UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Score-driven time series models

Andrew
Harvey

Abstract

The construction of score-driven filters for nonlinear time series models is described and it is shown how they apply over a wide range of disciplines. Their theoretical and practical advantages over other methods are highlighted. Topics covered include robust time series modeling, conditional heteroscedasticity, count data, dynamic correlation and association, censoring, circular data and switching regimes.

Reference Details

CWPE	2133
Published	7 April 2021
Key Words	copula, count data, directional data, generalized autoregressive conditional heteroscedasticity, generalized beta distribution of the second kind, observation-driven model, robustness
JEL Codes	C22, C32
Website	www.econ.cam.ac.uk/cwpe

Score-driven time series models

Andrew Harvey*

Faculty of Economics, Cambridge University

Prepared for Volume 9 of the Annual Review of Statistics and Its Application.

Submitted. doi: 10.1146/annurev-statistics-040120-021023

April 7, 2021

Abstract

The construction of score-driven filters for nonlinear time series models is described and it is shown how they apply over a wide range of disciplines. Their theoretical and practical advantages over other methods are highlighted. Topics covered include robust time series modeling, conditional heteroscedasticity, count data, dynamic correlation and association, censoring, circular data and switching regimes.

KEYWORDS: copula; count data; directional data; generalized autoregressive conditional heteroscedasticity; generalized beta distribution of the second kind; observation-driven model; robustness.

*ach34@cam.ac.uk

1 Introduction

One of the principal aims of time series modeling is to construct filters, that is functions of current and past observations, that estimate where we are now, where we were in the past and where we might be in the future. These objectives are called nowcasting, smoothing and forecasting respectively. When models are linear and Gaussian, the optimal solutions to all three problems are given by the Kalman filter and smoother. For nonlinear models, recent work has shown that observation-driven models based on the score of the conditional distribution provide an integrated solution to the forecasting problem that is theoretically sound and yields results that, on the whole, compare favourably with those obtained by competing methods.

The purpose of observation-driven nonlinear models is to estimate a changing moment and so they entail setting up a dynamic equation driven by a variable whose expectation is equal to that moment. The generalized autoregressive conditional heteroscedasticity (GARCH) model, which is widely used in finance, is a leading example. Many other nonlinear models have dynamics based on arbitrary forcing variables whose main appeal is a simple interpretation. While such models are important historically, they become less appealing once the score-driven solution becomes apparent. The use of the score guards against features of the data, such as extreme values, that might throw a filter off course. In contrast to moment-based models, robustness is automatically built into the filter. More generally the score leads to the construction of forcing variables in filters that respect the key features of the data and whose forms have a natural and intuitive interpretation even in cases where the analytic expression is complex; scores for dynamic copulas are a good illustration.

The defining characteristic of observation-driven models is that they are formulated in terms of the one-step ahead predictive distribution. Hence the likelihood function is immediately available. This is not the case with nonlinear parameter-driven models, a distinction due to Cox (1981). At one time parameter-driven models seemed the way forward as they could impose a meaningful structure that reflected the nature of the problem and could potentially impose the kind of restrictions on the parameter space inherent in linear state space models. More computing power and the attendent algorithmic development led to an increase in the use of computer-intensive approaches, especially Bayesian methods; see Durbin and Koopman (2012). However, when observation-driven models are driven by the score the balance shifts and, in many situations, they become more attractive than parameter-driven models. Indeed a score-driven model can be regarded as providing an approximation to the computer-intensive solution for the corresponding parameter-driven unobserved components model. Koopman, Lucas and Scharth (2016) demonstrate that the approximation is usually a very good one.

This article discusses the score-driven approach to modeling in a wide range of situations. The aim is to consolidate the new results that have appeared since the publication of the book[1] by Harvey (2013) and the articles by Creal, Koopman and Lucas (2011, 2013). However, it is not intended to be comprehensive in its coverage. A full list of papers can be found on the website at the Free University of Amsterdam, that is http://www.gasmodel.com/index.htm.

Some packages already exist for estimating score-driven models. The new TSL package of Lit et al (2020) is menu-driven and intended to be a companion to the STAMP package of Koopman et al (2020). Programs in R are provided by Ardia et al (2019).

2 Unobserved components and filters

A simple Gaussian signal plus noise model for T observations, $y, \ldots, y_T$, is

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}\left(0, \sigma_\varepsilon^2\right), \quad t = 1, \ldots, T, \\ \mu_{t+1} &= \phi\mu_t + \eta_t, \quad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2), \end{aligned} \tag{1}$$

where the irregular and level disturbances, ε_t and η_t respectively, are mutually independent and $\mathrm{NID}(0, \sigma^2)$ denotes normally and independently distributed

[1]Score-driven model are called dynamic conditional score (DCS) models in Harvey (2013) and generalized autoregressive score (GAS) models in Creal et al (2013).

with mean zero and variance σ^2. The autoregressive parameter is ϕ and the signal-noise ratio is $q = \sigma_\eta^2/\sigma_\varepsilon^2$.

The unobserved component model in (1) is in state space form and, as such, it may be handled by the Kalman filter. The parameters ϕ and q can be estimated by maximum likelihood, with the likelihood function constructed from the one-step ahead prediction errors. The Kalman filter can be expressed as a single equation which combines $\mu_{t|t-1}$, the optimal estimator of μ_t based on information at time $t-1$, with y_t in order to produce the best estimator of μ_{t+1}. Writing this equation together with an equation that defines the one-step ahead prediction error, v_t, gives the innovations form of the Kalman filter:

$$\begin{aligned} y_t &= \mu_{t|t-1} + v_t, \\ \mu_{t+1|t} &= \phi\mu_{t|t-1} + k_t v_t. \end{aligned} \tag{2}$$

The Kalman gain, k_t, depends on ϕ and q. In the steady-state, k_t is constant.

When the disturbance term, ε_t, in (1) is non-Gaussian, the Kalman filter is no longer optimal, unless attention is confined to linear filters. The main ingredient in the score-driven approach is the replacement of v_t in the Kalman filter by a variable, u_t, that is proportional to the score, $\partial \ln f(y_t;\mu)/\partial \ln \mu$, of an assumed conditional distribution, $f(y_t;\mu)$. Thus the second equation in (2) becomes

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa u_t, \tag{3}$$

where κ is treated as an unknown parameter. The dynamics in a score-driven model need not be confined to location. A filter such as (3) may be cast in terms of the score with respect to any parameter, θ. When the information matrix is time-invariant, and the model is identifiable, the asymptotic distribution for the maximum likelihood estimator may be derived; see Harvey (2013, ch 2).

