

Automated Model Selection in Finance: General-to-Specific Modelling of the Mean, Variance and Density*

Genaro Sucarrat[†] and Álvaro Escribano[‡]

1 February 2010

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. Making use of a recent result on log-GARCH Models, we provide methods 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 and log-GARCH terms, an asymmetry term, 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 three 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, specification search, model selection, finance, volatility, value-at-risk

*We are grateful to Luc Bauwens, André Alves Portela Santos, Enrique Sentana and seminar participants at BI (Norwegian School of Management) for helpful comments and suggestions. An earlier version of this paper circulated under the title “Automated Financial Multi-Path GETS Modelling” (Sucarrat and Escribano 2009). 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

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 increases the resources and efforts needed for reliable estimation and model validation multiply. Indeed, this may even become an obstacle to financial multi-path General-to-Specific (GETS) modelling, as for example argued by Granger and Timmermann (1999), and McAleer (2005) regarding automated GETS modelling of financial volatility.

A recent result in Sucarrat and Escribano (2010) means many of the estimation and inference complexity issues typically associated with financial models can readily be overcome. Specifically, Sucarrat and Escribano (2010) prove a result that enables consistent least squares estimation and inference of power log-GARCH models¹ for fixed power, and under very general assumptions on the standardised error. Moreover, their simulations suggest OLS estimation of power log-ARCH models compares favourably with (Gaussian) Quasi Maximum Likelihood (QML) estimation when the standardised error is non-Gaussian.² Since the exponential specification ensures positive variance and imposes less restrictions on the parameter space than standard ARCH models, this effectively means estimation and inference is greatly simplified. 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, an asymmetry term, 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

¹“GARCH” is short for generalised autoregressive conditional heteroscedasticity, and the origins of the acronym are Engle (1982) and Bollerslev (1986).

²Currently there does not exist a proof of the normal QML estimator being consistent for a general class of exponential variance models. To our knowledge, the proofs that exist are for very limited classes of models, see Straumann and Mikosch (2006), Linton (2008), and Dahl and Iglesias (2008).

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

It is well known that ARCH models may be consistently estimated by means of OLS, and that robust inference strategies hold asymptotically. However, for several reasons, the least squares estimation and inference procedures associated with the power log-ARCH model have much better properties. First, the error term in the ARCH regression is heteroscedastic. By contrast, the error term in power log-GARCH regressions is IID. Second, the distribution of the error term in the ARCH regression has an exponential-like shape, and takes on values in $[-1, \infty)$. By contrast, in the power log-ARCH regression it is almost symmetric, with the left-tail usually being “longer”, and takes on values in $(-\infty, \infty)$. This means estimators and test-statistics are likely to correspond much closer to their asymptotic approximations in finite samples, since the convergence to their asymptotic counterparts will be much faster. Also, coefficient tests will exhibit greater power under the alternative, since the error is “smaller” due to the log-transformation. Finally, power log-GARCH regressions impose much weaker restrictions on the parameter space due to the exponential variance specification. In ARCH regressions, by contrast, strong parameter restrictions might be needed in order to ensure positive variance.

GETS modelling consists of first specifying a general model, and then simplifying it while checking the statistical validity of the simplification process. *Single-path* GETS for example, the simplest implementation of GETS modelling, is thus similar to stepwise step-down (backwards) regression. The main attraction of GETS modelling is that it takes the influence of all the potential variables into account already at the outset, which ideally should result in better estimation and inference, since each variable’s impact is controlled for the impact of the others.

GETS modelling has witnessed major advances over the last decade thanks to the development of *multi-path* GETS specification search software, see amongst others Hoover and Perez (1999), Hendry and Krolzig (2001, 2005), Krolzig (2003) and Doornik (2009).³ An important result in this literature is that multi-path GETS can be vastly superior to single-path GETS and many other specification search algorithms. Key to the success is that estimation is essentially by means of ordinary least squares (OLS) procedures, something which renders automation of multi-path GETS specification search feasible in practice. Otherwise, if the multi-path GETS search were to be undertaken manually, practitioners and academics would often find it too resource demanding to implement it in practice.

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

³GETS 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.

three 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.

The rest of the paper is organised into four sections and one appendix. The next section, section 2, outlines the SEARCH model. Section 3 studies the properties of our multi-path 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 supporting information.

2 The SEARCH model

Several of the methods we employ are applicable not only to the SEARCH model, but also to more general structures. To accommodate this, and in order to establish a general and compact notation that will be useful throughout, we start by outlining a “generalised SEARCH” (GSEARCH) model, which nests the SEARCH model. The GSEARCH model is given by

$$r_t = \mu(\psi, \mathbf{v}_t) + \epsilon_t, \quad (1)$$

$$E(r_t | \mathcal{I}_t) = \mu(\psi, \mathbf{v}_t), \quad (2)$$

$$\epsilon_t = \sigma_t z_t, \quad z_t | \mathcal{I}_t \sim IID(0, 1), \quad \sigma_t > 0, \quad (3)$$

$$\log \sigma_t^2 = g(\gamma, \mathbf{w}_t). \quad (4)$$

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 (2) 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 limits⁴ 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 logarithmic variance specification. Specifically, the SEARCH model is given by

$$r_t = \phi_0 + \sum_{m=1}^M \phi_m r_{t-m} + \sum_{n=1}^N \eta_n x_{nt} + \epsilon_t \quad (5)$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \stackrel{IID}{\sim} GED(\tau), \quad \tau > 1 \quad (6)$$

$$\begin{aligned} \log \sigma_t^\delta = & \alpha_0 + \sum_{p=1}^P \alpha_p \log |\epsilon_{t-p}|^\delta + \sum_{q=1}^Q \beta_q \log \sigma_{t-q}^\delta + \lambda (\log |\epsilon_{t-1}|^\delta) I_{\{z_{t-1} < 0\}} \\ & + (\log \kappa^\delta) J_t + \omega_0 \log EWMA_{t-1} + \sum_{d=1}^D \omega_d y_{dt}, \quad P \geq Q, \quad \delta > 0 \end{aligned} \quad (7)$$

$$J_t \in \{0, 1\}, \quad J_t \sim IID, \quad J_t \perp z_t, \quad \kappa \geq 1 \quad (8)$$

In the mean specification (5) 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, in general the estimation and inference methods we employ for the log-variance specification will be applicable subject to general conditions if MA terms are included in the mean. One thing that our methods do not admit, though, is GARCH-in-mean terms.⁵ 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

⁴One should maybe add the qualifier “current”, because future developments are likely to broaden the class of models that are amenable to automated GETS modelling.

