\gamma}_k)$ is the kth. element of the diagonal of the ordinary covariance matrix estimate $\hat{\sigma}^2_{u_t^*}(\mathbf{W}'\mathbf{W})^{-1}$. The $\hat{\sigma}^2_{u_t^*}$ is the estimated standard error of u_t^* and equal to $\frac{1}{T-K}\sum_{t=1}^{T}\hat{u}_t^*$. In order to conduct asymptotic inference regarding α_0, we may proceed by means of a Wald parameter restriction test. Specifically, when $\delta = 2$, given the estimate $\hat{\alpha}_0^*$, we may straightforwardly test $\alpha = 0$ by testing whether $\hat{\alpha}_0^*$ is equal to $-\log\hat{E}[\exp(\hat{u}_t^*)]$. The Wald-statistic under the null of $\alpha = 0$ then becomes

$$\frac{\{\hat{\alpha}_0^* + \log\hat{E}[\exp(\hat{u}_t^*)]\}^2}{\widehat{Var}(\hat{\alpha}_0^*)},$$

where $\widehat{Var}(\hat{\alpha}_0^*)$ is the ordinary coefficient variance estimate of α_0^*.

When the logarithmic variance specification contains log-GARCH terms, then one might consider using the usual theory for inference regarding the parameters of the ARMA representation. However, it is doubtful that this theory will be of great value in practice, since the AR and MA coefficient estimates will typically be strongly correlated. An alternative approach is to use the property that a stationary log-GARCH specification is invertible. One may then approximate the log-GARCH specification by means of a (possibly long) log-ARCH specification, and next conduct inference on each of the lags. Another alternative approach is to use an information criterion (in the standardised error $\{z_t\}$) to select between alternative specifications. A third approach is to treat the regressor $\log EWMA_{t-1}$ as a local approximation to $\log\sigma_{t-1}^\delta$, and undertake ordinary inference on the associated parameter ω_0.

Table 3 contains the simulated finite sample size for five coefficient tests when a nominal size of 5% is used, and when $\delta = 2$. The simulations suggest least squares

inference is appropriately sized in finite samples for the first four tests, since the simulated rejection frequencies range between 4% and 6% for most sample sizes. There are signs of over-sizedness in very small samples ($T = 10$) in the test of $\omega = 0$. But this is not serious because the oversizedness is small (between 3% and 4% points), and because the convergence to around 5% is quick. Finally, the fifth coefficient test is a two-sided test of the hypothesis $\alpha_0 = 0$ when $r_t \overset{IID}{\sim} GED(\tau)$, which implies that $\alpha_0 = 0$. The simulations suggest the Wald test is undersized, since the simulated rejection frequencies are close to 0%. Deviations from the normal brings the size closer to the nominal, but the discrepancy is nevertheless still substantial. This might suggest that the test lacks power to reasonable departures from the null of $\alpha_0 = 0$. However, additional simulations (not reported) suggest this is not the case. Even though the Wald test is undersized under the null, the test carries reasonable power even when the departure from the null is small.

3 Financial multi-path GETS modelling

In this section we propose and study a simple but straightforward and very flexible algorithm for financial multi-path GETS modelling of the SEARCH model, in the case where $\delta = 2$. We underline that the algorithm can possibly be improved in numerous ways that suggest themselves naturally, but we leave this for future research. The results in this section should therefore be viewed as a "minimally" achievable starting point or lower bound.

The automated GETS algorithm we propose can be viewed as consisting of three stages, whose starting point is an overall General Unrestricted Model (GUM). That is, a model with general unrestricted mean (MGUM) and variance (VGUM) specifications. The first stage consists of multi-path GETS specification search of the MGUM specification, while the VGUM specification remains unchanged. The second stage consists of multi-path GETS specification search of the VGUM specification while the parsimonious mean specification remains unchanged. Finally, the third stage consists of fitting a GED density to the standardised residuals and testing against normality. If normality is rejected, then the estimated GED density is returned as the distribution of the standardised residuals. The purpose of this section is to study the properties of this algorithm through Monte Carlo simulations.

In the Monte Carlo simulations we will focus on three statistics. Let k_0 denote the number of relevant variables in the GUM, and let k_1 denote the number of irrelevant variables in the GUM. The first statistic is simply the probability $p(DGP)$ of recovering the DGP exactly, that is, the probability of selecting a model such that $p(\hat{k}_0 = k_0, \hat{k}_1 = 0)$. This statistic is intimately related to what in a multiple hypothesis testing context is called the Family-Wise Error (FWE), which is simply the probability of making one or more false rejections.[11] Specifically, in a GETS context

[11] The methods of White (2000), Hansen (2005), and Romano and Wolf (2005) are examples of

the FWE is 1-$p(DGP)$, and consistent model selection implies that $p(DGP)$ tends to 1 as the sample size goes to infinity, or alternatively that the FWE tends to 0. As pointed out by Romano et al. (2008), however, the FWE is rather conservative, and the FWE may at any rate not be the error rate of greatest interest. The two statistics of (arguably) greatest interest in a GETS context are the average relevance proportion $M(\hat{k}_0/k_0)$, which is analogous to statistical power in a hypothesis testing context and which Doornik (2009) calls "potency", an the average irrelevance proportion $M(\hat{k}_1/k_1)$, which is analogous to statistical size in a hypothesis testing context and which Doornik (2009) terms "gauge". These two statistics can be viewed as a more detailed characterisation of the expected value of the False Discovery Proportion (FDP), see Romano et al. (2008).

The section is divided into four subsections. In the first subsection we compare our algorithm with other multi-path GETS algorithms. Since the other algorithms have only been studied in modelling the mean with homoscedastic errors, the focus in that subsection will exclusively be on modelling the mean with homoscedastic errors. The second subsection studies the properties of our algorithm in modelling the mean when the errors are heteroscedastic. The third subsection studies the properties of our algorithm in modelling the conditional variance, whereas the final subsection studies the properties of the automated density modelling part of our algorithm.

3.1 Comparison with other multi-path GETS algorithms

Three multi-path GETS specification search algorithms have previously been studied in the academic literature: The algorithm proposed by Hoover and Perez (1999, henceforth HP), the PcGets algorithm proposed by Hendry and Krolzig (1999, 2001, 2005), and the Autometrics algorithm (Doornik and Hendry 2007, Doornik 2009). The way our algorithm undertakes multi-path GETS specification search in stages 1 and 2 is essentially a straightforward improvement of the HP algorithm. But in order to distinguish our algorithm from that of Hoover and Perez we will refer to our algorithm as AutoSEARCH.[12] The purpose of this subsection is to compare the specification search properties of AutoSEARCH with HP, PcGets and Autometrics. The latter three algorithms have all been developed for and studied in the modelling of a mean specification with homoscedastic residuals, so the simulations in this subsection will exclusively focus on modelling the mean under the assumption of homoscedastic variance. In this case, AutoSEARCH proceeds as follows:

approaches that seek to control the FWE.

[12]We intend to make the code developed for this paper freely available as an (open source) R package with the name AutoSEARCH. This is work in progress (Sucarrat 2009) and those who are interested in the code are encouraged to email the corresponding author. Notification will be given as soon as a beta version of the code—with documentation—is available.

Step 1. Formulate a general unrestricted mean GUM (MGUM) of the form

$$r_t = \mathbf{v}_t\psi + \epsilon_t, \quad \epsilon_t = \sigma z_t, \quad z_t \sim IID(0,1), \tag{16}$$

where $\mathbf{v}_t$ is $(1 \times k)$, and where $\mathbf{v}_t$ includes a constant and possibly lags of r_t and other explanatory variables. By assumption, $k_0 \geq 0$ of the k regressors are relevant, $k_1 \geq 0$ are irrelevant and $k_0 + k_1 + 1 = k$. The "+1" is due to the constant, which is restricted from removal in the simulations of AutoSEARCH.

Step 2. Define the number of paths to be searched as equal to the number of insignificant variables in the GUM. The first insignificant variable constitutes the first variable to be removed in path 1, the second insignificant variable constitutes the first variable to be removed in path 2, and so on.

Step 3. After removal of the first variable in a path in step 2, then subsequent simplification in each path is undertaken using a stepwise step-down procedure ("single-path" GETS), where the regressor with highest p-value is deleted at each removal until all remaining regressors are significant. At each removal the standardised residuals are checked for autocorrelation and ARCH using a Bonferroni correction.[13] If removal induces either autocorrelation or heteroscedasticity (or both), then the variable is re-included and subsequently restricted from removal in the simplification search in that path (but the variable is not restricted from removal in other paths). The single-path GETS is undertaken for each of the paths.

Step 4. Form a list of models that contain the distinct models of the multi-path simplification search in steps 2 and 3. The GUM and an empty model, which contains none of the $k_0 + k_1$ regressors under consideration, are included in the list. But the empty model is included in the list only if it passes the diagnostic tests.

Step 5. Select the best model from the list according to an information criterion (Schwarz is used in the simulations) that is computed using the log-likelihood of the standardised residuals.

The AutoSEARCH algorithm can be viewed as a modified version of the HP algorithm of Hoover and Perez (1999). The most important differences between the two algorithms are two. First, the HP algorithm is restricted to search a maximum of 10 paths, because this—in Hoover and Perez's view—resembled what users of the GETS methodology did in practice (prior to the existence of multi-path GETS specification search software). The number of paths in the AutoSEARCH algorithm

[13]For example, if a nominal level of 5% is chosen, then the autocorrelation and ARCH tests are each checked for significance using a level equal to the chosen nominal level divided by the number of diagnostic tests. Here, the number of tests is two and so the Bonferroni adjusted level is 2.5%. Simulations (not reported) suggests the Bonferroni correction is appropriate as long as the sample size is greater than 50.

by contrast is not limited to 10, but to the number of insignificant variables in the GUM. This change improves the ability to detect relevant variables. The second important difference compared with HP is that AutoSEARCH always tries to include an empty model in the list of models that are compared. The empty model contains none of the considered regressors, but it is included only if it passes the diagnostic tests. The effect of this is that the ability to exclude irrelevant variables is improved, and a much better capacity to recover the simulation DGP when the left-side variable is IID and not under the influence of any of the regressors considered. This capacity is a desirable property, since the IID assumption is a very common and useful structure in many financial applications. Further differences (of less importance) between HP and AutoSEARCH concern which and how many diagnostic checks that are performed.