Likelihood-based tests can be constructed. For example, a test of time-variation may be carried out, prior to fitting a model, by a portmanteau test constructed from the autocorrelations of the scores in the static model. A test of this kind can be derived as a Lagrange multiplier test of the null hypothesis that $\kappa_0 = \kappa_1 = \cdots = \kappa_{P-1} = 0$, against the alternative that $\kappa_i \neq 0$ for some $i = 0, ..., P-1$, in the dynamic model

$$\theta_{t|t-1} = \omega + \kappa_0 u_{t-1} = \cdots = \kappa_{P-1} u_{t-P}, \quad t = 1, ..., T; \tag{4}$$

see Harvey (2013, sect 2.5), Harvey and Thiele (2016) and Calvori et al (2017).

3 Distributions and scores

The *location-dispersion* model is

$$y_t = \mu + \varphi\varepsilon_t, \qquad -\infty < y_t < \infty, \quad t = 1, ..., T, \tag{5}$$

where μ is location, the scale, $\varphi > 0$, is called the dispersion and ε_t is a standardized variable with a probability density function (PDF) that depends on one or more shape parameters. With an exponential link function, $\varphi = \exp \lambda$, the score for λ is

$$\partial \ln f_t(y_t; \mu, \lambda)/d\lambda = (y_t - \mu)\partial \ln f_t(y_t; \mu, \lambda)/\partial\mu - 1, \qquad t = 1, ..., T.$$

For a non-negative variable, the *location/scale* model is

$$y_t = \varphi\varepsilon_t, \qquad y_t \geq 0, \qquad t = 1, ..., T, \tag{6}$$

where the standardized variable, ε_t, has unit scale. The location is proportional to the scale, φ. Provided the mean of ε_t exists, $E(y_t) = \varphi E(\varepsilon_t)$.

Many of the distributions used for *location-dispersion* and *location/scale* models are related. As a result there is a unity in much of the technical discussion as it pertains to score-driven models. The generalized beta distribution of the second kind (GB2) distribution plays a prominent role. Its PDF is

$$f(y_t;\varphi,\upsilon,\xi,\varsigma)=\frac{\upsilon(y_t/\varphi)^{\upsilon\xi-1}}{\alpha B(\xi,\varsigma)\left[(y_t/\varphi)^{\upsilon}+1\right]^{\xi+\varsigma}},\qquad y_t\geq 0,\qquad \varphi,\upsilon,\xi,\varsigma>0, \quad (7)$$

where φ is the scale parameter, υ,ξ and ς are shape parameters and $B(\xi,\varsigma)$ is the beta function; see Kleiber and Kotz (2003, Ch 6). GB2 distributions are fat-tailed[2] for finite ξ and ς with upper and lower tail indices of $\eta=\varsigma\upsilon$ and $\overline{\eta}=\xi\upsilon$ respectively. The GB2 distribution contains many important distributions as special cases, including the Burr ($\xi=1$) and log-logistic ($\xi=1,\varsigma=1$). Other distributions are derived by simple transformations, as in the cases of F, generalized$-t$ and exponential generalized beta of the second kind. All the scores are functions of a variable that has a standard beta distribution.

The GB2 distribution, (7), can be reparameterized so that the (upper) tail index replaces ς, that is we define $\eta=\upsilon\varsigma$. To get the generalized gamma as a limiting case as $\eta\rightarrow\infty$ it is necessary to redefine the scale parameter in the GB2 distribution as $\varphi\eta^{1/\upsilon}$ so that its PDF becomes

$$f(y_t;\varphi,\upsilon,\xi,\eta)=\frac{\upsilon(y_t/\varphi)^{\upsilon\xi-1}}{\varphi\eta^{\xi}B(\xi,\eta/\upsilon)\left[(y_t/\varphi)^{\upsilon}/\eta+1\right]^{\xi+\eta/\upsilon}},\quad y_t\geq 0,\qquad \varphi,\upsilon,\xi,\eta>0. \quad (8)$$

The generalized gamma distribution is thin-tailed and the distributions of the scores are functions of a variable that has a standard gamma distribution.

The stationary first-order score-driven model corresponds to the Gaussian innovations form, (2), and is

$$\begin{aligned} y_t &= \mu_{t|t-1}+v_t=\mu_{t|t-1}+\exp(\lambda)\varepsilon_t, && t=1,...,T,\\ \mu_{t+1|t} &= \delta+\phi\mu_{t|t-1}+\kappa u_t, && |\phi|<1, \end{aligned} \quad (9)$$

[2] Embrechts, Kluppelberg and Mikosch (1997) define various categories of tails.

where $\omega = \delta/(1-\phi)$ is the unconditional mean of $\mu_{t|t-1}$, ε_t is a serially independent, standardized variate and u_t is proportional to the conditional score. More generally, a quasi-ARMA-type model of order (p, r) is

$$\mu_{t+1|t} = \delta + \phi_1 \mu_{t|t-1} = \cdots = \phi_p \mu_{t-p+1|t-p} + \kappa_0 u_t + \kappa_1 u_{t-1} = \cdots = \kappa_r u_{t-r}. \quad (10)$$

More than one component is possible. These components may be nonstationary, as in the model with trend and seasonality estimated by Caivano et al (2016). Explanatory variables can be introduced as in Harvey and Luati (2014).

An important aspect of the score-driven model is to guard against outliers. The attractions of using the t-distribution for this purpose are discussed in Lange, Little and Taylor (1989). In this article the t-distribution is discussed in the wider context of the generalized-t and exponential GB2 distributions and the connections with the robustness literature, as described in Maronna et al (2006), explored.

3.1 Student-t and Generalized-t

The generalized Student t distribution proposed by McDonald and Newey (1988) contains the general error distribution (GED) and the Student t distribution as special cases. The PDF is

$$f(y_t; \mu, \upsilon, \eta) = \frac{\upsilon}{2\eta^{1/\upsilon}} \frac{1}{B(1/\upsilon, \eta/\upsilon)} \frac{1}{(1 + |(y_t - \mu)/\varphi|^{\upsilon} / \eta)^{(\eta+1)/\upsilon}}, \quad -\infty < y_t < \infty, \quad (11)$$

where υ and η are positive shape parameters and $\upsilon = 2$ gives Student's t with η degrees of freedom. The distribution has fat tails when the tail index, η, is finite. Letting $\eta \to \infty$ yields the GED, with $\upsilon = 1$ giving the Laplace or double exponential distribution and $\upsilon = 2$ the Gaussian distribution. The absolute value of a generalized-t variable has a GB2 density in the form (8), but with the constraint $\xi = 1/\upsilon$, so the mode is at zero. When location is dynamic, its conditional score is

$$\frac{\partial \ln f_t(y_t;\mu_{t|t-1},\upsilon,\eta)}{\partial \mu_{t|t-1}} = \frac{\eta+1}{\eta e^{\lambda}}(1-b_t)\left|\varepsilon_t\right|^{\upsilon-1}\operatorname{sgn}(y_t-\mu_{t|t-1}), \qquad (12)$$

where

$$b_t = \frac{(\left|y_t-\mu_{t|t-1}\right| e^{-\lambda})^{\upsilon}/\eta}{(\left|y_t-\mu_{t|t-1}\right| e^{-\lambda})^{\upsilon}/\eta+1}, \qquad 0 \leq b_t \leq 1, \quad 0 < \eta < \infty, \qquad (13)$$

is distributed as beta $(1/\upsilon, \eta/\upsilon)$; see Harvey and Lange (2017). Provided η is finite, the score (influence) function of location is redescending in that it approaches zero as y moves away from zero.