⁵This is not necessarily a serious drawback, since proxies for financial price 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.

normal. In particular, when $\tau \rightarrow 1$ then the GED tends to a double exponential distribution, and when $\tau \rightarrow \infty$ 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 exhibits finite moments, see Nelson (1991). By contrast, a commonly used alternative distribution for which this is not necessarily the case is the t distribution.⁶

In the logarithmic variance specification (7), P is the number of log-ARCH terms, Q is the number of log-GARCH terms and δ is the power. Throughout this paper we will set δ to 2. However, our methods and algorithms can also be applied when δ differs from 2, this is simply a programming issue. λ is the impact of the logarithmic asymmetry term analogous to that of Glosten et al. (1993),⁷ $\kappa \geq 1$ is the jump-size ($\kappa = 1$ means no jump) of the Bernoulli jump process $\{J_t\}$, $EWMMA_{t-1}$ is an equally weighted moving average of the past squared residuals $\{\epsilon_{t-1}^2, \epsilon_{t-2}^2, \dots\}$ 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.⁸ It should be noted that the term $\log EWMMA_{t-1}$ can be viewed as a local approximation to $\log \sigma_{t-1}^\delta$, that is, a volatility proxy. The attractive properties of $\log EWMMA_{t-1}$ compared with $\log \sigma_t^\delta$, though, is that it is simpler to estimate, and that 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 $\delta = 2$: $E(\epsilon_t | \mathcal{I}_t^{sv}) = E(\sigma_t | \mathcal{I}_t^{sv}) E(z_t | \mathcal{I}_t^{sv}) = 0$,

⁶On the other hand, an advantage with the t distribution relative to the GED distribution is that the t distribution allows for fatter tails.

⁷One 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.

⁸The 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 $\exp(\alpha_0/\delta)$.

$Var(r_t|\mathcal{I}_t^{sv}) = Var(\epsilon_t|\mathcal{I}_t^{sv}) = E(\sigma_t^2|\mathcal{I}_t^{sv})E(z_t^2|\mathcal{I}_t^{sv}) = E(\sigma_t^2|\mathcal{I}_t^{sv})$, 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 $\lambda = \kappa = \omega_1 = \dots = \omega_D = 0$, then modulus greater than one for all the roots of the polynomial equation $[1 - (\alpha_1 + \beta_1)c - \dots - (\alpha_P + \beta_P)c^P] = 0$ is a sufficient condition for stability in the variance of r_t .

Under standard assumptions (stability, etc.) the parameters of both the mean and variance specifications can be estimated consistently by means of least squares. In particular, with no GARCH terms then the log-variance specification can be estimated by means of a modified OLS procedure, see Sucarrat and Escibano (2010) for the details. Furthermore, if the mean is zero or if it is estimated with sufficiently high precision in an appropriate sense, then ordinary OLS inference in the log-variance is asymptotically valid for all the parameters apart from the constant α_0 . For further details see Sucarrat and Escibano (2010).

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 the power $\delta = 2$. We underline that the algorithm might 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 of the overall parsimonious specification and testing for 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 $\hat{k}_0 = k_0$ and $\hat{k}_1 = 0$. 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.⁹ Specifically, in a GETS context the FWE is $1-p(DGP)$, and consistent model selection takes place when $p(DGP)$ 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”, and 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 A comparison of 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

⁹The methods of White (2000), Hansen (2005), and Romano and Wolf (2005) are examples of approaches that seek to control the FWE.

our algorithm as AutoSEARCH.¹⁰ 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:

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

$$r_t = \mathbf{v}_t\psi + \epsilon_t, \quad \epsilon_t = \sigma z_t, \quad z_t \sim IID(0, 1), \quad (9)$$

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.¹¹ 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.

¹⁰We 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.

¹¹For 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.

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 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 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 should tend to a value somewhere between 0 and 0.05. The exact value will depend on the nature of the actual DGP. Similarly, $p(DGP)$ should

tend to 1 - *the chosen overall diagnostic level*. For example, for an overall diagnostic level equal to 5% (as in the simulations), then $p(DGP)$ should tend to $1 - 0.05 = 0.95$.

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. In table 1 the experiments are labelled HP0, HP2' and HP7', and the results are contained in table 2. The first important feature that emerges from the results is that AutoSEARCH compares well in deleting irrelevant variables. In experiment HP0, where none of the regressors matter, AutoSEARCH recovers the DGP about 49% of time, which is as good or better than the other algorithms. 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 actually quite good. The simulation results also suggest that there is a clear gain in including the empty model (if it passes the diagnostic tests) in the list of considered models, since $p(DGP)$ then doubles from 22% to 46%. The effect of the parsimonious encompassing test, however, is minor. A second important feature of the simulation results is that the average irrelevance proportion $M(\hat{k}_1/k_1)$, a measure of "overfitting", in general is as low or lower for AutoSEARCH as for the other algorithms 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 mean 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 mean errors are heteroscedastic.¹² 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.¹³ Second, we turn off ARCH diagnostic checking and designate all the diagnostic checking significance level (5%) to the test for serial correlation. Third

¹²To our knowledge, no one has studied the properties of multi-path GETS specification search of the mean when the errors are heteroscedastic.

¹³There are numerous heteroscedasticity consistent standard errors to choose from. We choose White (1980) simply because it is the most common in econometric software.

and finally, we use the Gaussian likelihood of the mean residuals when computing the information criterion.