The first main difference between the PcGets algorithm of Hendry and Krolzig (2005) on the one hand and the HP and AutoSEARCH algorithms on the other, is that PcGets is a "multi-round" algorithm, whereas HP and AutoSEARCH are "single-round" algorithms. Whereas HP and AutoSEARCH select between models from a first-round multi-path GETS simplification search, PcGets goes on to do a second round multi-path search if more than one model results from the first round. Starting from a GUM made up of the union of the models from the first round, PcGets continues the multi-round search until only a single model results, or alternatively until the resulting models are equal to the models from the previous round. The main effect of multi-round search is an increased ability to retain relevant variables. However, it does to some extent come at the cost of excluding irrelevant variables. The Autometrics of Doornik (2009) proceeds similarly to PcGets, but Autometrics searches more paths (by means of a "tree search" method) than PcGets at each round. A second main difference between HP and AutoSEARCH on the one hand and PcGets and Autometrics on the other, is that the latter two are calibrated such that the average irrelevance proportion is equal to the chosen nominal size. In HP and AutoSEARCH by contrast the design philosophy is that the average irrelevance proportion should tend to a value bounded from above by means of the chosen overall diagnostic level as the sample goes to infinity. For example, if the overall chosen diagnostic level is 5% (as in the simulations), then the average irrelevance proportion will tend to a value somewhere between 0 and 0.05. The exact value will depend on the exact nature of the DGP. Similarly, $p(DGP)$ will tend to a value about 1 minus the chosen overall diagnostic level. For example, for an overall diagnostic level is 5%, then $p(DGP)$ will tend to about 1 - 0.05 = 0.95.

In order to compare AutoSEARCH with HP, PcGets and Autometrics, we study AutoSEARCH in Monte Carlo experiments that have previously been run for the three other algorithms. The experiments are listed in table 4, and the results of experiments HP0, HP2' and HP7' are contained in table 5. The first important feature that emerges from the results is that AutoSEARCH fares very well in deleting irrelevant variables compared with the other algorithms. In experiment HP0, where none of the regressors matter, AutoSEARCH recovers the DGP about 94% of time.

Having in mind that the GUM contains as many as 40 variables (all irrelevant) and that the sample size is (only) 139 observations, this is impressive. By contrast, HP and PcGets recover the simulation DGP a maximum of 45% of the time. The simulation results also suggest quite clearly why AutoSEARCH fares so much better than the other algorithms in deleting irrelevant variables. The most important reason is that it includes an empty model (if it passes the diagnostic tests) in the list of considered models. The other reason, which is less important, is that AutoSEARCH by default does not undertake a parsimonious encompassing test (PET). A second important feature of the simulation results is that the average irrelevance proportion $M(\hat{k}_1/k_1)$, a measure of "overfitting", is as low or lower for AutoSEARCH as for the other algorithms. And this is the case across experiments. The ability to remove irrelevant variables is likely to come at the cost of the ability to retain variables that matter. However, in the experiments that contain variables that matter, experiments HP2' and HP7', AutoSEARCH performs equally well in retaining relevant variables as measured by the average relevance proportion $M(\hat{k}_0/k_0)$. Nevertheless, it should be pointed out though that in experiments HP2' and HP7' the signal of the variables that matter is relatively high. So possibly a different experimental design is needed in order to provide a more accurate comparison of the relative potency of the algorithms.

3.2 Multi-path GETS of the mean with heteroscedastic errors

When modelling financial returns the errors very often remain heteroscedastic even after including explanatory information in the mean specification. So it is of interest to study the properties of multi-path GETS when the errors are heteroscedastic.[14] In doing so, we modify our multi-path algorithm from the previous section in three straightforward ways. First, we use White (1980) standard errors when computing the coefficient test-statistics instead of the ordinary standard errors. Second, we turn off ARCH diagnostic checking and designate all the diagnostic checking significance level (5% in the simulations) to the test for serial correlation. Third and finally, we use the Gaussian likelihood of the standardised residuals when computing the information criterion.[15]

Table 6 contains the results of four Monte Carlo experiments, all with a persistent log-GARCH(1,1) specification on the errors as part of the simulation DGP. The first two experiments, HP0* and HP2'*, are essentially equal to HP0 and HP2', respectively, but for the heteroscedastic errors.[16] The results of HP0* are very

[14]To our knowledge, no one has studied the properties of multi-path GETS specification search of the mean when the errors are heteroscedastic.

[15]In additional simulations we studied whether modelling the variance and fitting a GED log-likelihood to the standardised residuals improved the algorithm. The simulations did not suggest a clear performance gain.

[16]In HP0* and HP2'* the parameter α_0 of the log-GARCH specification has been calibrated.

similar to those of HP0. This suggests that persistent heteroscedasticity of the ARCH type does not affect multi-path GETS specification search very much when no regressor in the GUM matters. When regressors do matter, though, as in HP2'*, then heteroscedasticity worsens the ability to delete irrelevant variables. Specifically, the irrelevance proportion increases and $p(DGP)$ falls. The relevance proportion is unaffected, but this result may be misleading since the signal of the regressor is very high. Moreover, there are two other aspects of the DGP in HP2'* that suggest the results may not be so indicative of how multi-path GETS actually performs in modelling financial returns. Firstly, the DGP in HP2'* is very persistent in the mean since the AR(1) coefficient is as high as 0.75 (this affects the relevance and irrelevance proportions). In empirical finance, however, the AR(1) coefficient is typically equal to zero or rarely higher than 0.1 in absolute value. Secondly, one may argue that a sample size of 139 observations is small in financial contexts. These two aspects motivates two additional Monte Carlo experiments, namely SE1 and SE2, which better correspond to a modelling situation of financial return. In the DGP of SE1 there are no relevant regressors, whereas in SE2 there is one, namely an AR(1) term with coefficient equal to 0.1. When no regressors matter, as in SE1, then the previous results of HP2'* are confirmed. Multi-path GETS recovers the DGP with a probability equal to about 0.9, and the irrelevance proportion remains below the chosen regressor significance level of 5%. Also in SE2 does the irrelevance proportion remain below the chosen regressor significance level. However, $p(DGP)$ is relatively low in small samples. For example, a $p(DGP)$ equal to 17% when $T = 200$ in SE2 does not suggest the algorithm works very differently from HP2'*, where $p(DGP)$ is equal to 15% when $T = 139$. In other words, although both the DGPs and GUMs differ substantially between HP2'* and SE2, the algorithm exhibits similar properties in small samples.

Specifically, the limit of the unconditional variance of a model with log-GARCH(1,1) specification $\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \epsilon_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2$ is

$$\exp\left(\frac{\alpha_0}{1-\alpha_1-\beta_1}\right) \cdot \lim_{Q\to\infty} \prod_{q=1}^{Q} E\left[z_{t-q}^{2\alpha_1(\alpha_1+\beta_1)^{q-1}}\right].$$

Numerical simulation suggests the limit of the power term is approximately equal to 0.3167 when $z_t \sim N(0,1)$ for $\alpha_1 = 0.1$ and $\beta_1 = 0.8$. In order to calibrate the DGP in HP0* such that its unconditional variance equals the constant variance of the DGP in HP0, we thus need to solve

$$130^2 = \exp\left(\frac{\alpha_0}{1-\alpha_1-\beta_1}\right) \cdot 0.3167$$

for α_0, which yields $\alpha_0 \approx 1.0885$. Similarly, in HP2'* we obtain $\alpha_0 \approx 1.0058$.

3.3 Multi-path GETS modelling of the log-variance

The purpose of this subsection is to study the properties of AutoSEARCH in modelling the log-variance specification. Specification search of the variance proceeds in a similar way to specification search of the mean, but for one difference. AutoSEARCH undertakes diagnostic checks of the standardised residuals $\{\hat{z}_t\}$ instead of the residuals $\{\hat{u}_t^*\}$ of the fitted ARMA representation of $\log \epsilon_t^2$. Table 7 contains the simulation results of the four experiments SE3 to SE6. In experiment SE3 the simulation DGP contains no relevant variables, whereas in SE4 the simulation DGP contains a single relevant variable, a log-ARCH(1) term, with a moderate value of $\alpha_1 = 0.2$. These two experiments start from a GUM that contains five log-ARCH terms, an asymmetry term, the contemporaneous and lagged variable of an unrelated but strongly persistent AR(1) process, two standard normal IID processes and two exponentially distributed IID processes with shape parameter equal to 1. The strongly persistent process may be interpreted as proxying variables that may explain the level of volatility (say, the level of trading volume), the normally distributed variables may be interpreted as proxying "short-term" relative changes in, say, trading volume, whereas the exponentially distributed variables may be interpreted as proxying the square of, say, stock market return, interest rate changes, or similarly. The objective of experiments SE5 and SE6 is to study the properties of a very common situation in financial practice, namely that of fitting a GARCH structure to a financial return series. In both SE5 and SE6 the GUM is a log-GARCH(1,1) model, but the simulation DGPs differ. In the former $r_t \overset{IID}{\sim} GED(\tau)$ which implies that $\alpha_0 = 0$, whereas in the latter the simulation DGP is a persistent log-GARCH(1,1) with $\log \sigma_t^2 = \alpha_0 = 0, \alpha_1 = 0.1, \beta_1 = 0.8$ and $z_t \overset{IID}{\sim} GED(\tau)$.

The first main property of table 7 is that the irrelevance proportion is either approximately equal to or lower than the chosen nominal level across experiments, and that this property is robust to the tail-thickness of the standardised error. In other words, the asymptotic property holds in finite samples across the four experiments. The second main property is that in SE3 and SE5 the DGPs are recovered with a probability equal to or almost equal to 1 - *the overall diagnostic significance level*, even in small samples. With an overall diagnostic significance level of 5%, the asymptotic probability of recovering the DGP is 0.95, and the results suggests that this probability is reached very fast when there are no variables that matter. In particular, the probability of retaining a log-GARCH(1,1) model when in fact returns are IID—thick-tailed or not, is a maximum of 6.4% when the sample is as small as 200 observations. By contrast, in SE4 and SE6, where there are variables that matter, then $p(DGP)$ is more variable as it depends to a greater extent on sample size, tail thickness and experiment. Specifically, the minimum $p(DGP)$ is 0.30 in SE4 when $T = 200$, and the maximum is 0.97 in SE6 when $T = 1000$. The third and final main property of the results is that tail-thickness has a notable bearing upon the ability to recover the DGP. The more thick-tailed, the more difficult. This is the case both when the DGP contains relevant variables and

when it does not, but the effect is greater when there are variables that matter.