Because the $u_t's$ are IID$(0,\sigma_u^2)$, that is independent and identically distributed with zero mean and variance σ_u^2, $\mu_{t|t-1}$ is weakly and strictly stationary so long as $|\phi| < 1$. All moments of u_t exist and the existence of moments of y_t is not affected by the dynamics. The autocorrelations can be found from the infinite moving average representation. The patterns are as they would be for a Gaussian model; see Harvey (2013, chapter 3). Maximum likelihood estimation is straightforward and for a first-order dynamic equation, as in (9), an analytic expression for the information matrix is available.

3.2 Exponential generalized beta distribution of the second kind (EGB2)

The EGB2 distribution results from taking the logarithm of a variable with a GB2 distribution, (7). It has light (exponential) tails. When $\xi = \varsigma$, it is symmetric, with $\xi = \varsigma = 1$ giving a logistic distribution. The normal distribution is obtained as a limiting case when $\xi = \varsigma \to \infty$.

The score function is bounded for positive ξ and ς, giving a gentle form

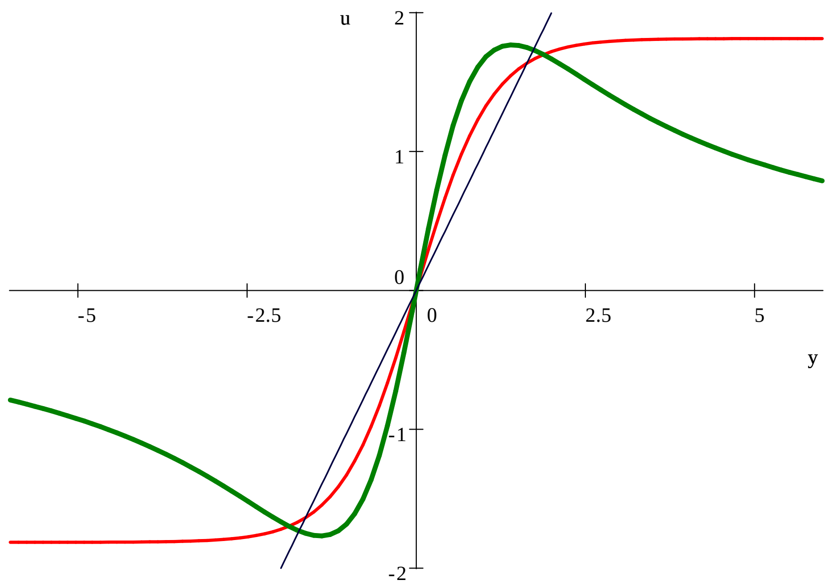

Figure 1: Location score for t_4 (thick line), logistic (thin) and normal (straight line). All these distributions have unit standard deviation.

of Winsorizing. Specifically,

$$\partial \ln f_t(y_t; \mu_{t|t-1}, \varphi, \upsilon, \xi, \varsigma)/\partial \mu_{t|t-1} = \upsilon(\xi + \varsigma)b_t(\xi, \varsigma) - \upsilon\xi, \qquad t = 1, ..., T,$$

where

$$b_t(\xi, \varsigma) = e^{(y_t - \mu_{t|t-1})\upsilon}/(e^{(y_t - \mu_{t|t-1})\upsilon} + 1). \tag{14}$$

has a beta distribution with parameters ξ and ς. Because $0 \leq b_t(\xi, \varsigma) \leq 1$, it follows that as $y_t \to \infty$, the score approaches an upper bound of $\upsilon\varsigma$, whereas $y_t \to -\infty$ gives a lower bound of $-\upsilon\xi$; see Caivano and Harvey (2014). Figure 1 constrasts the score with that of a t_4 distribution. The shapes are unaltered if the scores are divided by their information quantities.

4 Scale

Since the 1980's, the generalized autoregressive conditional heteroscedasticity (GARCH) model has been the standard way of modeling changes in the volatility of returns; see Bollerslev et al (1994). It is a moment-based observation-driven model in which the conditional variance is a linear function of past squared observations. The first-order case, the GARCH $(1,1)$ model, is

$$y_t = \mu + \sigma_{t|t-1}\varepsilon_t, \quad \varepsilon_t \sim \mathrm{IID}(0,1), \quad t = 1,..,T, \tag{15}$$

and

$$\sigma^2_{t|t-1} = \delta + \beta\sigma^2_{t-1|t-2} + \alpha y^2_{t-1}, \quad \delta > 0, \beta \geq 0, \alpha \geq 0. \tag{16}$$

The GARCH-t model introduced by Bollerslev (1987) has long been an industry standard. The restrictions on the parameters ensure that the variance remains positive. An alternative way of achieving this objective is to set up the dynamic equation in terms of the logarithm of $\sigma^2_{t|t-1}$. This is the exponential GARCH (EGARCH) model of Nelson (1991). In the corresponding parameter-driven stochastic volatility (SV) model, the logarithm of the standard deviation, λ_t in

$$y_t = \mu + \sigma_t\varepsilon_t, \qquad \sigma^2_t = \exp(2\lambda_t), \qquad \varepsilon_t \sim \mathrm{IID}(0,1), \tag{17}$$

is an unobserved component. It is usually set up as a Gaussian first-order autoregessive process, though it seems that here is no compelling reason for an assumption of Gaussianity. The likelihood function is not available in closed form, so computer-intensive methods are needed to estimate it efficiently. Nelson showed that the EGARCH model could be regarded as an approximate filter for the SV model. However, in his classic formulation the conditions for the existence of the moments of the observations do not normally hold when the conditional distribution is Student t with finite degrees of freedom. Using the score to define the forcing variables solves this problem[3].

The score-driven EGARCH model is set up as

$$y_t = \mu + \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ..., T, \tag{18}$$

where the $\varepsilon_t' s$ are independently and identically distributed with location zero and unit scale[4]. The stationary first-order dynamic model for $\lambda_{t|t-1}$, the logarithm of the scale, is

$$\lambda_{t+1|t} = \omega(1-\phi) + \phi\lambda_{t|t-1} + \kappa u_t, \quad |\phi| < 1, \tag{19}$$

where u_t is the score of the distribution of y_t conditional on past observations, $\lambda_{1|0} = \omega$ and ϕ and κ are parameters. When the conditional distribution is fat tailed, the score is bounded and so extreme observations are downweighted. As is clear from (16), this does not happen with the GARCH-t model. Letting the conditional distribution be Student t leads to a model known as Beta-t-EGARCH. The stationarity conditions are straightforward because in (19) all that is required is that $|\phi| < 1$ and, as Harvey and Lange (2017) show, the invertibility conditions of Blasques et al (2018) will be satisfied in most practical situations. This model has now been widely applied and shown to be more attractive than the standard GARCH-t model from both the practical and theoretical points of view; see for example, Harvey (2013, ch 4) and Catania and Nonejad (2019).