Table 3 contains the results of four Monte Carlo experiments, all with a reasonably 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.¹⁴ The results of HP0* are very similar to those of HP0. This suggests that 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 slightly and $p(DGP)$ about halves. 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 AutoSEARCH recovers the DGP with a probability equal to about 0.8 (slowly increasing in T), and the irrelevance proportion remains below the chosen regressor significance level of 5%. In other words, multi-path GETS works better for financial data than the HP0* experiment suggests. This is also the case for SE2, since the irrelevance proportion also here remains below the chosen regressor significance level, and since $p(DGP)$ is higher than in HP1*. One caveat is that the HP experiments

¹⁴In HP0* and HP2'* the parameter α_0 of the log-GARCH specification has been calibrated. 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

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

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 $\alpha_0 \approx 1.0058$.

consist of sample sizes equal to about 140 observations, whereas SE1 and SE2 take place for $T \in \{200, 500, 1000\}$. So the difference could simply be due to the difference in sample size. Nevertheless, the conclusions of SE1 and SE2 that multi-path GETS works better than suggested by the HP experiments remain, because the SE1 and SE2 experiments better reflect financial modelling situations.

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 4 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 $\alpha_1 = 0.2$. 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 volatility proxies (say, the level of trading volume or a high-frequency based measure), 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 $r_t \stackrel{IID}{\sim} GED(\tau)$ 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 $z_t \stackrel{IID}{\sim} GED(\tau)$.

The first main property of table 4 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 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 17% in SE4 when $T = 200$ and $\tau = 1.1$, and the maximum is 0.97 in SE6 when $T = 1000$ and $\tau = 2$. 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 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

$$VI = \frac{\sqrt{m_2}}{m_1} = \frac{\sqrt{\Gamma(1/\tau)\Gamma(3/\tau)}}{\Gamma(2/\tau)} \quad (10)$$

is an index of kurtosis, where m_r is the absolute moment of grade r . The sample counterpart of (10) is

$$\widehat{VI} = \frac{\sqrt{T \sum_{t=1}^T z_t^2}}{\sum_{t=1}^T |z_t|}, \quad (11)$$

and the simulations of Mineo (2003) shows that the inverse of (11) 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 inference regarding the value of τ in terms of an LR-test. Table 5 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 applications

In this section we illustrate and assess the methods and algorithm further through three empirical applications.

4.1 Explanatory modelling of the mean and variance

In the empirical illustration of Sucarrat (2009) an explanatory model of weekly exchange rate return with homoscedastic errors outperforms standard volatility models (including realised volatility) in predicting variability *ex post*, both in-sample and out-of-sample. The principal source of the unusually high explanatory power (about 40% in terms of R^2) is forward order flow. Here, we revisit the model by extending the data sample and by adding more explanatory variables to the GUM, both in the mean and variance specifications. The overall GUM we start from contains 14 deletable regressors in the mean and 14 in the variance, a specification that would surely cause a lost of headache for anyone trying to implement multi-path GETS of an ML-estimated GARCH model. Next, we undertake multi-path GETS specification search by means of the AutoSEARCH algorithm.

The overall GUM we start from is (p -values in square brackets)

$$\begin{aligned} r_t = & \psi_0 + \psi_1 r_{t-1} + \psi_2 r_{t-1}^2 + \psi_3 \Delta x_t + \psi_4 \Delta x_{t-1} + \psi_5 \Delta(ir_t^{no} - ir_t^{eu}) \\ & + \psi_6 \Delta(ir_{t-1}^{no} - ir_{t-1}^{eu}) + \psi_7 \Delta oil p_t + \psi_8 \Delta oil p_{t-1} + \psi_9 \Delta ose_t + \psi_{10} \Delta ose_{t-1} \\ & + \psi_{11} \Delta sp100_t + \psi_{12} \Delta sp100_{t-1} + \psi_{13} \widehat{ECM}_{t-1} + \epsilon_t, \end{aligned}$$

$$\widehat{ECM}_t = s_t - 2.16 + 0.03 ir_t^{no} - 0.06 ir_t^{eu},$$

$$\epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0, 1),$$

$$\begin{aligned} \log \sigma_t^2 = & \alpha_0 + \sum_{p=1}^5 \alpha_p \log \epsilon_{t-p}^2 + \lambda (\log |\epsilon_{t-1}|^2) I_{z_{t-1} < 0} + \omega_0 \log EWMA(8)_{t-1} \\ & + \omega_1 \Delta v_t + \omega_2 v_{t-1} + \omega_3 (\Delta ir_t^{no})^2 + \omega_4 (\Delta ir_t^{eu})^2 + \omega_5 (\Delta oil p_t)^2 + \omega_6 (\Delta ose_t)^2 \\ & + \omega_7 (\Delta sp100_t)^2 \end{aligned}$$

$$R^2 : 0.35 \quad AR_{1-2} : \begin{matrix} 1.60 \\ [0.45] \end{matrix} \quad ARCH_{1-6} : \begin{matrix} 1.55 \\ [0.96] \end{matrix} \quad JB : \begin{matrix} 7.37 \\ [0.03] \end{matrix} \quad T = 190$$

The sample goes from 2 October 2005 to 5 July 2009 (197 end-of-week observations), where $r_t = (s_t - s_{t-1}) \times 100$ is log-return in % of the NOK/EUR exchange rate (an increase means the NOK depreciates), Δx_t is the associated forward order flow (buy initiated volume - sell initiated volume) in billions of NOK, ir_t^{no} is the 1-week Norwegian interbank money market interest rate in %, ir_t^{eu} is the 1-week