3.4 Modelling of the density of the standardised error

A consistent estimate of τ that is efficient and numerically robust can be obtained as the inverse of a moment estimator of a kurtosis index, see Mineo (2003, 2008), and Mineo and Ruggieri (2005). The estimator is based on the observation that

$$VI = \frac{\sqrt{m_2}}{m_1} = \frac{\sqrt{\Gamma(1/\tau)\Gamma(3/\tau)}}{\Gamma(2/\tau)} \tag{17}$$

is an index of kurtosis, where m_r is the absolute moment of grade r. The sample counterpart of (17) is

$$\widehat{VI} = \frac{\sqrt{T\sum_{t=1}^{T} z_t^2}}{\sum_{t=1}^{T} |z_t|}, \tag{18}$$

and the simulations of Mineo (2003) shows that the inverse of (18) compares favourably to other estimators of τ, including maximum likelihood (ML) methods.

Given consistent estimates $\{\hat{z}_t\}$ of the standardised errors $\{z_t\}$, we may thus undertake asymptotically valid inference regarding the value of τ in terms of an LR-test. Table 8 compares the finite sample properties of the LR-test with three other normality tests, namely Jarque and Bera (1980), Anscombe and Glynn (1983), and Bonett and Seier (2002). The first is the commonly used joint test for skewness and kurtosis in excess of that of the normal, whereas the next two are tests for only kurtosis in excess of that of the normal. Overall, the results suggest that the likelihood ratio test is the preferred test. Although somewhat undersized under the null in small samples (up to $T = 100$), the test has substantially greater power than the three other tests to detect fat-tailed departures ($\tau < 2$) from the null. This comes to some extent at the cost of detecting thin-tailed distributions ($\tau > 2$), since in these cases the Anscombe-Glynn test is more powerful. However, the Anscombe-Glynn test is not as powerful in detecting fat-tailedness. Since financial data typically are more fat-tailed than the normal, a test that is more powerful to detect fat-tailed alternatives is preferred.

4 Empirical evaluation

In this section we assess the methods and algorithm through two empirical applications.

4.1 How well do volatility proxies forecast variability?

Volatility is by definition a conditional forecast of price variability when the conditional mean is zero. The economic meaning of a mean equal to zero is essentially

that the direction of the financial price change is unpredictable, which explains the importance of volatility forecasts in derivative pricing. Indeed, volatility forecasts are arguably the most important inputs in derivative pricing, and so volatility forecasting is of great importance in the financial industry.

The volatility forecasting literature has experienced major developments over the last decade. One of the developments is the increased production, dispersion and availability of high-frequency data, and the increased and cheaper computing power to handle the larger datasets. A second development of great importance is theoretical. The last ten years have witnessed many theoretical contributions that enables efficient volatility forecasting by making use of high-frequency data. The most well-known of the estimators is realised volatility (RV, sums of squared intra-period high-frequency returns), but numerous relatives have also been proposed and studied. How well do all these volatility proxies actually forecast price variability? If the underlying continuous time model is a valid or "true" representation of the DGP in some appropriate sense—this is effectively the assumption RV and its cousins rely upon, then this has three important implications. First, the standardised residuals defined as $\hat{z}_t = r_t/\sqrt{RV}_t$ should be serially uncorrelated and exhibit no ARCH. Second, the coefficient restrictions $\alpha_0 = 0$ and $\omega = 1$ in the SEARCH specification $\log \sigma_t^2 = \alpha_0 + \omega \log RV_t$ should not be rejected. Third, RV_t should parsimoniously encompass models that make use of the same data. If it does not, then this means the other models make more efficient use of the data.

The first two implications are readily investigated via logarithmic Mincer-Zarnowitz regressions (MZ), which amounts to fitting

$$\log \sigma_t^2 = \alpha_0 + \omega \log RV_t. \tag{19}$$

Next, the hypotheses of no serial correlation and ARCH in the standardised error, and whether $\alpha_0 = 0$ and $\omega = 1$, can readily be tested. Table 9 contains logarithmic MZ-regressions of daily stock return (IBM) on three different volatility proxies. This data series is of interest because Patton (2008) use them to illustrate how volatility proxies can improve volatility forecast evaluation. However, table 9 shows that the first and second proxies employed by Patton both invalidate the hypotheses of no serial correlation and no ARCH in the standardised residuals. And at 6% percent or lower the nulls of $\alpha_0 = 0$ and $\omega = 1$ are rejected for the third volatility proxy as well. In other words, the basic diagnostic tests and the coefficient restriction tests do not suggest that the theory upon which the volatility proxies is based on holds. The only candidate that comes close to satisfy the basic diagnostics is the third proxy, that is, RV made up of 5-minute intra-day returns.

Whether a volatility proxy parsimoniously encompasses other models that make use of the same data is readily investigated by means of automated multi-path GETS modelling. Table 10 contains the results of a parsimonious encompassing test for the third volatility proxy. MGUM is the general and unrestricted mean specification, whereas VGUM1 and VGUM2 are two different variance GUMs. VGUM1 contains

only the constant and log RV_t^{5m} as regressor, and the ARCH diagnostic test suggests $\log RV_t^{5m}$ does not capture all the volatility persistence. This motivates VGUM2 where we add log-ARCH lags and an asymmetry term to the VGUM. This improves the ARCH diagnostics. MSPEC and VSPEC are the specifications obtained after multi-path GETS specification search of MGUM and VGUM2: First MSPEC is obtained by holding the variance specification constant and equal to VGUM2, and next VSPEC is obtained by holding the mean specification constant and equal to MSPEC. The conclusion is that RV_t^{5m} does not parsimoniously encompasses the other terms. A positive and significant constant is retained in the specific mean equation, which means there is a slight daily drift upwards of about 0.075% in the IBM stock value over the sample. Also, a 4th. order log-ARCH term is statistically significant in the variance specification in addition to $\log RV_t^{5m}$.

These two simple exercises show that the SEARCH model can be useful in testing volatility proxies for bias and parsimonious encompassing. Also, if RV does not pass the diagnostic tests, and/or if RV does not parsimoniously encompass the alternative information available, then it can still be used as a regressor in the development of *ex post* volatility models that passes the diagnostic tests of relevance.

4.2 Value-at-Risk forecasting

Value-at-Risk (VaR) analysis plays an important role in financial decision making, and here we evaluate the methods and algorithm in forecasting the VaR of the daily Standard & Poor's 500 (SP500) US stock market index. This index appears frequently in VaR forecast evaluations, so it suggests itself naturally for comparison purposes. In our forecast evaluation we divide the sample in two. The estimation and design sample runs from Monday 1 January 2001 to Friday 30 December 2005 (1305 observations), and the forecast evaluation sample runs from Monday 2 January 2006 to Tuesday 24 February 2009 (821 observations).[17] No re-estimation is undertaken after 2005, so the forecast comparison is a true out-of-sample evaluation. Six models are included in our comparison. The first model is a constant model of volatility equal to the sample variance in the estimation and design sample. The second model is a twenty-day equally weighted moving average (EWMA) of the sample variance. Twenty trading days corresponds to four calendar weeks, and EWMA at day t is computed as $(\sum_{i=1}^{20} r_{t-i}^2)/19$, where r_t is the daily end-of-day log-return of SP500 at t. It should be noted that the EWMA can be viewed as a restricted SEARCH model with α_0 equal to zero and the coefficient value of

[17]The choice of the sample dates is the result of several considerations. The sample is relatively recent in that it does not include periods that differ substantially from today's financial regulatory and institutional architecture, and the sample contains both upswings and downturns. So the results of the forecast evaluation exercise are likely to be indicative of how the models would fare in the future. Moreover, both the estimation and forecast samples are sufficiently large for the asymptotic theory that several of the methods rely upon to be adequate. Finally, including the periods of recent financial turmoil means the evaluation provides a real test of the methods.

$\log \text{EWMA}_t$ equal to one. The third model is an integrated GARCH(1,1) model with parameter values known as the "RiskMetrics" specification. The fourth model is a stationary but highly persistent GARCH(1,1) model, whereas the fifth model is a stationary but highly persistent log-GARCH(1,1) model. The final specification is a parsimonious SEARCH model obtained through multi-path GETS specification search of the general and unrestricted log-variance specification

$$\begin{aligned}\log \sigma_t^2 &= \alpha_0 + \sum_{p=1}^{5} \alpha_p \log \epsilon_{t-p}^2 + \lambda(\log \epsilon_{t-1}^2) I_{z_{t-1}<0} \\ &\quad + \omega_1 \log \text{EWMA}_t + \omega_2 tue_t + \omega_3 wed_t + \omega_4 thu_t + \omega_5 fri_t.\end{aligned}$$

In words, the GUM nests the EWMA and a log-ARCH(5) model with asymmetry, and in addition the GUM also contains the impulse dummies tue_t, wed_t, thu_t and fri_t that accommodates the possibility of deterministic day-of-the-week dependency in the conditional variance.

Table 11 contains the estimation results and associated diagnostic tests of the six models, and figure 1 plots the standardised residuals of the six models over the whole sample. The figure is particularly informative. As expected, the figure shows that the constant model does not capture the time-varying conditional variance of SP500 returns, since there is clear evidence of ARCH in the standardised residuals both in-sample and out-of-sample. The two simplest of the time-varying volatility models, however, the EWMA and RiskMetrics, perform remarkably well. Visually, the variance of their standardised residuals is constant across *both* the estimation and forecast samples. This is impressive when one haves in mind all the financial turmoil that has occurred during the forecast period. By contrast, the two more "advanced" models, the GARCH(1,1) and log-GARCH(1,1) specifications, both exhibit clear structural breaks going from the estimation sample to the forecast sample. Finally, the parsimonious SEARCH model obtained through multi-path GETS, which is essentially an unrestricted EWMA, also exhibits relatively constant variance (visually) in the standardised residuals. This suggest that the EWMA, RiskMetrics and SEARCH specifications are more likely to do well in the out-of-sample comparison.