4.1 Generalized-t EGARCH

A generalized Student t distribution in (18) gives what Harvey and Lange (2017) call the Beta-Gen-t-EGARCH model. The GED is a limiting case, but the problem with the resulting Gamma-GED-EGARCH model is that the

[3]The solution is implied in Bollerslev et al (1994), where a forcing variable based on the generalized-t distribution is proposed. However, the idea was not followed up.

[4]The variance need not exist.

score is not bounded and so is vulnerable to fat tails. The classic EGARCH model of Nelson (1991) is a special case of the Gamma-GED-EGARCH model obtained when the distribution of the observations is Laplace. The conditional score for the logarithm of the dynamic scale parameter of the generalized-t distribution is

$$u_t = \partial \ln f_t(y_t; \lambda_{t|t-1}, \upsilon, \eta)/\partial \lambda_{t|t-1} = (\eta + 1)\, b_t - 1, \tag{20}$$

where b_t is as in (13), but with scale dynamic, rather than location. As $|y_t| \to \infty$, $u_t \to \eta$, so the score is bounded for finite η. This reflects a general result that in a location-scale model with a fat-tailed distribution, the score for location is re-descending whereas the score for scale is not.

The fact that b_t is distributed as beta $(1/\upsilon, \eta/\upsilon)$ enables exact expressions for the moments and autocorrelations of $|y_t|^c$, $-1 < c < \eta$, to be found and the information matrix to be constructed. Much of the theory can be further extended to handle skewness and asymmetry. The advantage of the generalized t distribution is that it has a sharp peak when $\upsilon < 2$ and this turns out to be characteristic of many series of returns; for example Harvey and Lange (2017) estimate υ to be 1.34 for silver returns.

4.2 Asymmetric impact curves (leverage)

The response of volatility to a change in asset price, y_t, is often asymmetric. This asymmetry, or leverage, can be captured in an EGARCH model by the modification

$$\lambda_{t+1|t} = \omega\,(1 - \phi) + \phi\,\lambda_{t|t-1} + \kappa\, u_t + \kappa^* u_t^*, \tag{21}$$

where u_t is the scale score of (20), u_t^* =sgn$(-\varepsilon_t)(u_t + 1)$ and κ^* is a new parameter which, because the negative of the sign of the return is taken, is usually expected to be positive. When the distribution of ε_t is symmetric, u_t^* has zero mean and $E(u_t u_t^*) = 0$. The information matrix for a Beta-t-EGARCH model with dynamics as in (21) is given in Harvey (2013, pp.

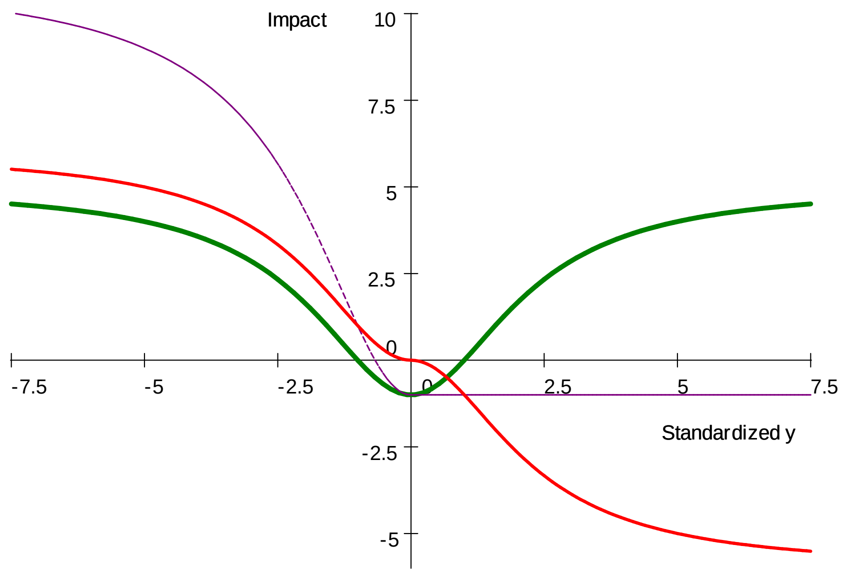

Figure 2: Impact curve for t_5 against a standardized y. Thick line is symmetric, thin line has $\kappa = \kappa^*$ and medium line is anti-symmetric.

121-4). Identifiability requires only that either κ or κ^* be non-zero, so Wald and likelihood-ratio tests of the null hypothesis that one of them is zero can be carried out.

Figure 2 shows the impact curves, $\kappa\, u_t + \kappa^* u_t^*$, for the Beta-t-EGARCH model for a t_5 distribution. The curves, which are plotted against the standardized variable, y_t, range from the symmetric, in which $\kappa = 1$ and $\kappa^* = 0$, to the anti-symmetric in which $\kappa = 0$ and $\kappa^* = 1$. In the intermediate case, when $\kappa = \kappa^* = 1$, positive values of y_t have no effect on volatility.

The standard way of incorporating leverage effects into GARCH models is to include a variable in which the squared observations are multiplied by an indicator, $I(y_t < 0)$, taking the value one for $y_t < 0$ and zero otherwise; see Glosten, Jagannathan and Runkle (1993, p 1788). This model is unable to allow for the asymmetric response in Figure 2. The sign is not used because it could give a negative conditional variance.

4.3 Components and long memory

Long memory in scale may be modelled by a fractionally integrated process. For example, Janus et al (2014) fit the ARFIMA score-driven model

$$(1-L)^d(\lambda_{t+1|t}-\omega)=\phi(1-L)^d\,(\lambda_{t|t-1}-\omega)+\kappa\,u_t,$$

where L is the lag operator, to four stocks and find values of d between 0.43 and 0.75. A more appealing approach is to fit two components. Thus with leverage included,

$$\begin{aligned}\lambda_{t|t-1} &= \omega+\lambda_{1,t|t-1}+\lambda_{2,t|t-1}, & t&=1,...,T,\\ \lambda_{i,t+1|t} &= \phi_i\,\lambda_{i,t|t-1}+\kappa_i\,u_t+\kappa_i^*\,u_t^*, & i&=1,2,\end{aligned}\tag{22}$$

where $\phi_1 > \phi_2$ if $\lambda_{1,t|t-1}$ denotes the long-run component. Identifiability requires $\phi_1 \neq \phi_2$, which is implicitly imposed by setting $\phi_1 > \phi_2$, together with $\kappa_1 \neq 0$ or $\kappa_1^* \neq 0$ and $\kappa_2 \neq 0$ or $\kappa_2^* \neq 0$.