Eurozone interbank money market interest rate in %, $oilp_t$ is the log of oilprice, ose_t is the log of the main stock market index at the Oslo stock exchange, $sp100_t$ is the log of the Standard and Poor's 100 index of US stocks, \widehat{ECM}_t is the estimate of an error correction model, $EWMA(8)_t$ is an eight week equally weighted moving average of the squared residuals $\{\hat{\epsilon}_t^2\}$, v_t is the log of weekly forward volume (buy initiated volume + sell initiated volume), AR_{1-2} is a Ljung and Box (1979) test for serial correlation in the standardised residuals up to the 2nd. order, $ARCH_{1-6}$ is a Ljung and Box (1979) test for serial correlation up to the sixth order in the squared standardised residuals, JB is the Jarque and Bera (1980) test for non-normality, and T is the number of observations used in the estimation.¹⁵ GETS specification search of the mean is undertaken while holding the variance specification constant, with the mean constant restricted from deletion. Two diagnostic checks of the standardised residuals are undertaken after each regressor removal, AR_{1-2} and $ARCH_{1-6}$, both at 2.5%, and the Schwarz information criterion (made up of a Gaussian log-likelihood in the residuals of the mean) is used as tie-breaker. The search yields four parsimonious terminal models, the smallest one being the empty model and the least parsimonious containing four regressors. Among these the least parsimonious specification yields the smallest information criterion. Next, GETS specification search of the variance specification while holding the mean specification fixed yields (p -values in square brackets)

$$\hat{r}_t = \underset{[0.77]}{1.89} + \underset{[0.02]}{1.80}r_{t-1} - \underset{[0.02]}{3.25}\Delta x_{t-1} - \underset{[0.00]}{9.88}\Delta oilp_t - \underset{[0.01]}{6.39}\widehat{ECM}_{t-1},$$

$$\hat{\epsilon}_t = \hat{\sigma}_t \hat{z}_t, \quad \hat{z}_t \stackrel{IID}{\sim} GED(\hat{\tau}), \quad \hat{\tau} = \underset{[0.08]}{1.49},$$

$$\log \hat{\sigma}_t^2 = \underset{[0.01]}{-9.88} + \underset{[0.00]}{1.44}\Delta v_t + \underset{[0.02]}{0.89}v_{t-1}$$

$$R^2 : 0.27 \quad AR_{1-2} : \underset{[0.76]}{0.56} \quad ARCH_{1-6} : \underset{[0.92]}{2.00} \quad JB : \underset{[0.06]}{5.59} \quad T = 190$$

Also here are two diagnostic tests undertaken at each regressor deletion, AR_{1-2} and $ARCH_{1-6}$ at 2.5%, but the Schwarz criterion is now computed in terms of the standardised residuals.

The mean specification differs in two ways compared with Sucarrat (2009). The lagged order flow Δx_{t-1} is retained instead of the contemporaneous order flow, and the contemporaneous relative change in oil price is now found to be significant. Further investigation suggests the most important reasons for the difference is that the impact of contemporaneous forward order flow changed in the course of the

¹⁵The rawdata of s_t , ir_t^{no} , ir_t^{eu} , $oilp_t$, ose_t and $sp100_t$ are the daily series ew:nor19101, ew:nor14265, ew:emu14313, ew:com20220, ew:nor15565 and ew:usa15100200, respectively, from Reuters - EcoWin. The Norwegian order flow data can be downloaded via the url http://www.norges-bank.no/templates/reportroot___60389.aspx and are described in more detail in Meyer and Skjelvik (2006).

financial events of 2008-2009, and that both the oil price and the value of the NOK fell substantially over the same period. The resulting variance specification contains two terms: Changes in market activity as measured by the relative change in forward volume, and the lagged (log)level of volume. Finally, the estimate of the shape parameter τ is equal to 1.49, which suggests a slight departure of normality.

4.2 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

$$\log \sigma_t^2 = \alpha_0 + \omega \log RV_t. \quad (12)$$

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 6 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 6 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 convincingly 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 7 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. In the specification search of the mean the standardised residuals are checked after each deletion for serial correlation up to the 3rd. order, and the squared standardised residuals are checked for serial correlation up to the 5th. order. Schwarz's information criterion, computed in terms of a Gaussian log-likelihood made up of the residuals of the mean, is used as tie-breaker for terminal models in the mean. The Schwarz criterion is also used as a tie-breaker between log-variance specifications. However, in this case the standardised residuals are used for the Gaussian log-likelihood. The conclusion is that RV_t^{5m} does not parsimoniously encompass all the other terms, since a 4th. order log-ARCH term is retained in addition to $\log RV_t^{5m}$ in the log-variance specification. The constants in the mean and log-variance specifications are retained because they are restricted to not be deleted.

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.

4.3 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).¹⁶ 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 of 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 in general to four calendar weeks, and $EWMA_{t-1}$ 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 (in %) 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 $\log EWMA_{t-1}$ 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

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

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 8 contains the estimation results and associated diagnostic tests of the six models, figure 1 plots the standardised residuals of the six models both in-sample and out-of-sample, and figure 2 plots the fitted conditional standard deviations both in-sample and out-of-sample. The last figure suggests that the conditional standard deviations are relatively similar for all models, apart from those of the constant model. As expected, the first 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. By contrast, the two simplest of the time-varying volatility models, the EWMA and RiskMetrics, perform reasonably well. Visually, the variance of

¹⁶The 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—in particular during the autumn of 2008—means the evaluation provides a real test of the methods.

their standardised residuals appears constant across *both* the estimation and forecast samples in figure 1. The two more “advanced” models, the GARCH(1,1) and log-GARCH(1,1) specifications, also exhibit reasonably constant variances, with the GARCH(1,1) the more so. 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.

Table 9 contains the VaR 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 . For the 1% quantile there is a clear gain in using a GED distribution with estimated shape parameter, in the sense that this brings the unconditional coverages \hat{p}_1 closer to 0.01. However, there is no clear gain for the 5% and 10% quantiles. With respect to a comparison amongst models in the second panel, the results are very similar and all are usefully close to the density quantiles. However, if we were to rank the models according to their position for each quantile, then the GARCH(1,1) comes out of top and the SEARCH model comes second. Overall, then, the out-of-sample VaR forecast evaluation suggests the methods and algorithms can be of great value in practice. 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, for example by introducing skewness.