Table 12 contains the forecast evaluation results, which follows the methodology proposed by Christoffersen (1998). The upper panel contains the results when the forecasted density of the standardised error is specified as standard normal, whereas the lower panel contains the results when the forecasted density is specified as a GED with estimated shape parameter τ. The best performing models in both panels are EWMA, RiskMetrics and SEARCH, that is, the models that exhibit relatively stable standardised residuals over the whole sample. The GARCH(1,1) and log-GARCH(1,1) specifications are the worst, even worse than the constant variance model. This underlines the importance of *properly* accounting for the time-varying ARCH in the conditional variance. The results strongly suggest that stationary but strongly persistent GARCH(1,1) and log-GARCH specifications are not capable of this. The results of EWMA, RiskMetrics and SEARCH are very similar across the

panels, and they are usefully close to the density quantiles. However, since the estimated out-of-sample coverages are sometimes significantly different (at conventional significance levels) from their quantile values in both panels, there is still room for further improvement of the density modelling part of the algorithm. But overall the out-of-sample VaR forecast evaluation suggests the methods and algorithms can be of great value in practice, as long as the variance GUM nests a specification that adequately describes the time-varying volatility process.

5 Conclusions

We have proposed methods that resolves many of the problems earlier faced in the implementation of automated multi-path GETS specification search of financial models. The simulations and empirical applications suggest the methods can be of great value in practice. However, there is still substantial room for further improvement, generalisation and exploration. Here, we only suggest five lines of further research and exploration, several of which is already work in progress by the authors. First, a line of research that naturally suggests itself is to improve the efficiency of the least squares estimation procedures, both in the variance and in the mean specifications, for example through feasible generalised least squares and/or iterated least squares procedures. Similarly, there are numerous possibilities to explore in an attempt to improve inference while trying to keep estimation simple. Indeed, as recently stressed by Hamilton (2008), the importance of appropriately accounting for ARCH in the estimation and inference about the mean is possibly even more important in macroeconomics. Second, key to our methods is the proposition that enables consistent estimation of the power log-GARCH structures by means of standard estimation procedures, subject to very general conditions on the distribution of the standardised error. A consequence of this is that certain classes of multivariate power log-GARCH models can be straightforwardly estimated by means of least squares through the associated VARMA representation. This may prove very useful since multivariate ARCH models are very difficult to estimate in practice. A third line of possible research is to study in more depth the properties of logarithmic Mincer-Zarnowitz regressions. Such regression are straightforwardly implemented due to our results, as in our first empirical application, since they are nested within the SEARCH model. A fourth line of possible research is to experiment with the search order. For our algorithm we chose the simplest and most straightforward order, namely to first model the mean, then the variance, and then finally the density. Another option that seems natural to consider is to model the variance first, and then use the variance-specification to improve inference in the mean, before finally modelling the density. Or, alternatively, a third option is to model the mean and variance simultaneously, before modelling the density. There are numerous options and strategies that one may consider. Finally, the ultimate test for any new method is usefulness in econometric practice, and there are nu-

merous applications that suggest themselves in addition to those of section 4. An example of particular interest is the modelling of electricity prices, where both the mean and variance specifications typically require many terms in both the mean and variance equations, see Escribano et al. (2002).

References

Anscombe, F. and W. Glynn (1983). Distribution of kurtosis statistic for normal statistics. *Biometrica 70*, 227–234.

Bauwens, L. and G. Sucarrat (2008). General to Specific Modelling of Exchange Rate Volatility: A Forecast Evaluation. Forthcoming in the *International Journal of Forecasting*. UC3M Working Paper version: WP 08-18 in the Economic Series (`http://hdl.handle.net/10016/2591`).

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327.

Bonett, D. and E. Seier (2002). A test of normality with high uniform power. *Computational Statistics and Data Analysis 40*, 435–445.

Box, G. E. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. London: Addison-Wesley Publishing Company.

Campos, J., D. F. Hendry, and N. R. Ericsson (Eds.) (2005). *General-to-Specific Modeling. Volumes 1 and 2*. Cheltenham: Edward Elgar Publishing.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review 39*, 841–862.

Dahl, C. M. and E. M. Iglesias (2008). The limiting properties of the QML in a general class of asymmetric volatility models. Unpublished manuscript, currently available as `https://www.msu.edu/~iglesia5/DahlIglesias2009.pdf`.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford University Press.

Ding, Z., R. Engle, and C. Granger (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*, 83–106.

Doornik, J. (forthcoming in 2009). Autometrics. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press. Manuscript (11 March 2008 version), Economics Department, University of Oxford.

Doornik, J. A. and D. F. Hendry (2007). *Empirical Econometric Modelling - PcGive 12: Volume I*. London: Timberlake Consultants Ltd.

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations. *Econometrica 50*, 987–1008.

Escribano, Á., J. I. Peña, and P. Villaplana (2002). Modelling Electricity Prices: International Evidence. UC3M Working Paper 02-27 in the Economic Series. Available as `http://docubib.uc3m.es/WORKINGPAPERS/WE/we022708.pdf`.

Geweke, J. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 57–61.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 1779–1801.

Granger, C. W. and A. Timmermann (1999). Data mining with local model specification uncertainty: a discussion of Hoover and Perez. *Econometrics Journal 2*, 220–225.

Hamilton, J. D. (2008). Macroeconomics and ARCH. In T. Bollerslev, J. R. Russell, and M. Watson (Eds.), *Festschrift in Honor of Robert F. Engle*. Forthcoming.

Hansen, P. R. (2005). A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics 23*, 365–380.

Harvey, A. C. (1976). Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica 44*, 461–465.

Harvey, A. C. (1990). *The Econometric Analysis of Time Series*. London: Philip Allan. Second Edition.

Hendry, D. F. and H.-M. Krolzig (1999). Improving on 'Data Mining Reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal 2*, 202–219.

Hendry, D. F. and H.-M. Krolzig (2001). *Automatic Econometric Model Selection using PcGets*. London: Timberlake Consultants Press.

Hendry, D. F. and H.-M. Krolzig (2005). The Properties of Automatic Gets Modelling. *Economic Journal 115*, C32–C61.

Hoover, K. D. and S. J. Perez (1999). Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. *Econometrics Journal 2*, 167–191. Dataset and code: `http://www.csus.edu/indiv/p/perezs/Data/data.htm`.

Jarque, C. and A. Bera (1980). Efficient Tests for Normality, Homoskedasticity, and Serial Independence. *Economics Letters 6*, pp. 255–259.

Krolzig, H.-M. (2003). General-to-Specific Model Selection Procedures for Structural Vector Autoregressions. *Oxford Bulletin of Economics and Statistics 65*, 803–819.

Linton, O. (2008). ARCH Models. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics Online, 2nd. Edition*. The New Palgrave. `http://www.dictionaryofeconomics.com/`.

Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika 66*, 265–270.

McAleer, M. (2005). Automated inference and learning in modeling financial volatility. *Econometric Theory 21*, 232–261.

Mikosch, T. and D. Straumann (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach. *The Annals of Statistics 34*, 2449–2495.

Milhøj, A. (1987). A Multiplicative Parametrization of ARCH Models. Research Report 101.

Mineo, A. M. (2003). On the Estimation of the Structure Parameter of a Normal Distribution of Order *p*. *Statistica 63*, 109–122. Available as `http://rivista-statistica.cib.unibo.it/article/viewFile/342/335/`.

Mineo, A. M. (2008). *The Normalp Package*. Available as `http://cran.r-project.org/web/packages/normalp/normalp.pdf`.

Mineo, A. M. and M. Ruggieri (2005). A Software Tool for the Exponential Power Distribution: The normalp package. *Journal of Statistical Software 12*, 1–24. Available as `http://www.jstatsoft.org/v12/i04/paper/`.

Mizon, G. (1995). Progressive Modeling of Macroeconomic Time Series: The LSE Methodology. In K. D. Hoover (Ed.), *Macroeconometrics. Developments, Tensions and Prospects*. Kluwer Academic Publishers.

Nelson, D. B. (1991). Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica 51*, 485–505.

Pantula, S. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 71–73.

Patton, A. J. (2008). Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies. Forthcoming in *The Journal of Econometrics*. Code and data: `http://www.economics.ox.ac.uk/members/andrew.patton/Patton_robust_loss_apr06.zip`.

Romano, J. P., A. Shaikh, and M. Wolf (2008). Formalized data snooping based on generalized error rates. *Econometric Theory 24*, 404–447.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*, 1237–1282.

Sucarrat, G. (2009). Forecast Evaluation of Explanatory Models of Financial Variability. *Economics – The Open-Access, Open-Assessment E-Journal 3*. Available via: `http://www.economics-ejournal.org/economics/journalarticles/2009-8`.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. *Econometrica 48*, 817–838.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica 68*, 1097–1126.