A two-component model allows different asymmetric effects in the short-run and the long-run. There seems to be a growing body of evidence suggesting that an asymmetric response is confined to the short-run volatility component. Indeed short-run volatility may even decrease after a good day, as in Figure 2, because it calms the market.

4.4 ARCH in Mean

The EGARCH-M model is a modification of the ARCH in mean model of Engle et al (1987) in which a time-varying risk premium, $\alpha\exp(\lambda_{t|t-1})$, is added to the right-hand side of (18). A two-component model not only deals with differing leverage effects in the long and short run, but also makes it possible to separate out the effects of long-run and short-run movements in volatility on the mean. Thus the model generalizes to

$$y_t = \mu + \alpha_1 \exp(\omega + \lambda_{1,t|t-1}) + \alpha_2[\exp(\lambda_{2,t|t-1}) - 1] + \varepsilon_t \exp(\lambda_{t|t-1}), \quad (23)$$

where μ, α_1, ω and α_2 are parameters. The equity risk premium is then captured by the long-run component, with an equilibrium level of $\mu + \alpha_1 \exp \omega$. Harvey and Lange (2018) demonstrate that a two-component score-driven model with symmetric long-run volatility, that is $\kappa_1^* = 0$, coupled with anti-symmetric short-run volatility, that is $\kappa_2 = 0$, provides a good fit and yields a plausible interpretation of market behaviour. This accords with the conclusion of Adrian and Rosenberg (2008, p 3015), in that the short-run component appears to capture shocks to market skewness, whereas the long-run component is related to business cycle risk.

5 Location/Scale

In the location/scale model, the structure is as for EGARCH, that is

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1}), \quad y_t \geq 0, \quad t = 1, ..., T. \quad (24)$$

With the GB2 parameterization of (8), $\varphi = \exp(\lambda_{t|t-1})$ and

$$\frac{\partial \ln f_t(y_t; \lambda_{t|t-1}, \varphi, \upsilon, \xi, \eta)}{\partial \lambda_{t|t-1}} = u_t = (\upsilon\xi + \eta) b_t(\xi, \eta) - \upsilon\xi, \quad (25)$$

where

$$b_t(\xi, \eta) = \frac{(y_t e^{-\lambda_{t|t-1}})^{\upsilon}/\eta}{(y_t e^{-\lambda_{t|t-1}})^{\upsilon}/\eta + 1}, \qquad t = 1, ..., T,$$

is distributed as beta$(\xi, \eta/\upsilon)$. As $y \to \infty$, the score approaches an upper bound of η. The corresponding score for the generalized gamma distribution is $(y_t e^{-\lambda_{t|t-1}})^{\upsilon} - \upsilon\xi$, with $(y_t e^{-\lambda_{t|t-1}})^{\upsilon}$ distributed as a gamma variate with unit scale and shape parameter ξ.

Special cases of the GB2 distribution have been used in finance to model time series on the range of daily stock prices and the daily realized variance (or volatility); see, for example, Harvey (2013, ch 5) and Opschoor et al (2016). Realized variance may exhibit long memory and leverage effects, as well as having fat tails. A GB2 location/scale model, (24), can capture these characteristics by employing two components, as in (22), with the leverage determined by the sign of (demeaned) returns, r_t, that is $u_t^* = \mathrm{sgn}(-r_t)(u_t + \upsilon\xi)$.

When the GB2 is parametrized as in (7) taking logarithms gives the location-scale model

$$\ln y_t = \lambda_{t|t-1} + \ln \varepsilon_t, \qquad t = 1, ..., T,$$

with $\ln \varepsilon_t$ having an EGB2 distribution. The fact that the EGB2 distribution tends to a normal as $\xi, \varsigma \to \infty$ shows the link with unobserved component models such as the one used by Alizadeh et al (2002) for modeling the logarithm of the intra-day range in the logarithm of an asset price or exchange rate.

6 Count data and qualitative observations

Time series models for count data and qualitative observations need to take account of the nature of the data in constructing dynamic equations. Despite the lack of a general asymptotic theory for maximum likelihood estimation, such evidence as there is lends support to the dynamics being driven by the standardized score.

6.1 Count data

The probability mass function of the Poisson distribution is

$$p(y_t;\mu) = \mu^{y_t} e^{-\mu}/y_t! \qquad \mu > 0, \quad y_t = 0,1,2,...., \tag{26}$$

where the parameter μ is both mean and variance. When the mean changes over time, an exponential link function, $\mu_{t|t-1} = \exp\theta_{t|t-1}$, ensures that it remains positive even though $\theta_{t|t-1}$ is unconstrained. The conditional score of $\theta_{t|t-1}$ is $y_t - \exp\theta_{t|t-1}$, which, when divided by the information quantity, gives $u_t = y_t \exp(-\theta_{t|t-1}) - 1 = y_t/\mu_{t|t-1} - 1$.

The negative binomial distribution allows for overdispersion. It is convenient to parameterize it in terms of the mean so the probability mass function for the dynamic model is

$$p(y_t;\mu_{t|t-1},\upsilon) = \frac{\Gamma(\upsilon+y_t)}{y_t!\Gamma(\upsilon)}\mu_{t|t-1}^{y_t}(\upsilon+\mu_{t|t-1})^{-y_t}(1+\mu_{t|t-1}/\upsilon)^{-\upsilon}, \quad y_t = 0,1,2,....,$$

where $\upsilon > 0$; the Poisson distribution is obtained by letting $\upsilon \to \infty$. With the exponential link function, dividing the score for $\theta_{t|t-1}$ by the information quantity gives $u_t = y_t/\mu_{t|t-1} - 1$, just as for the Poisson distribution.

Zucchini et al (2016) give weekly data on firearm homicides in Capetown over the period 1986 to 1991. Models were fitted to the first 305 observations, assuming a random walk dynamic equation,

$$\mu_{t+1|t} = \mu_{t|t-1} + \kappa u_t, \quad t = 1,..,T,$$

with $\mu_{1|0}$ estimated as a fixed parameter. The negative binomial gave a log-likelihood of -589.15, as opposed to -610.41 for the Poisson distribution. The estimates were $\widetilde{\kappa} = 0.097$ and $\widetilde{\upsilon} = 4.137$. The (predictive) filtered estimates, $\mu_{t\imath t-1}$, and multi-step forecasts, computed using the TSL package of Lit et al (2020), are shown in Figure 3 in the shaded area. As can be seen, the forecasts have adapted to the higher level towards the end of the sample.