5 Conclusions

By making use of the recent results in Sucarrat and Escibano (2010), we have proposed methods and algorithms that resolves many of the problems earlier faced in the implementation of automated multi-path General-to-Specific (GETS) specification search of financial models. The simulations and empirical applications suggest the methods can be of great value in practice for the following reasons: First, our simulations show that the GETS algorithm we propose is equally or less likely to retain irrelevant variables compared with the other currently available GETS algorithms when no variables matter. This feature is particularly desirable in models for financial economics and business finance. Moreover, the low irrelevance proportion does not seem to compromise the capability to retain relevant variables in a notable way. Second, our algorithm exhibits two desirable properties related to the chosen significance levels. That the irrelevance proportion is equal to or below the regressor significance level, and that the probability of recovering the simulation DGP tends to 1 - *overall diagnostic significance level* as the sample size increases. Third, in our first empirical application it takes us only a couple of seconds on an ordinary computer to undertake multi-path GETS modelling of a Stochastic Exponential ARCH

(SEARCH) model with a total of 28 deletable regressors in the mean and variance specifications. By contrast, multi-path GETS specification search of the GARCH counterpart of our SEARCH model, with joint ML estimation of the mean, variance and density of the standardised errors, is probably not feasible in practice, and would in any case require substantial effort and time by the modeller, in addition to numerous subjective decisions throughout the modelling process about starting values, convergence criteria, multiple optima, sensible estimates and so on. Fourth, our second empirical application show that our methods can be of great use in both evaluating the forecasts of volatility proxies, and in improving them. Fifth, our third empirical application show that the SEARCH model is also useful in density forecasting, as in Value-at-Risk forecasting. Nevertheless, there is still room for further improvement, generalisation and exploration of the methods and algorithms we have proposed in this paper. For example, the efficiency of the estimation procedures may be improved through (say) feasible generalised least squares (FGLS) procedures and/or iterative least squares procedures, and common outlier detection algorithms (developed for the identification of outliers in the mean) can readily be adapted to search for Bernoulli jumps in the log-variance specification.

References

- Anscombe, F. and W. Glynn (1983). Distribution of kurtosis statistic for normal statistics. *Biometrika* 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>.
- 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 Inflation. *Econometrica* 50, 987–1008.
- 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.
- Hansen, P. R. (2005). A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics* 23, 365–380.
- 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.
- Meyer, E. and J. Skjelvik (2006). Statistics on foreign exchange transactions — new insight into foreign exchange markets. *Norges Bank Economic Bulletin* (2/06), 80–88.
- 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.
- 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.
- Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach. *The Annals of Statistics* 34, 2449–2495.
- 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>.
- Sucarrat, G. and Á. Escribano (2009). Automated Financial Multi-Path GETS Modelling. UC3M Working Paper 09-36 in the Economic Series. Available as: <http://www.eco.uc3m.es/temp/09-36-20.pdf>.
- Sucarrat, G. and Á. Escribano (2010). The Power Log-GARCH Model. Work in progress. Current version available as <http://www.eco.uc3m.es/sucarrat/research/loggarch.pdf>.
- 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

$$f(z; \tau) = \frac{\tau \Gamma(3/\tau)^{1/2}}{2\Gamma(1/\tau)^{3/2}} \exp \left[-|z|^\tau \cdot \left(\frac{\Gamma(3/\tau)}{\Gamma(1/\tau)} \right)^{\tau/2} \right], \quad -\infty < z < \infty, \quad \tau > 1, \quad (13)$$

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 (13) equals the standard normal, when $\tau < 2$ then (13) has thicker tails than the standard normal, and when $\tau > 2$ then (13) has thinner tails than the standard normal. In particular, when $\tau \rightarrow 1$ then (13) tends to the double exponential distribution, and when $\tau \rightarrow \infty$ then (13) tends to a uniform distribution on the interval $[-3^{(1/2)}, 3^{(1/2)}]$ (Nelson 1991, pp. 352-353). The distribution (13) 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).¹⁷ In that package the density of the exponential power distribution is given by

$$f(x, \mu, \sigma, p) = \frac{1}{2p^{1/p}\Gamma(1 + 1/p)\sigma} \exp\left(\frac{-|x - \mu|^p}{p\sigma^p}\right), \quad (14)$$

and the GED(τ) density given by (13) is obtained by setting $\tau = p$, $\mu = 0$ and

$$\sigma = \frac{\Gamma(1/\tau)^{(1/2)}}{\tau^{(1/\tau)}\Gamma(3/\tau)^{(1/2)}}. \quad (15)$$

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

¹⁷Further 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).