Appendix

The generalised error distribution

The generalised error distribution (GED) normalised to have a mean of zero and a variance of one can be parametrised as

$$f(z;\tau) = \frac{\tau\Gamma(3/\tau)^{1/2}}{2\Gamma(1/\tau)^{3/2}} \exp\left[-|z|^{\tau} \cdot \left(\frac{\Gamma(3/\tau)}{\Gamma(1/\tau)}\right)^{\tau/2}\right], \quad -\infty < z < \infty, \;\; \tau > 1, \qquad (20)$$

where $\Gamma(1/\tau)$ is the Gamma-function equal to $\int_0^{\infty} t^{(1/\tau)-1}\exp(-t)dt$. The parameter τ is a shape or "tail-thickness" parameter. When $\tau = 2$ then (20) equals the standard normal, when $\tau < 2$ then (20) has thicker tails than the standard normal, and when $\tau > 2$ then (20) has thinner tails than the standard normal. In particular, when $\tau \to 1$ then (20) tends to the double exponential distribution, and when $\tau \to \infty$ then (20) tends to a uniform distribution on the interval $[-3^{(1/2)}, 3^{(1/2)}]$ (Nelson 1991, pp. 352-353). The distribution (20) is a special case of the exponential power distribution, and the generation of random variates, computation of distribution values, etc., are readily performed by means of the R package normalp, see Mineo (2008).[18] In that package the density of the exponential power distribution is given by

$$f(x,\mu,\sigma,p) = \frac{1}{2p^{1/p}\Gamma(1+1/p)\sigma} \exp\left(\frac{-|x-\mu|^p}{p\sigma^p}\right), \qquad (21)$$

[18]Further analysis of the exponential power distribution is contained in, amongst others, Box and Tiao (1973, pp. 156-160), Harvey (1990, pp. 117-118), Nelson (1991, pp. 352-353), Mineo (2008), and Mineo and Ruggieri (2005).

and the GED(τ) density given by (20) is obtained by setting $\tau = p$, $\mu = 0$ and

$$\sigma = \frac{\Gamma(1/\tau)^{(1/2)}}{\tau^{(1/\tau)}\Gamma(3/\tau)^{(1/2)}}. \tag{22}$$

For estimation of τ see Mineo (2003), and Mineo and Ruggieri (2005).

Proof of proposition 1

In proving a), we first show that $\log E[\exp(u_t^*)] = -E(\log|z_t^*|^\delta)$, then that $\frac{1}{T}\sum_{t=1}^T \exp(\hat{u}_t^*) \xrightarrow{P} E[\exp(u_t^*)]$. Since $u_t^* = \log|z_t^*|^\delta - E(\log|z_t^*|^\delta)$ straightforward algebra yields

$$\begin{aligned}
\log E[\exp(u_t^*)] &= \log E\{\exp[\log|z_t^*|^\delta - E(\log|z_t^*|^\delta)]\} \\
&= \log E\left\{\frac{|z_t^*|^\delta}{\exp[E(\log|z_t^*|^\delta)]}\right\} \\
&= \log\left\{\frac{E|z_t^*|^\delta}{\exp[E(\log|z_t^*|^\delta)]}\right\} \\
&= \log E|z_t^*|^\delta - E(\log|z_t^*|^\delta) \\
&= -E(\log|z_t^*|^\delta),
\end{aligned}$$

since $E|z_t^*|^\delta = 1$ and since $|E(\log|z_t^*|^\delta)| < \infty$. The latter follows from the assumptions $0 < E|z_t|^\delta < \infty$ and $|E(\log|z_t|)| < \infty$. Accordingly, $(-1)\cdot\log E[\exp(u_t^*)] = E(\log|z_t^*|^\delta)$. We now turn to the proof of $\frac{1}{T}\sum_{t=1}^T \exp(\hat{u}_t^*) \xrightarrow{P} E[\exp(u_t^*)]$. We have that $\frac{1}{T}\sum_{t=1}^T \exp(u_t^*) \xrightarrow{P} E[\exp(u_t^*)]$ due to Khinshine's theorem (see for example Davidson 1994, theorem 23.5) since $\{u_t^*\}$ is IID, and the properties $E|z_t^*|^\delta = 1$ and $|E(\log|z_t^*|^\delta)| < \infty$ ensure that $E[\exp(u_t^*)]$ exists. Consider $\frac{1}{T}\sum_{t=1}^T \exp(\hat{u}_t^*) - \frac{1}{T}\sum_{t=1}^T \exp(u_t^*)$, which can be rewritten as $\frac{1}{T}\sum_{t=1}^T[\exp(\hat{u}_t^*) - \exp(u_t^*)]$. Since $\hat{u}_t^* \xrightarrow{P} u_t^*$ for each t, we have that $\exp(\hat{u}_t^*) \xrightarrow{P} \exp(u_t^*)$ for each t due to the continuity of the $\exp(\cdot)$ function. Accordingly, $\frac{1}{T}\sum_{t=1}^T \exp(\hat{u}_t^*) \to \frac{1}{T}\sum_{t=1}^T \exp(u_t^*)$ as $T \to \infty$, and since $\frac{1}{T}\sum_{t=1}^T \exp(u_t^*) \to E[\exp(u_t^*)]$ as $T \to \infty$ it follows that $\frac{1}{T}\sum_{t=1}^T \exp(\hat{u}_t^*) \xrightarrow{P} E[\exp(u_t^*)]$.

We now prove b). Due to the continuity of the $\exp(\cdot)$ operator, the assumption of consistent estimation of the ARMA representation ensures that the fitted values $\{\hat{\sigma}_t^{*\delta}\}$ are consistent estimates of their true counterparts. Next, taking the δth. square root and dividing each ϵ_t by means of $\sqrt[\delta]{\hat{\sigma}_t^{*\delta}}$ means the $\{\hat{z}_t^*\}$ are consistent estimates of their true counterparts $\{z_t^*\}$. Finally, using a similar argument to the proof of a) yields that $\frac{1}{T}\sum_{t=1}\hat{z}_t^{*2} \xrightarrow{P} 1/E(|z_t|^\delta)^{2/\delta}$, and so $-\frac{\delta}{2}\log(\frac{1}{T}\sum_{t=1}^T \hat{z}_t^{*2}) \xrightarrow{P} \log E|z_t|^\delta$.

Table 1: Finite sample precision of estimation methods

Model:															
ϕ_1	α_0	α_1	β_1	Meth.	T	$M(\hat{\phi}_1)$	$V(\hat{\phi}_1)$	$M(\hat{\alpha}_0)$	$V(\hat{\alpha}_0)$	$M(\hat{\alpha}_1)$	$V(\hat{\alpha}_1)$	$M(\hat{\beta}_1)$	$V(\hat{\beta}_1)$	$M[\hat{E}(\log z_t^2)]$	$V[\hat{E}(\log z_t^2)]$
	0			OLS	200			-0.004	0.009					-1.263	0.014
					500			-0.003	0.004					-1.270	0.006
					1000			-0.001	0.002					-1.271	0.003
	0	0.10		OLS	200			-0.013	0.020	0.093	0.005			-1.268	0.015
					500			-0.008	0.008	0.096	0.002			-1.269	0.006
					1000			-0.005	0.004	0.098	0.001			-1.272	0.003
	0	0.10	0.8	2OLS	200			-0.396	0.825	0.100	0.006	0.478	0.504	-1.297	0.018
					500			-0.131	0.114	0.108	0.002	0.673	0.084	-1.297	0.007
					1000			-0.077	0.030	0.109	0.001	0.715	0.025	-1.293	0.003
				NLS	200			-0.432	0.454	0.104	0.004	0.443	0.244	-1.294	0.047
					500			-0.103	0.059	0.105	0.001	0.705	0.041	-1.276	0.016
					1000			-0.030	0.009	0.102	0.001	0.771	0.008	-1.275	0.005
	0	0.05	0.9	2OLS	200			-0.822	1.864	0.049	0.005	0.255	1.009	-1.293	0.019
					500			-0.396	0.549	0.061	0.002	0.559	0.382	-1.295	0.007
					1000			-0.216	0.213	0.057	0.001	0.711	0.146	-1.297	0.003
				NLS	200			-0.794	0.692	0.048	0.007	0.272	0.356	-1.292	0.090
					500			-0.405	0.399	0.057	0.001	0.557	0.247	-1.283	0.007
					1000			-0.128	0.127	0.057	0.001	0.783	0.076	-1.288	0.016
0.1	0	0.10		OLS	200	0.097	0.006	-0.023	0.022	0.082	0.005			-1.262	0.014
					500	0.101	0.002	-0.013	0.008	0.092	0.002			-1.274	0.006
					1000	0.102	0.001	-0.004	0.004	0.097	0.001			-1.271	0.003

In all DGPs $z_t \overset{IID}{\sim} N(0,1)$. $M(\cdot)$ and $V(\cdot)$ are the sample mean and sample variance of the estimates, respectively. The 2OLS method consists of estimating an (assumed invertible) ARMA representation in two steps. First, use OLS to estimate the residuals by means of a (possibly) long AR-regression that approximates the ARMA representation. Second, estimate by means of OLS the ARMA representation using the estimated residuals from the first step as MA regressors. NLS is the non-linear least squares estimation procedure. Simulations in R with 1000 replications, and a prior burn-in sample of 100 observations were discarded at each replication in order to avoid initial value issues.

Table 2: Finite sample precision of estimation methods: OLS *vs.* ML for a log-ARCH(1)

Method	$f(z_t)$	T	$M(\hat{\alpha}_0)$	$V(\hat{\alpha}_0)$	$M(\hat{\alpha}_1)$	$V(\hat{\alpha}_1)$
OLS	$N(0,1)$	200	-0.013	0.020	0.093	0.005
		500	-0.008	0.008	0.096	0.002
		1000	-0.005	0.004	0.098	0.001
	$GED(1.1)$	200	-0.025	0.039	0.094	0.005
		500	-0.012	0.017	0.097	0.002
		1000	-0.001	0.008	0.099	0.001
	$st(4.1)$	200	-0.040	0.065	0.093	0.005
		500	-0.019	0.024	0.099	0.002
		1000	-0.015	0.014	0.097	0.001
ML	$N(0,1)$	200	0.013	0.014	0.107	0.003
		500	0.000	0.005	0.104	0.001
		1000	0.001	0.003	0.101	0.000
	$GED(1.1)$	200	0.038	0.039	0.114	0.004
		500	0.035	0.015	0.112	0.002
		1000	0.027	0.009	0.109	0.001
	$st(4.1)$	200	0.023	0.078	0.116	0.005
		500	0.035	0.039	0.115	0.002
		1000	0.027	0.026	0.111	0.001

The DGP of the simulation is $r_t = \sigma_t z_t, z_t \overset{IID}{\sim} f(z_t), \log \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2$ with $\alpha_0 = 0$ and $\alpha_1 = 0.1$. $N(0,1)$ is short for standard normal, $GED(1.1)$ is short for Generalised Error Distribution with shape parameter 1.1 and $st(4.1)$ is short for standardised t-distribution with 4.1 degrees of freedom. ML estimation consists of Gaussian maximum likelihood estimation with initial parameter values provided by OLS. ML estimation is implemented as a Newton-Raphson algorithm with analytical gradient and Hessian, unit step-size and 0.0001 as convergence criterion in the log-likelihood. $M(\cdot)$ and $V(\cdot)$ are the sample mean and variances of the estimates, respectively. Simulations in R with 1000 replications, and a prior burn-in sample of 100 observations were discarded at each replication in order to avoid initial value issues.