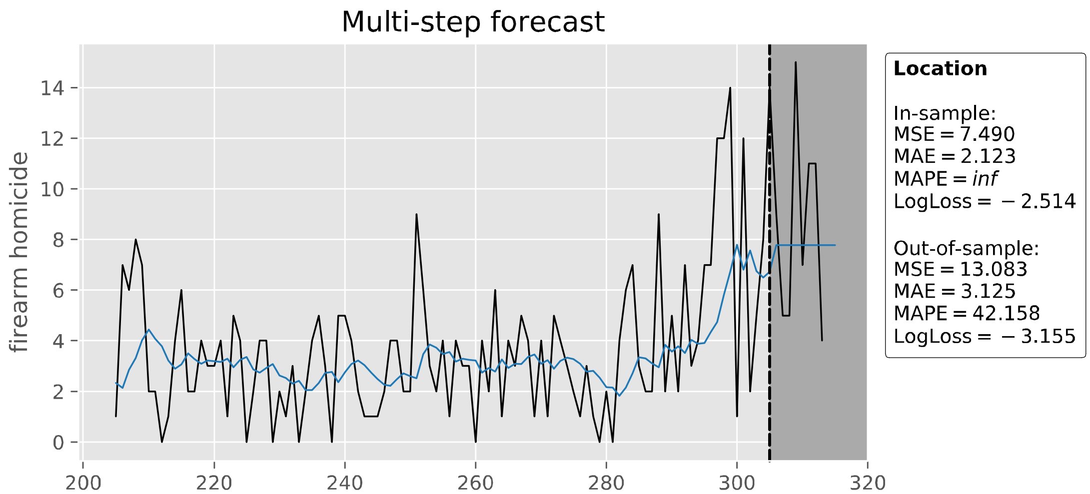

Figure 3: Weekly firearm homicides in Capetown, together with predictive filter; weekly data 1986-91. Out-of-sample period is shaded.

Harvey and Kattuman (2020) describe the fitting of a score-driven negative binomial model to data on deaths from coronavirus. The application is described in more detail in Lit et al (2020).

The Skellam distribution is used to model the difference between two counts. Koopman and Lit (2019) set up a score-driven model to predict the goal difference in football matches.

6.2 Binomial, multinomial and ordered categorical data

In the binomial model, where the probability that $y_t = 1$ is π and the probability that $y_t = 0$ is $1 - \pi$, the usual link function is the logistic. Thus with time-variation

$$\pi_{t|t-1} = 1/(1 + \exp(-\theta_{t|t-1})), \qquad t = 1, ..., T.$$

The score divided by the information quantity is

$$u_t = \frac{y_t - \pi_{t|t-1}}{\pi_{t|t-1}(1 - \pi_{t|t-1})}.$$

Lit et al (2020) apply the model to the annual Oxford-Cambridge boat race.

Multinomial observations can be handled by extending the binomial model; the general logistic transformation described in Catania (2019) could be used. Ordered categorical data require a different treatment. The observations are defined in terms of intervals on a continuous distribution for a variable, x_t. The probability of being in a given interval is obtained from the CDF of x_t and these probabilities define the probability mass function of the discrete distribution of y_t. Koopman and Lit (2019) model the results of football matches using the ordered categorical variables win, draw and lose.

7 Multivariate models

This section describes dynamic models for a set of N variables in a vector $\mathbf{y}_t$ under the simplifying assumption that they have zero mean. The emphasis is on changing correlation and association.

7.1 Multivariate scale and dynamic correlation

The dynamics in a general GARCH model depend on the elements of the filtered covariance matrix, $\mathbf{V}_{t|t-1}$, for a multivariate$-t$ or Gaussian distribution. As such $\mathbf{V}_{t|t-1}$ contains a large number of parameters and it is not clear how best to impose restrictions. Furthermore there is no guarantee that $\mathbf{V}_{t|t-1}$ will be positive definite at all points in time. A better way forward is to follow Creal et al (2011) and work with a scale matrix, $\mathbf{\Omega}_{t|t-1}$, that allows volatility and correlation parameters to be separated by the decomposition

$$\mathbf{\Omega}_{t|t-1} = \mathbf{D}_{t|t-1}\mathbf{R}_{t|t-1}\mathbf{D}_{t|t-1}, \tag{27}$$

where $\mathbf{D}_{t|t-1}$ is diagonal and $\mathbf{R}_{t|t-1}$ is a positive-definite correlation matrix with diagonal elements equal to unity. An exponential link function may be used for the volatilities in $\mathbf{D}_{t|t-1}$. Joint modeling of dynamic scale and correlation can be based on a dynamic equation of the form

$$\boldsymbol{\theta}_{t+1|t}= (\mathbf{I}-\boldsymbol{\Phi})\boldsymbol{\omega}+\boldsymbol{\Phi}\boldsymbol{\theta}_{t|t-1}+\mathbf{K}\mathbf{u}_t,$$

where $\boldsymbol{\theta}_{t|t-1}= (\boldsymbol{\lambda}'_{t|t-1}, \boldsymbol{\gamma}'_{t|t-1})'$, with the $N(N-1)/2$ vector $\boldsymbol{\gamma}_{t|t-1}$ determining $\mathbf{R}_{t|t-1}$ and $\boldsymbol{\lambda}_{t|t-1}$ modeling the EGARCH effects. The parameters are contained in the $\boldsymbol{\Phi}$ and $\mathbf{K}$ matrices and the $\boldsymbol{\omega}$ vector; these are typically restricted.

The attraction of implementing the score-driven approach in this way becomes clear when modeling changing correlation. Consider the simple set-up of a bivariate model with a conditional Gaussian distribution and let the variances be time-invariant. Dividing the observations by their standard deviations gives variables x_{1t} and x_{2t}. It might be thought that the product of x_{1t} and x_{2t} provides the information needed to drive the dynamics of correlation, but this turns out not to be the case. In order to keep the correlation coefficient, $\rho_{t|t-1}$, in the range, $-1 \leq \rho_{t|t-1} \leq 1$, the link function

$$\rho_{t|t-1} = \tanh(\gamma_{t|t-1}) = (\exp(2\gamma_{t|t-1}) - 1)/(\exp(2\gamma_{t|t-1}) + 1) \qquad (28)$$

may be used. The dynamic equation for the unconstrained variable $\gamma_{t|t-1}$ depends on the score, which, when written in terms of $\rho_{t|t-1}$, is

$$u_{\gamma t} = (1-\rho^2_{t|t-1})^{-1}(x_{1t}-\rho_{t|t-1}x_{2t})(x_{2t}-\rho_{t|t-1}x_{1t}) - \rho_{t|t-1}.$$

The score only reduces to $x_{1t}x_{2t}$ when $\rho_{t|t-1} = 0$. On the other hand, when $\rho_{t|t-1}$ is close to one, the weight given to $(x_{1t}+x_{2t})^2$ is small and the second term dominates. As a result $u_{\gamma t}$ is negative and so $\rho_{t+1|t}$ falls unless x_{1t} and x_{2t} are close; see Figure 4 and the discussion in Creal et al (2011) and Harvey