Table 1: List of multi-path GETS experiments

k_0	k_1	Simulation	DGP	GUM
HP0	0	40	$r_t = \epsilon_t$, $\epsilon_t = 130z_t$, $z_t \stackrel{IID}{\sim} GED(\tau)$	$r_t = \sum_{k=1}^{41} \psi_k x_{kt}^{HP} + \epsilon_t$, 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	$r_t = \epsilon_t$, $\epsilon_t = \sigma z_t$, $z_t \stackrel{IID}{\sim} GED(2)$, $E(\epsilon_t^2) \approx 130^2$ $\log \sigma_t^2 = 1.0885 + 0.1 \log \epsilon_{t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in HP0
HP2'	1	39	$r_t = 0.75r_{t-1} + \epsilon_t$, $\epsilon_t = 85.99z_t$, $z_t \stackrel{IID}{\sim} GED(\tau)$	Same as in HP0
HP2**	1	39	$r_t = 0.75r_{t-1} + \epsilon_t$, $\epsilon_t = \sigma z_t$, $z_t \stackrel{IID}{\sim} GED(2)$, $E(\epsilon_t^2) \approx (85.99)^2$, $\log \sigma_t^2 = 1.0058 + 0.1 \log \epsilon_{t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in HP0
HP7'	3	37	$r_t = 0.75r_{t-1} + 1.33x_{11t}^{HP} + 0.9975x_{20t}^{HP} + \epsilon_t$, $\epsilon_t = 85.99z_t$, $z_t \stackrel{IID}{\sim} GED(\tau)$	Same as in HP0
SE1	0	9	$r_t = \epsilon_{1t}$, $\epsilon_{1t} = \sigma_t z_t$, $z_t \stackrel{IID}{\sim} GED(2)$ $\log \sigma_t^2 = 0.1 \log \epsilon_{1t-1}^2 + 0.8 \log \sigma_{t-1}^2$	$r_t = \phi_0 + \sum_{m=1}^2 \phi_m r_{t-m} + \sum_{n=0}^2 \eta_n x_{t-n} + \eta_3 x_{1t}^{N(0,1)} + \eta_4 x_{2t}^{N(0,1)} + \eta_5 x_{1t}^{EXP(1)} + \eta_6 x_{2t}^{EXP(1)} + \epsilon_{1t}$, where $x_t = 0.1 + 0.9x_{t-1} + \epsilon_{2t}$ with $\epsilon_{2t} \stackrel{IID}{\sim} N(0,1)$
SE2	1	8	$r_t = 0.1r_{t-1} + \epsilon_{1t}$, $\epsilon_{1t} = \sigma_t z_t$, $z_t \stackrel{IID}{\sim} GED(2)$ $\log \sigma_t^2 = 0.1 \log \epsilon_{1t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in SE1
SE3	0	12	$r_t = \epsilon_{1t}$, $\epsilon_{1t} = \sigma_t z_t$, $\sigma_t = 1$, $z_t \stackrel{IID}{\sim} GED(\tau)$	$\log \sigma_t^2 = \alpha_0 + \sum_{p=1}^P \alpha_p \log \epsilon_{1t-p}^2 + \lambda_1 (\log \epsilon_{1t-1}^2) I_{z_{t-1} < 0} + \omega_1 y_t + \omega_2 y_{t-1} + \omega_3 x_{1t}^{EXP(1)} + \omega_4 x_{2t}^{EXP(1)} + \omega_5 x_{3t}^{N(0,1)} + \omega_6 x_{4t}^{N(0,1)}$, where $y_t = 0.9y_{t-1} + \epsilon_{2t}$ with $\epsilon_{2t} \stackrel{IID}{\sim} N(0,1)$
SE4	1	11	$r_t = \epsilon_{1t}$, $\epsilon_{1t} = \sigma_t z_t$, $z_t \stackrel{IID}{\sim} GED(\tau)$ $\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \epsilon_{1t-1}^2$	Same as in SE3
SE5	0	2	$r_t = \epsilon_{1t}$, $\epsilon_{1t} = \sigma_t z_t$, $\sigma_t = 1$, $z_t \stackrel{IID}{\sim} GED(\tau)$	$\log \sigma_t^2 = \alpha_0 + \alpha_1 \log \epsilon_{t-1}^2 + \beta_1 \log \sigma_{t-1}^2$
SE6	2	0	$r_t = \epsilon_{1t}$, $\epsilon_{1t} = \sigma_t z_t$, $z_t \stackrel{IID}{\sim} GED(\tau)$ $\log \sigma_t^2 = 0.1 \log \epsilon_{1t-1}^2 + 0.8 \log \sigma_{t-1}^2$	Same as in SE5

The design of the experiments HP0, HP2' and HP7' are based on Hoover and Perez (1999), and make use of their data x_{1t}, \dots, 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 2: 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		0.031	0.486
			AutoSEARCH w/PET			0.035	0.459
			AutoSEARCH w/PET, no empty			0.052	0.215
			HP			0.045	0.292
			PcGets			≈ 0.04	≈ 0.45
HP2'	1	39	AutoSEARCH	139	1.000	0.050	0.247
			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	1.000	0.052	0.249
			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 actual 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 actual 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 and “no empty” means no attempt was made to include the empty model. In the diagnostic checks both the AR and ARCH test of the standardised residuals were undertaken at lag 2 using a significance level of 2.5% for each, and as tiebreaker the Schwarz information criterion is used with a Gaussian log-likelihood made up of the mean residuals $\{\hat{\epsilon}_t\}$.

Table 3: Properties of AutoSEARCH algorithm: 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		0.037	0.534
HP2**	1	39	AutoSEARCH	139	1.000	0.079	0.127
SE1	0	9	AutoSEARCH	200		0.034	0.777
				500		0.026	0.811
				1000		0.021	0.855
SE2	1	8	AutoSEARCH	200	0.285	0.034	0.198
				500	0.562	0.038	0.385
				1000	0.815	0.039	0.600

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. Only one diagnostic check (AR) of the standardised residuals was undertaken at lag 2 using a significance level of 5%, and as tiebreaker the Schwarz information criterion is used with a Gaussian log-likelihood made up of the mean residuals $\{\hat{\epsilon}_t\}$.

Table 4: 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		0.004	0.961
					1.1		0.018	0.838
				500	2.0		0.004	0.969
					1.1		0.018	0.832
				1000	2.0		0.005	0.966
					1.1		0.016	0.865
SE4	1	11	AutoSEARCH	200	2.0	0.430	0.036	0.262
					1.1	0.284	0.032	0.166
				500	2.0	0.916	0.046	0.568
					1.1	0.633	0.035	0.390
				1000	2.0	0.997	0.046	0.629
					1.1	0.896	0.046	0.526
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 SE6 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 simulation 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 5: 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 6: Logarithmic Mincer-Zarnowitz regressions of variability (squared return) on volatility proxies

Model	$\hat{\alpha}$ [<i>p-val</i>]	$\hat{\beta}$ [<i>p-val</i>]	$\chi^2(2)$ [<i>p-val</i>]	<i>AR</i> (1) [<i>p-val</i>]	<i>ARCH</i> (5) [<i>p-val</i>]	<i>JB</i> [<i>p-val</i>]
$\log \sigma_t^2 = \alpha + \beta \log RV_t^{65m}$	-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^{15m} 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 7: Parsimonious encompassing test of RV_t^{5m}