Table 3: Finite sample size in the logarithmic variance specification, using a nominal level of 5%

H_0	H_1	Fitted specification	T	$\tau = 1.1$	$\tau = 2$	$\tau = 3$
$\alpha_1 = 0$	$\alpha_1 \neq 0$	$\alpha_0 + \alpha_1 \log \epsilon_{t-1}^2$	10	0.054	0.049	0.047
			100	0.047	0.046	0.044
			1000	0.052	0.049	0.051
			10000	0.048	0.049	0.048
$\lambda = 0$	$\lambda \neq 0$	$\alpha_0 + \lambda(\log \epsilon_{t-1}^2) I_{\{z_{t-1}<0\}}$	15	0.054	0.052	0.055
			100	0.048	0.047	0.043
			1000	0.050	0.051	0.044
			10000	0.050	0.051	0.051
$\omega = 0$	$\omega \neq 0$	$\alpha_0 + \omega y_t^{N(0,1)}$	10	0.089	0.081	0.083
			100	0.056	0.052	0.051
			1000	0.054	0.049	0.050
			10000	0.050	0.051	0.048
$\omega = 0$	$\omega \neq 0$	$\alpha_0 + \omega y_t^{EXP(1)}$	10	0.085	0.079	0.100
			100	0.051	0.047	0.040
			1000	0.050	0.046	0.057
			10000	0.044	0.050	0.046
$\alpha_0 = 0$	$\alpha_0 \neq 0$	α_0	10	0.070	0.044	0.027
			100	0.027	0.004	0.001
			1000	0.015	0.001	0.000
			10000	0.020	0.001	0.002

The simulation DGP is $r_t = \epsilon_t$, $\epsilon_t = \sigma_t z_t$, $z_t \overset{IID}{\sim} GED(\tau)$, $\log \sigma_t^2 = 0$, for $t = 1, \ldots, T$. Tests are two-sided and $\{y_t^{(\cdot)}\}$ is IID and mutually independent with $\{z_t\}$. "N(0,1)" denotes standard normal and "EXP(1)" denotes an exponentially distributed variable with shape parameter equal to 1. Simulations in R with 10 000 replications.

Table 4: List of multi-path GETS experiments

	k_0	k_1	Simulation DGP	GUM
HP0	0	40	$r_t = \epsilon_t, \quad \epsilon_t = 130 z_t, \quad z_t \overset{IID}{\sim} GED(\tau)$	$r_t = \sum_{k=1}^{41} \psi_k x_{kt}^{HP} + \epsilon_t$, where $x_{37t}^{HP} = r_{t-1}$, $x_{38t}^{HP} = r_{t-2}$, $x_{39t}^{HP} = r_{t-3}$, $x_{40t}^{HP} = r_{t-4}$ and $x_{41t}^{HP} = 1$
HP0*	0	40	$r_t = \epsilon_t, \quad \epsilon_t = \sigma z_t, \quad z_t \overset{IID}{\sim} GED(2), \quad E(\epsilon_t^2) \approx 130^2$ $\log \sigma_t^2 = 1.0885 + 0.1 \log \epsilon_{t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in HP0
HP2'	1	39	$r_t = 0.75 r_{t-1} + \epsilon_t, \quad \epsilon_t = 85.99 z_t, \quad z_t \overset{IID}{\sim} GED(\tau)$	Same as in HP0
HP2'*	1	39	$r_t = 0.75 r_{t-1} + \epsilon_t, \quad \epsilon_t = \sigma z_t, \quad z_t \overset{IID}{\sim} GED(2),$ $E(\epsilon_t^2) \approx (85.99)^2, \quad \log \sigma_t^2 = 1.0058 + 0.1 \log \epsilon_{t-1}^2 +$ $0.8 \log \sigma_{t-1}^2$	Same as in HP0
HP7'	3	37	$r_t = 0.75 r_{t-1} + 1.33 x_{11t}^{HP} + 0.9975 x_{29t}^{HP} + \epsilon_t,$ $\epsilon_t = 85.99 z_t, \quad z_t \overset{IID}{\sim} GED(\tau)$	Same as in HP0
SE1	0	9	$r_t = \epsilon_{1t}, \quad \epsilon_{1t} = \sigma_t z_t, \quad z_t \overset{IID}{\sim} GED(2)$ $\log \sigma_t^2 = 0.1 \log \epsilon_{1t-1}^2 + 0.8 \log \sigma_{t-1}^2$	$r_t = \phi_0 + \sum_{m=1}^{2} \phi_1 r_{t-m} + \sum_{n=0}^{2} \eta_n x_{t-n} + \eta_3 x_{1t}^{N(0,1)} +$ $\eta_4 x_{2t}^{N(0,1)} + \eta_5 x_{1t}^{EXP(1)} + \eta_6 x_{2t}^{EXP(1)} + \epsilon_{1t}$, where $x_t = 0.1 +$ $0.9 x_{t-1} + \epsilon_{2t}$ with $\epsilon_{2t} \overset{IID}{\sim} N(0,1)$
SE2	1	8	$r_t = 0.1 r_{t-1} + \epsilon_{1t}, \quad \epsilon_{1t} = \sigma_t z_t, \quad z_t \overset{IID}{\sim} GED(2)$ $\log \sigma_t^2 = 0.1 \log \epsilon_{1t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in SE1
SE3	0	12	$r_t = \epsilon_{1t}, \quad \epsilon_{1t} = \sigma_t z_t, \quad \sigma_t = 1, \quad z_t \overset{IID}{\sim} GED(\tau)$	$\log \sigma_t^2 = \alpha_0 + \sum_{p=1}^{P} \alpha_p \log \epsilon_{1t-p}^2 + \lambda_1 (\log \epsilon_{1t-1}^2) I_{z_{t-1}<0} +$ $\omega_1 y_t + \omega_2 y_{t-1} + \omega_3 x_{1t}^{EXP(1)} + \omega_4 x_{2t}^{EXP(1)} + \omega_5 x_{3t}^{N(0,1)} +$ $\omega_6 x_{4t}^{N(0,1)}$, where $y_t = 0.9 y_{t-1} + \epsilon_{2t}$ with $\epsilon_{2t} \overset{IID}{\sim} N(0,1)$
SE4	1	11	$r_t = \epsilon_{1t}, \quad \epsilon_{1t} = \sigma_t z_t, \quad z_t \overset{IID}{\sim} GED(\tau)$ $\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \epsilon_{1t-1}^2$	Same as in SE3
SE5	0	2	$r_t = \epsilon_{1t}, \quad \epsilon_{1t} = \sigma_t z_t, \quad \sigma_t = 1, \quad z_t \overset{IID}{\sim} GED(\tau)$	$\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \epsilon_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2$
SE6	2	0	$r_t = \epsilon_{1t}, \quad \epsilon_{1t} = \sigma_t z_t, \quad z_t \overset{IID}{\sim} GED(\tau)$ $\log \sigma_t^2 = 0.1 \log \epsilon_{1t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in SE5

Note: The design of the experiments HP0, HP2' and HP7' are based on Hoover and Perez (1999), and make use of their data $x_{1t}, \ldots, x_{36t}$ which are available via `http://www.csus.edu/indiv/p/perezs/Data/data.htm`. The number of relevant variables in the GUM is k_0, the number of irrelevant variables in the GUM is k_1 and the total number of variables (the constant included) in the GUM is $k = k_0 + k_1 + 1$.

Table 5: Comparison of multi-path GETS algorithms: Specification search in the mean with Gaussian homoscedastic errors

	k_0	k_1	Algorithm	T	$M(\hat{k}_0/k_0)$	$M(\hat{k}_1/k_1)$	$\hat{p}(DGP)$
HP0	0	40	AutoSEARCH	139		0.004	0.942
			AutoSEARCH w/PET			0.038	0.825
			AutoSEARCH w/PET, no empty			0.068	0.280
			HP			0.045	0.292
			PcGets			≈ 0.04	≈ 0.45
HP2'	1	39	AutoSEARCH	139	1.000	0.043	0.285
			HP		1.000	0.107	0.000
			PcGets		≈ 0.97	≈ 0.05	≈ 0.32
			Autometrics		1.000	0.063	0.119
HP7'	3	37	AutoSEARCH	138	0.999	0.044	0.309
			HP		0.967	0.082	0.040
			PcGets		≈ 1.00	≈ 0.04	≈ 0.37
			Autometrics		0.999	0.021	0.111

Simulations of the AutoSEARCH algorithm are in R with 1000 replications using the data and DGPs of Hoover and Perez (1999). $M(\hat{k}_0/k_0)$ is the average proportion of relevant variables $\hat{k}_0$ retained relative to the number of relevant variables k_0 in the DGP. $M(\hat{k}_1/k_1)$ is the average proportion of irrelevant variables $\hat{k}_1$ retained relative to the number of irrelevant variables k_1 in the GUM. $\hat{p}(DGP)$ is the proportion of times the DGP is found exactly. The properties of the HP algorithm are from Hoover and Perez (1999, table 3 on p. 179). The properties of the PcGets algorithm are from Hendry and Krolzig (2005, figure 1 on p. C39), and the properties of the Autometrics algorithm are from Doornik (2009). A constant is included in all the regressions but ignored in the evaluation of successes and failures. This is in line with Hoover and Perez (1999) but counter to Hendry and Krolzig (2005), and Doornik (2009). PET is short for Parsimonious Encompassing Test.