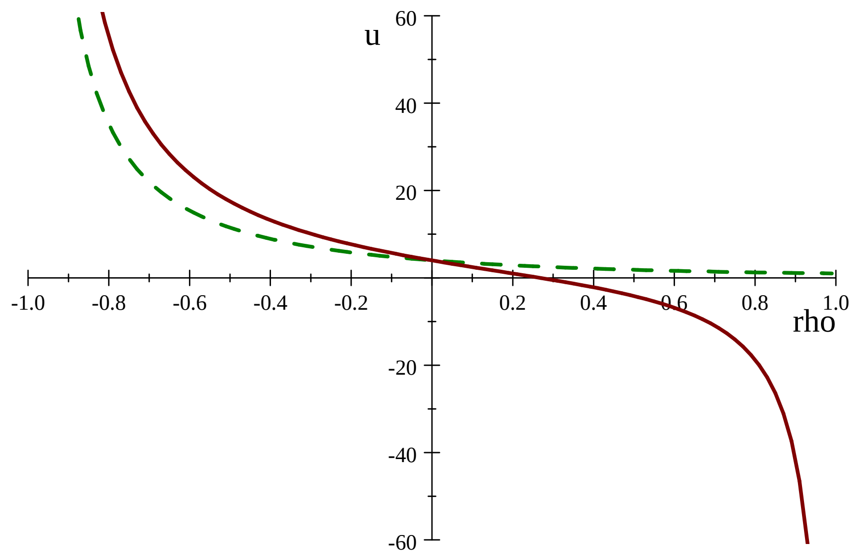

Figure 4: Plot of score, u, against correlation, rho, for bivariate Gaussian distribution with $x_1 = x_2 = 2$ (dash) and $x_1 = 4, x_2 = 1$

(2013, ch 7).

The scores may be computed under the null hypothesis of constant correlation $\rho_{t|t-1} = r$, where r is the maximum likelihood estimator of ρ, and used in a portmanteau test. When $r = 0$, moment-based tests are obtained because $u_t = x_{1t}x_{2t}$, but when $r \neq 0$ the score-based tests can be much more powerful; see Harvey and Thiele (2016). The tests can be modified for a bivariate t-distribution with estimated EGARCH models.

7.2 Dynamic Copulas

A copula models the association between two variables independently of their marginal distributions. It is a joint distribution function of standard uniform random variables, that is

$$C(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2), \quad 0 \leq u_1, u_2 \leq 1.$$

As such it provides a comprehensive measure of dependence. The upper and lower coefficients of tail dependence are often of special interest in the context of risk; see McNeil et al (2005). When two variables have a bivariate normal distribution, they are asymptotically independent in the tails because the coefficients of tail dependence are both zero (unless $\rho = 1$). On the other hand, a t-copula does exhibit tail dependence.

Time-varying copulas are best modeled using the conditional score to drive a dynamic equation for the shape parameter; see Patton (2013, pp 931-2). The viability of this approach was first explored by Creal et al (2011) in an application of dynamic Gaussian copulas to exchange rate data. Expressions for the conditional score can be quite elaborate. However, a graph of the score can show that, once obtained, it has a natural and intuitive interpretation. The score for a Clayton copula in Harvey (2013, p 229) provides an illustration.

Janus et al (2014) use the t-copula with t-marginals, thereby allowing the degrees of freedom to be different in the marginal distributions as well as in the joint distribution. More applications can be found in De Lira Salvatierra and Patton (2015), Oh and Patton (2017), Lucas et al (2017) and Bernardi and Catania (2019).

7.3 Spatial correlation, count data and location/scale models

A number of other models for different aspects of multivariate time series have been proposed.

7.3.1 Spatial correlation

Blasques et al (2016) set up a dynamic model for spatial autocorrelation as

$$\mathbf{y}_t = \rho_{t|t-1}\mathbf{W}\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}), \quad t = 1, ..., T.$$

where $\mathbf{W}$ is the spatial weight matrix. The time-varying correlation, $\rho_{t|t-1}$, is a scalar kept in the range by the transformation of (28), that is $\rho_{t|t-1} = \tanh(\gamma_{t|t-1})$. The score with respect to $\gamma_{t|t-1}$ is

$$u_t = [\mathbf{y}'\mathbf{W}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \rho_{t|t-1}\mathbf{W}\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}) - tr\{(\mathbf{I} - \rho_{t|t-1}\mathbf{W})^{-1}\mathbf{W}\}](1 - \rho_{t|t-1}^2).$$

See also Billé and Catania (2019).

7.3.2 Bivariate Poisson distribution

The bivariate Poisson distribution can be fitted to two series of count data. The parameterization is similar to that of the football example discussed for the Skellam distribution. Koopman and Lit (2019) found the forecasting performance of the model to be at least as good as that of the corresponding parameter-driven model, but with only a fraction of the estimation time.

7.3.3 Dynamic location/scale model

Realized covariance can be measured in a similar way to realized variance, leading to the construction of $N \times N$ realized volatility covariance matrices, $\mathbf{Y}_t$, $t = 1, ..., T$. Opschoor et al (2016) propose a location/scale model that uses a multivariate F distribution. The PDF is

$$f(\mathbf{Y}_t \mid \boldsymbol{\Omega}_{t|t-1}, \nu_1, \nu_2) = K(\nu_1, \nu_2)\frac{\left|\boldsymbol{\Omega}_{t|t-1}\right|^{-\nu_1/2}\left|\mathbf{Y}_t\right|^{(\nu_1-N-1)/2}}{\left|\mathbf{I} + \boldsymbol{\Omega}_{t|t-1}^{-1}\mathbf{Y}_t\right|^{(\nu_1+\nu_2)/2}}, \qquad \nu_1, \nu_2 > N-1,$$

where $\boldsymbol{\Omega}_{t|t-1} = (\nu_2 - N - 1)/\nu_1)\mathbf{V}_{t|t-1}$ is a scale matrix, such that $\mathbf{V}_{t|t-1} = E(\mathbf{Y}_t)$ for $\nu_1, \nu_2 > N-1$, and $K(\nu_1, \nu_2) = \Gamma_N((\nu_1+\nu_2)/2)\Gamma_N(\nu_1/2)\Gamma_N(\nu_2/2)$, where $\Gamma_N(.)$ is the multivariate gamma function. When $\nu_2 \to \infty$, the distribution becomes a Wishart distribution, which is the multivariate generalization of the chi–squared distribution; see Gorgi et al (2019). A single entry on the diagonal of $\mathbf{Y}_t$, that is $y_{ii,t}$, $i = 1, ..., N$, is distributed as $F(\nu_1, \nu_2 - N - 2)$.