MGUM:	$\hat{r}_t = 0.048 - 0.036r_{t-1} - 0.028r_{t-2} + 0.004RV_{t-1}^{5m} + 0.002RV_{t-2}^{5m}$
VGUM 1:	$\log \hat{\sigma}_t^2 = -0.192 + 1.08 \log RV_t^{5m}$
	$AR(3) : -0.02 \quad ARCH(5) : 0.05$
	$[p.val.] \quad [0.40] \quad [p.val.] \quad [0.03]$
VGUM 2:	$\log \hat{\sigma}_t^2 = -0.191 - 0.027 \log \hat{\epsilon}_{t-1}^2 - 0.001 \log \hat{\epsilon}_{t-2}^2 - 0.015 \log \hat{\epsilon}_{t-3}^2 + 0.031 \log \hat{\epsilon}_{t-4}^2$
	$+ 0.024(\log \hat{\epsilon}_{t-1}^2) \hat{I}_{\{\epsilon_{t-1} < 0\}} + 1.077 \log RV_t^{5m}$
	$AR(3) : -0.02 \quad ARCH(5) : 0.05$
	$[p.val.] \quad [0.40] \quad [p.val.] \quad [0.20]$
MSPEC:	$\hat{r}_t = 0.075$
	$[p.val.] \quad [0.07]$
VSPEC:	$\log \hat{\sigma}_t^2 = -0.170 + 0.035 \log \hat{\epsilon}_{t-4}^2 + 1.065 \log RV_t^{5m}$
	$AR(3) : -0.01 \quad ARCH(5) : 0.05$
	$[p.val.] \quad [0.18] \quad [p.val.] \quad [0.04]$

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. $AR(3)$ is a Ljung and Box (1979) test of serial correlation in the standardised residuals $\{z_t\}$ up to order 3 and $ARCH(5)$ is a Ljung and Box (1979) test of serial correlation in the squared standardised residuals $\{z_t^2\}$ up to order 5.

Table 8: Estimation results of VaR models

Model	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}$	$\hat{\omega}$	AR_1 [<i>p.val.</i>]	$ARCH_1$ [<i>p.val.</i>]	$ARCH_2$ [<i>p.val.</i>]	$ARCH_5$ [<i>p.val.</i>]	$\hat{\tau}$ [<i>p.val.</i>]
Constant	1.27				-0.040 [0.15]	0.176 [0.00]	0.247 0.000	0.207 0.000	1.040 [0.00]
EWMA	0.00			1.00	-0.046 [0.10]	-0.028 [0.31]	0.025 [0.40]	0.028 [0.23]	1.484 [0.00]
RiskMetrics		0.06	0.94		-0.045 [0.10]	-0.028 [0.31]	0.023 [0.42]	0.013 [0.57]	1.492 [0.00]
GARCH	0.01	0.06	0.93		-0.040 [0.14]	-0.041 [0.13]	0.025 [0.21]	0.013 [0.43]	1.488 [0.00]
log-GARCH	0.07	0.05	0.93		-0.021 [0.46]	-0.003 [0.90]	0.065 [0.06]	0.091 [0.00]	1.336 [0.00]
SEARCH	0.09			0.89	-0.045 [0.11]	-0.029 [0.29]	0.033 [0.29]	0.036 [0.08]	1.500 [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. All computations in R.

Table 9: Out-of-sample VaR forecast evaluation results of SP500 returns

Model	1%		5%		10%	
	\hat{p}_1 [<i>p.val.</i>]	$\hat{\pi}_{11}$ [<i>p.val.</i>]	\hat{p}_1 [<i>p.val.</i>]	$\hat{\pi}_{11}$ [<i>p.val.</i>]	\hat{p}_1 [<i>p.val.</i>]	$\hat{\pi}_{11}$ [<i>p.val.</i>]
Constant- $N(0, 1)$	0.050 [0.00]	0.013 [1.00]	0.082 [0.00]	0.025 [1.00]	0.110 [0.36]	0.032 [1.00]
EWMA- $N(0, 1)$	0.030 [0.00]	0.000 [1.00]	0.071 [0.01]	0.003 [1.00]	0.110 [0.36]	0.015 [1.00]
RiskMetrics- $N(0, 1)$	0.034 [0.00]	0.000 [1.00]	0.068 [0.02]	0.006 [1.00]	0.110 [0.36]	0.015 [1.00]
GARCH(1,1)- $N(0, 1)$	0.028 [0.00]	0.000 [1.00]	0.065 [0.07]	0.003 [1.00]	0.107 [0.50]	0.015 [1.00]
log-GARCH(1,1)- $N(0, 1)$	0.039 [0.00]	0.000 [1.00]	0.088 [0.00]	0.012 [1.00]	0.118 [0.09]	0.024 [1.00]
SEARCH- $N(0, 1)$	0.030 [0.00]	0.000 [1.00]	0.065 [0.07]	0.003 [1.00]	0.108 [0.43]	0.012 [1.00]
Constant- $GED(\hat{\tau})$	0.037 [0.00]	0.013 [1.00]	0.084 [0.00]	0.025 [1.00]	0.110 [0.36]	0.032 [1.00]
EWMA- $GED(\hat{\tau})$	0.024 [0.00]	0.000 [1.00]	0.068 [0.02]	0.003 [1.00]	0.118 [0.09]	0.015 [1.00]
RiskMetrics- $GED(\hat{\tau})$	0.026 [0.00]	0.000 [1.00]	0.068 [0.02]	0.006 [1.00]	0.118 [0.09]	0.018 [1.00]
GARCH(1,1)- $GED(\hat{\tau})$	0.022 [0.00]	0.000 [1.00]	0.065 [0.07]	0.003 [1.00]	0.114 [0.17]	0.015 [1.00]
log-GARCH(1,1)- $GED(\hat{\tau})$	0.028 [0.00]	0.000 [1.00]	0.088 [0.00]	0.000 [1.00]	0.129 [0.01]	0.034 [1.00]
SEARCH- $GED(\hat{\tau})$	0.023 [0.00]	0.000 [1.00]	0.067 [0.03]	0.003 [1.00]	0.118 [0.09]	0.012 [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 $\hat{\pi}_{11}$ 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.