Table 6: Comparison of multi-path GETS algorithms: Specification search in the mean with heteroscedastic errors

	k_0	k_1	Algorithm	T	$M(\hat{k}_0/k_0)$	$M(\hat{k}_1/k_1)$	$\hat{p}(DGP)$
HP0*	0	40	AutoSEARCH	139		0.010	0.893
HP2'*	1	39	AutoSEARCH	139	1.000	0.076	0.147
SE1	0	9	AutoSEARCH	200		0.014	0.922
				500		0.019	0.883
				1000		0.018	0.895
SE2	1	8	AutoSEARCH	200	0.251	0.019	0.172
				500	0.548	0.030	0.387
				1000	0.825	0.039	0.586

Simulations of the AutoSEARCH algorithm are in R with 1000 replications. In each replication a prior burn-in sample of 100 observations is discarded in order to avoid initial value issues.

Table 7: Properties of multi-path GETS in experiments SE3 - SE6

	k_0	k_1	Algorithm	T	τ	$M(\hat{k}_0/k_0)$	$M(\hat{k}_1/k_1)$	$p(DGP)$
SE3	0	12	AutoSEARCH	200	2.0		0.005	0.951
					1.1		0.016	0.850
				500	2.0		0.003	0.976
					1.1		0.017	0.866
				1000	2.0		0.003	0.971
					1.1		0.016	0.860
SE4	1	11	AutoSEARCH	200	2.0	0.464	0.035	0.303
					1.1	0.265	0.034	0.140
				500	2.0	0.911	0.044	0.571
					1.1	0.609	0.034	0.353
				1000	2.0	0.995	0.046	0.624
					1.1	0.902	0.045	0.537
SE5	0	2	AutoSEARCH	200	2.0		0.047	0.953
					1.1		0.064	0.936
				500	2.0		0.055	0.945
					1.1		0.057	0.943
				1000	2.0		0.045	0.955
					1.1		0.051	0.949
SE6	2	0	AutoSEARCH	200	2.0	0.471		0.471
					1.1	0.267		0.267
				500	2.0	0.824		0.824
					1.1	0.487		0.487
				1000	2.0	0.979		0.979
					1.1	0.783		0.783

1000 replications in R, each replication with a prior burn-in sample of 100 observations in SE3, SE4 and S6 in order to avoid initial value issues (there are no initial value issues in SE5). In SE3 and SE4 the simulations are undertaken using a nominal regressor significance level of 5%, and in SE4 the parameter values of the simulations DGP are $(\alpha_0, \alpha_1) = (0, 0.2)$. In SE5 and SE6 no regressor is tested for significance but two specifications (The empty model and the GUM) are evaluated in terms of diagnostics and information criterion, and in SE6 the parameter values of the simulation DGP are $(\alpha_0, \alpha_1, \beta_1) = (0, 0.1, 0.8)$. The nominal level used for the two diagnostic tests Ljung-Box AR(1) and Ljung-Box ARCH(1), respectively, is 2.5%. Estimation in SE3 and SE4 is by OLS, and in SE5 and SE6 by means of 2-step OLS. k_0 is the number of relevant variables (apart from the constant) in the simulation DGP, and k_1 is the number of irrelevant variables in the GUM.

Table 8: Rejection probabilities of normality ($\tau = 2$) for various values of τ using a nominal level of 5%

Jarque-Bera:

T	$\tau = 1.1$	$\tau = 1.3$	$\tau = 1.5$	$\tau = 1.7$	$\tau = 2$	$\tau = 2.3$	$\tau = 2.5$	$\tau = 3$
25	0.2320	0.1300	0.0710	0.0580	0.0260	0.0170	0.0080	0.0050
100	0.6730	0.4310	0.2390	0.1390	0.0390	0.0100	0.0090	0.0020
200	0.9010	0.6720	0.4020	0.1980	0.0500	0.0090	0.0210	0.1090
500	0.9990	0.9620	0.7290	0.3610	0.0500	0.0660	0.1820	0.7330
1000	1.0000	1.0000	0.9460	0.5530	0.0490	0.1930	0.5770	0.9920
10000	1.0000	1.0000	1.0000	1.0000	0.0550	1.0000	1.0000	1.0000

Anscombe-Glynn:

T	$\tau = 1.1$	$\tau = 1.3$	$\tau = 1.5$	$\tau = 1.7$	$\tau = 2$	$\tau = 2.3$	$\tau = 2.5$	$\tau = 3$
25	0.2480	0.1330	0.0750	0.0600	0.0550	0.0490	0.0550	0.0690
100	0.6600	0.4100	0.2000	0.1130	0.0510	0.0880	0.1470	0.3400
200	0.8960	0.6630	0.3680	0.1690	0.0510	0.1380	0.2570	0.6560
500	0.9980	0.9580	0.7300	0.3370	0.0660	0.2630	0.5370	0.9580
1000	1.0000	1.0000	0.9520	0.5540	0.0520	0.4650	0.8380	0.9990
10000	1.0000	1.0000	1.0000	1.0000	0.0570	1.0000	1.0000	1.0000

Bonett-Seier:

T	$\tau = 1.1$	$\tau = 1.3$	$\tau = 1.5$	$\tau = 1.7$	$\tau = 2$	$\tau = 2.3$	$\tau = 2.5$	$\tau = 3$
25	0.2940	0.1730	0.1010	0.0630	0.0610	0.0510	0.0590	0.0750
100	0.8020	0.5350	0.2570	0.1190	0.0350	0.0760	0.1140	0.2510
200	0.9740	0.8270	0.4780	0.1970	0.0530	0.1090	0.2070	0.5060
500	1.0000	0.9950	0.8470	0.4230	0.0470	0.2350	0.4310	0.8970
1000	1.0000	1.0000	0.9880	0.6560	0.0410	0.4120	0.7840	0.9980
10000	1.0000	1.0000	1.0000	1.0000	0.0500	1.0000	1.0000	1.0000

Likelihood-Ratio (LR):

T	$\tau = 1.1$	$\tau = 1.3$	$\tau = 1.5$	$\tau = 1.7$	$\tau = 2$	$\tau = 2.3$	$\tau = 2.5$	$\tau = 3$
25	0.3340	0.1710	0.0960	0.0600	0.0390	0.0430	0.0390	0.0600
100	0.8200	0.5270	0.2490	0.1250	0.0270	0.0520	0.1110	0.2760
200	0.9750	0.8350	0.4640	0.1980	0.0520	0.1100	0.2380	0.5820
500	1.0000	0.9970	0.8510	0.4220	0.0580	0.2570	0.5020	0.9450
1000	1.0000	1.0000	0.9900	0.6660	0.0400	0.4560	0.8460	0.9990
10000	1.0000	1.0000	1.0000	1.0000	0.0520	1.0000	1.0000	1.0000
Kurtosis	≈ 5.3		≈ 3.9		$= 3.0$		≈ 2.6	≈ 2.4

Simulations in R with 1000 replications.

Table 9: Logarithmic Mincer-Zarnowitz regressions of variability (squared return) on volatility proxies

Model	$\hat{\alpha}$ $[p-val]$	$\hat{\beta}$ $[p-val]$	$\chi^2(2)$ $[p-val]$	$AR(1)$ $[p-val]$	$ARCH(5)$ $[p-val]$	JB $[p-val]$
$\log \sigma_t^2 = \alpha + \beta \log RV_t^{65m}$	-0.04 [0.85]	2.17 [0.00]	72.81 [0.00]	-0.04 [0.04]	0.06 [0.00]	6301.47 [0.00]
$\log \sigma_t^2 = \alpha + \beta \log RV_t^{15m}$	-0.58 [0.03]	1.87 [0.00]	19.22 [0.00]	-0.01 [0.45]	0.11 [0.00]	58.80 [0.00]
$\log \sigma_t^2 = \alpha + \beta \log RV_t^{5m}$	-0.64 [0.05]	1.55 [0.02]	5.69 [0.06]	-0.01 [0.76]	0.02 [0.36]	4.68 [0.10]

All calculations and tests are based on the assumptions that $r_t = \sigma_t z_t, z_t \sim IID(0,1)$ and $\log \sigma_t^2 = \alpha + \beta \log RV_t^{(\cdot)}$, where r_t is daily IBM return 4 January 1993 - 31 December 2003 (2772 observations). The data are from Patton (2008), where RV_t^{65m}, RV_t^{15} and RV_t^{5m} are realised volatilities made up of 65-minute, 15-minute and 5-minute intra-day returns. The p-values in the $\hat{\alpha}$ and $\hat{\beta}$ columns are from Wald coefficient restriction tests of $\alpha = 0$ and $\beta = 1$, respectively, whereas the p-values in the $\chi^2(2)$ column are from the joint test. The ordinary variance-covariance matrix is used for the Wald tests. $AR(1)$ is a Ljung and Box (1979) test of 1st. order serial correlation in the standardised residuals $\{z_t\}$, $ARCH(5)$ is a Ljung and Box (1979) test of 5th. order serial correlation in the squared standardised residuals $\{z_t^2\}$, and JB is the Jarque and Bera (1980) test for normality.

Table 10: Parsimonious encompassing test of RV_t^{5m}

MGUM: $\underset{[p.val.]}{\hat{r}_t} = \underset{[0.25]}{0.048} - \underset{[0.12]}{0.036} r_{t-1} - \underset{[0.22]}{0.028} r_{t-2} + \underset{[0.08]}{0.004} RV_{t-1}^{5m} + \underset{[0.54]}{0.002} RV_{t-2}^{5m}$

VGUM 1: $\underset{[p.val.]}{\log \hat{\sigma}_t^2} = \underset{[0.01]}{-0.192} + \underset{[0.00]}{1.08} \log RV_t^{5m}$

$AR(3) : \underset{[p.val.]}{-0.02}$ $\underset{[0.40]}{}$ $ARCH(5) : \underset{[p.val.]}{0.05}$ $\underset{[0.03]}{}$

VGUM 2: $\underset{[p.val.]}{\log \hat{\sigma}_t^2} = \underset{[0.01]}{-0.191} - \underset{[0.31]}{0.027} \log \hat{\epsilon}_{t-1}^2 - \underset{[0.95]}{0.001} \log \hat{\epsilon}_{t-2}^2 - \underset{[0.41]}{0.015} \log \hat{\epsilon}_{t-3}^2 + \underset{[0.09]}{0.031} \log \hat{\epsilon}_{t-4}^2$

$+\underset{[0.49]}{0.024} (\log \hat{\epsilon}_{t-1}^2) \hat{I}_{\{\epsilon_{t-1}<0\}} + \underset{[0.00]}{1.077} \log RV_t^{5m}$

$AR(3) : \underset{[p.val.]}{-0.02}$ $\underset{[0.40]}{}$ $ARCH(5) : \underset{[p.val.]}{0.05}$ $\underset{[0.20]}{}$

MSPEC: $\underset{[p.val.]}{\hat{r}_t} = \underset{[0.07]}{0.075}$

VSPEC: $\underset{[p.val.]}{\log \hat{\sigma}_t^2} = \underset{[0.02]}{-0.170} + \underset{[0.05]}{0.035} \log \hat{\epsilon}_{t-4}^2 + \underset{[0.05]}{1.065} \log RV_t^{5m}$

$AR(1) : \underset{[p.val.]}{-0.01}$ $\underset{[0.61]}{}$ $ARCH(5) : \underset{[p.val.]}{0.05}$ $\underset{[0.04]}{}$

r_t is daily IBM return 4 January 1993 - 31 December 2003 (2772 observations). The data are from Patton (2008), where RV_t^{5m} is realised volatility made up of 5-minute intra-day returns. MGUM is short for mean GUM, VGUM is short for variance GUM, MSPEC is short for specific mean specification and VSPEC is short for specific variance specification. In all the mean specifications White (1980) standard errors are used in the computations of the p-values, whereas in all the variance specifications the ordinary variance-covariance matrix is used. $AR(1)$ is a Ljung and Box (1979) test of 1st. order serial correlation in the standardised residuals $\{z_t\}$ and $ARCH(5)$ is a Ljung and Box (1979) test of 5th. order serial correlation in the squared standardised residuals $\{z_t^2\}$.

Table 11: Estimation results of VaR models

Model	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}$	$\hat{\omega}_1$	AR_1 [$p.val.$]	$ARCH_1$ [$p.val.$]	$ARCH_2$ [$p.val.$]	$ARCH_5$ [$p.val.$]	$\hat{\tau}$ [$p.val.$]
Constant	1.27				-0.040	0.176	0.247	0.207	1.040
					[0.15]	[0.00]	0.000	0.000	[0.00]
EWMA	0.00			1.00	-0.046	-0.028	0.025	0.028	1.484
					[0.10]	[0.31]	[0.40]	[0.23]	[0.00]
RiskMetrics		0.06	0.94		-0.045	-0.028	0.023	0.013	1.492
					[0.10]	[0.31]	[0.42]	[0.57]	[0.00]
GARCH	0.01	0.07	0.92		-0.049	-0.069	0.019	0.012	1.891
					[0.07]	[0.01]	[0.04]	[0.13]	[1.00]
log-GARCH	0.07	0.05	0.93		-0.021	-0.003	0.065	0.091	1.336
					[0.46]	[0.90]	[0.06]	[0.00]	[0.00]
SEARCH	0.09			0.89	-0.045	-0.029	0.033	0.036	1.500
					[0.11]	[0.29]	[0.29]	[0.08]	[0.00]

The models are of daily SP500 return (in %) variability 1 January 2001 - 30 December 2005 (1305 observations). The AR_1 column contains the first order autocorrelation of the standardised residuals together with the p-value of the associated Ljung and Box (1979) test, the $ARCH_{(\cdot)}$ columns contain the first, second and fifth order autocorrelations of the squared standardised residuals together with the p-values of the associated Ljung and Box (1979) tests, and the $\hat{\tau}$ column contains the estimate of the GED-shape parameter together with the p-value of an LR-test for normality. Estimation and computations in R and EViews.

Table 12: Out-of-sample VaR forecast evaluation results of SP500 returns

Model	1%		5%		10%	
	$\hat{p}_1$ [$p.val.$]	$\hat{\pi}_{11}$ [$p.val.$]	$\hat{p}_1$ [$p.val.$]	$\hat{\pi}_{11}$ [$p.val.$]	$\hat{p}_1$ [$p.val.$]	$\hat{\pi}_{11}$ [$p.val.$]
Constant-$N(0,1)$	0.050	0.013	0.082	0.025	0.110	0.032
	[0.00]	[1.00]	[0.00]	[1.00]	[0.36]	[1.00]
EWMA-$N(0,1)$	0.030	0.000	0.071	0.003	0.110	0.015
	[0.00]	[1.00]	[0.01]	[1.00]	[0.36]	[1.00]
RiskMetrics-$N(0,1)$	0.034	0.000	0.068	0.006	0.110	0.015
	[0.00]	[1.00]	[0.02]	[1.00]	[0.36]	[1.00]
GARCH(1,1)-$N(0,1)$	0.000	0.000	0.001	0.000	0.001	0.000
	[0.00]	[1.00]	[0.00]	[1.00]	[0.00]	[1.00]
log-GARCH(1,1)-$N(0,1)$	0.000	0.000	0.001	0.000	0.004	0.000
	[0.00]	[1.00]	[0.00]	[1.00]	[0.00]	[1.00]
SEARCH-$N(0,1)$	0.030	0.000	0.065	0.003	0.108	0.012
	[0.00]	[1.00]	[0.07]	[1.00]	[0.43]	[1.00]
Constant-$GED(\hat{\tau})$	0.037	0.013	0.084	0.025	0.110	0.032
	[0.00]	[1.00]	[0.00]	[1.00]	[0.36]	[1.00]
EWMA-$GED(\hat{\tau})$	0.024	0.000	0.068	0.003	0.118	0.015
	[0.00]	[1.00]	[0.02]	[1.00]	[0.09]	[1.00]
RiskMetrics-$GED(\hat{\tau})$	0.026	0.000	0.068	0.006	0.118	0.018
	[0.00]	[1.00]	[0.02]	[1.00]	[0.09]	[1.00]
GARCH(1,1)-$GED(\hat{\tau})$	0.000	0.000	0.001	0.000	0.001	0.000
	[0.00]	[1.00]	[0.00]	[1.00]	[0.00]	[1.00]
log-GARCH(1,1)-$GED(\hat{\tau})$	0.000	0.000	0.001	0.000	0.004	0.000
	[0.00]	[1.00]	[0.00]	[1.00]	[0.00]	[1.00]
SEARCH-$GED(\hat{\tau})$	0.030	0.000	0.065	0.003	0.108	0.012
	[0.00]	[1.00]	[0.07]	[1.00]	[0.43]	[1.00]

The VaR forecasts are of daily SP500 returns (in %) 2 January 2006 - 24 February 2009 (821 observations).Estimates and tests are based on the methodology proposed by Christoffersen (1998). $\hat{p}_1$ is the estimated unconditional probability of returns being lower than the quantile in question (unconditional coverage), 1%, 5% or 10%, and $\hat{p}_{11}$ is the estimated probability of a sequential pair of returns both being lower than the quantile in question. The p-values are $\chi^2(1)$ distributed tests of unconditional coverage and first-order Markov independence, respectively. All computations in R.

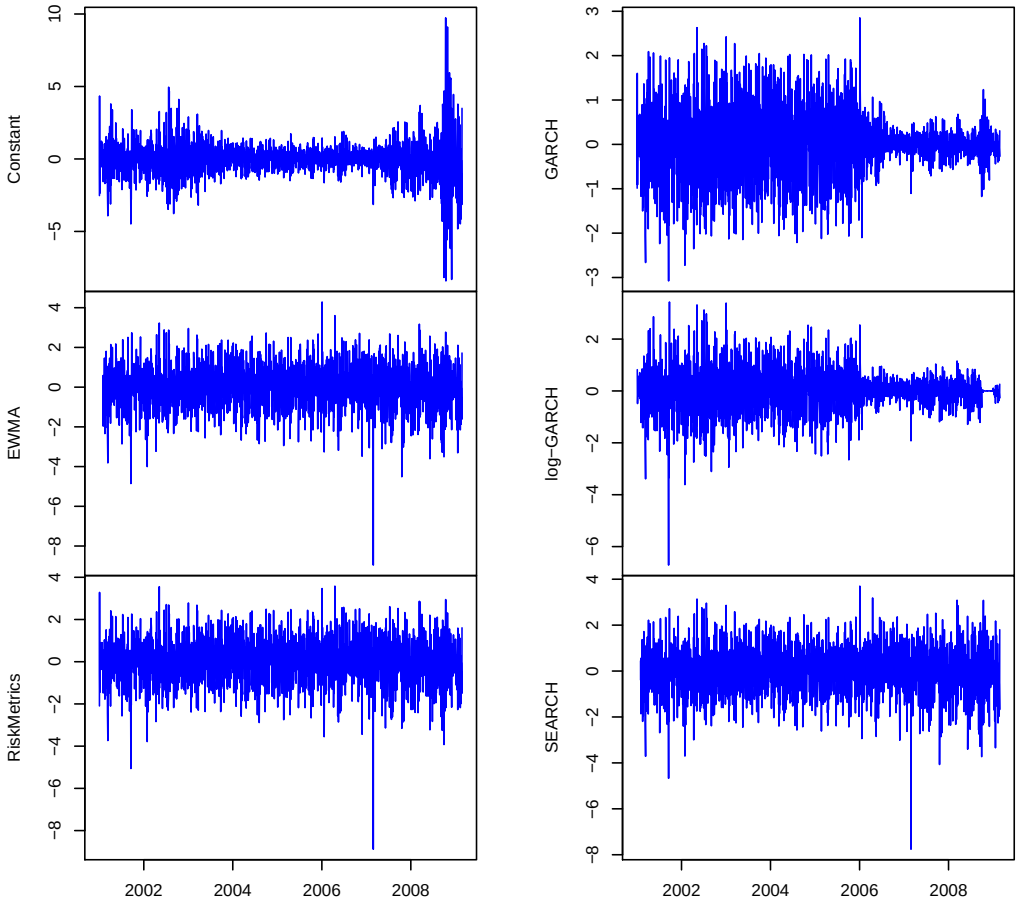

Figure 1: Standardised residuals of forecast models over the estimation and design sample 1 January 2001 - 30 December 2005, and over the forecast sample 2 January 2006 - 24 February 2009