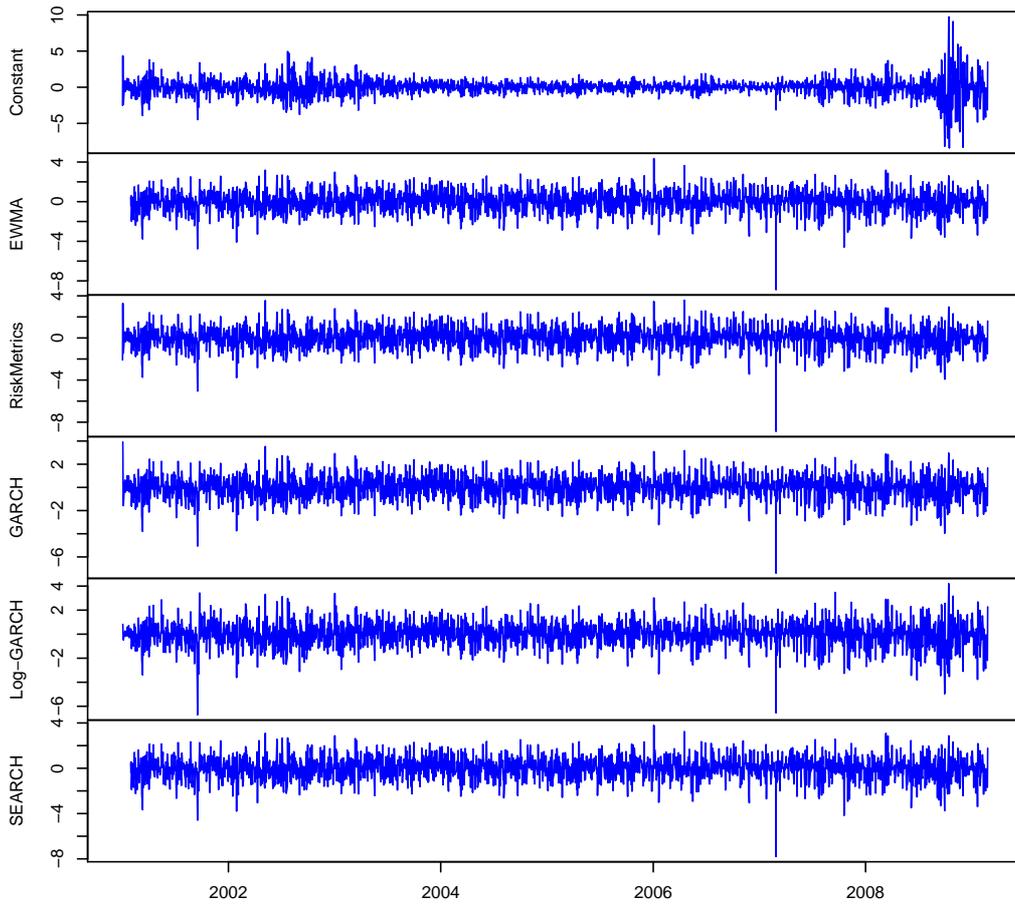


Figure 1: Standardised residuals of VaR forecast models over the estimation and design sample 1 January 2001 - 30 December 2005, and over the forecast sample 2 January 2006 - 24 February 2009

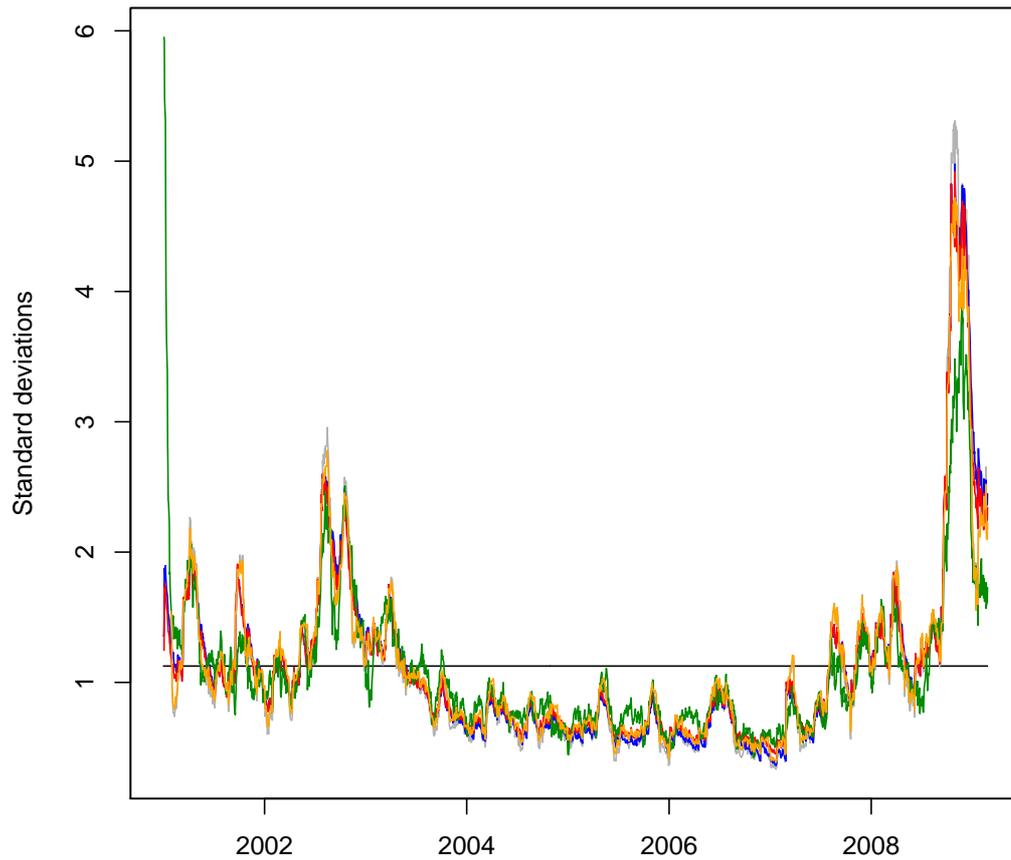


Figure 2: Fitted conditional standard deviations of VaR forecast models over the estimation and design sample 1 January 2001 - 30 December 2005, and over the forecast sample 2 January 2006 - 24 February 2009. Black line: Constant; Grey line: EWMA; Blue line: RiskMetrics; Red line: GARCH(1,1); Green line: Log-GARCH(1,1); Orange line: SEARCH