When $\nu_2 \to \infty$, the distribution becomes Wishart, the multivariate generalization of the chi–squared distribution. Opschoor et al (2016) model the dynamics of the covariance matrix directly with the filter

$$\mathbf{V}_{t+1\mid t} = \mathbf{V} + \phi \mathbf{V}_{t|t-1} + \kappa \mathbf{U}_t, \qquad t = 1, ..., T,$$

where $\mathbf{U}_t$ is a score matrix for $\mathbf{V}_{t|t-1}$. An alternative would be to decompose $\mathbf{\Omega}_{t|t-1}$ as in (27).

8 Extensions

The score provides a solution to constructing viable dynamic models in non-standard situations. Some examples are set out below.

8.1 Censoring and dynamic Tobit models

Censoring takes place when a variable above or below a certain value is set equal to that value. When location changes over time the challenge is how to formulate a dynamic Tobit model. A number of researchers, beginning with Zeger and Brookmayer (1986), have addressed this problem when the underlying (uncensored) observations are Gaussian. However, there is no computational disadvantage to adopting other, more flexible, distributions. It is in this spirit that Lewis and McDonald (2014) propose the use of generalized$-t$ and EGB2 distributions for censored static regression and these distributions may be similarly employed for dynamic Tobit models. The score-driven model automatically solves the problem of how to weight the censored observations in the dynamic location equation.

Let x_t be a variable for which the observations are subject to censoring from below, that is

$$y_t = \begin{cases} x_t, & x_t > c, \\ c, & x_t \leq c, \end{cases} \qquad -\infty < x_t < \infty. \tag{29}$$

The lower bound, c, is usually known. Then $\Pr(y_t = c) = \Pr(x_t \leq c\ ) = F_x(c)$, where F_x is the CDF of x_t. Let $I(y_t > c)$ be an indicator that is zero when $y_t = c$ and one when $y_t > c$. The distribution of y_t is a discrete-continuous mixture, with a point mass at c, so

$$\ln f(y_t; c) = (1 - I(y_t > c)) \ln F_x(c) + I(y_t > c) \ln f_x(y_t). \tag{30}$$

The score with respect to a changing location is therefore

$$\frac{\partial \ln f(y_t; c)}{\partial \mu_{t|t-1}} = (1 - I(y_t > 0)) \frac{\partial \ln F_x(c)}{\partial \mu} + I(y_t > 0) \frac{\partial \ln f_x(y_t)}{\partial \mu}. \tag{31}$$

The logistic distribution, which is a special case of the EGB2 distribution, has a shape close to that of the normal but with slightly heavier tails. The score, (31), is $I(y_t > c)e^{-\lambda}b_t - e^{-\lambda}(1 - b_t)$, where b_t is as defined in (13). The fact that the CDF of a logistic distribution has a simple closed form makes it an attractive choice, and it has an additional robustness benefit because, in the absence of censoring, as when y_t is positive in (31), the score implies Winsorizing.

Dynamic volatility can also be modeled when the observations are censored. Harvey and Liao (2019) illustrate the viability of the method with data on Chinese stock returns that are subject to an upper limit on the daily change.

Harvey and Ito (2019) use similar techniques for modeling time series with a variable that is continuous, except for a significant number of zeroes. They do this by shifting a continuous location/scale distribution to the left and censoring all the negative observations so that they are assigned a value of zero.

8.2 Circular data

Observations on direction are circular. When circular observations are recorded in radians, they are usually assumed to have a von Mises density

$$f(y_t;\mu,\upsilon)=\frac{1}{2\pi I_0(\upsilon)}\exp\{\upsilon\cos(y_t-\mu)\},\qquad -\pi<y_t,\mu\leq\pi,\quad \upsilon\geq 0,\tag{32}$$

where $I_k(\upsilon)$ denotes a modified Bessel function of order k, μ denotes location (directional mean) and υ is a non-negative concentration parameter that is inversely related to scale. When $\upsilon=0$ the distribution is uniform, whereas y_t is approximately $N(\mu,1/\upsilon)$ for large υ.

Data generated by a time series model over the real line, that is $-\infty<z_t<\infty$, can be converted into wrapped circular time series observations in the range $[-\pi,\pi)$ by letting $y_t=z_t \bmod(2\pi)-\pi$, $t=1,...,T$, as in Breckling (1989). The score-driven model for directional data is

$$z_t=\mu_{t|t-1}+\varepsilon_t,\qquad t=1,....,T,\tag{33}$$

where the $\varepsilon_t's$ are IID random variables from a standardized circular distribution with location zero and the forcing variable, u_t, in the dynamic equation for $\mu_{t|t-1}$ is proportional to the conditional score. A key property of a (continuous) circular distribution is that it satisfies the periodicity condition $f(y\pm 2\pi k;\theta)=f(y;\theta)$, where k is an integer and θ denotes parameters. Provided the derivatives of the log-density with respect to the elements of θ are continuous, they too are circular in that the periodicity condition is satisfied. The conditional distribution of the wrapped observations, y_t, is therefore the same as that of z_t in (33) and so the likelihood function is the same. The problem of estimating a wrapped model, as posed by Breckling (1989), is therefore solved and the resulting class of models has considerable advantages over those currently in use; see Fisher and Lee (1994).

When ε_t has a von Mises distribution with $\mu = 0$, the score is $u_t = \upsilon \sin(z_t - \mu_{t|t-1}) = \upsilon \sin(y_t - \mu_{t|t-1})$. For first-order dynamics, Harvey et al (2019) derive the asymptotic distribution of the maximum likelihood estimator.

8.3 Switching regimes and dynamic adaptive mixture models

The dynamic adaptive mixture model (DAMM) of Catania (2019) has the probability of being in a given regime changing over time. The dynamics are modeled using the scores of the regime probabilities in the conditional distribution. The model may be extended so that the locations and/or scales in each of the regimes contained in the mixture are also dynamic. Again the conditional scores are used, thus providing a unified approach based on well-established principles. The scores for locations and scale, like the scores for the regime probabilities, have a natural and intuitive interpretation.

The DAMM is designed for situations similar to those addressed by the textbook regime-switching model of Hamilton (1989). That model introduces dynamics by a Markov chain in which there is a fixed probability of staying in the current regime or moving to another. The regime is not observed: hence the term hidden Markov chain, as in Zucchini et al (2016). However, unusually for a nonlinear parameter-driven model, the probability of being in a particular regime is ultimately given by a filter that depends on past observations, just as the Kalman filter is a function of past observations in a linear model. These probabilities yield a conditional distribution for the current observation, as in the DAMM, but with the difference that the DAMM is formulated as observation-driven at the outset.

The conditional distribution in a two-state DAMM is

$$f_{t|t-1}(y_t) = \xi_{t|t-1} f_{1,t|t-1}(y_t) + (1 - \xi_{t|t-1}) f_{2,t|t-1}(y_t), \qquad t = 1, ..., T, \quad (34)$$

where $\xi_{t|t-1}$ is the probability of being in state one at time t, based on information available up to and including time $t-1$. A logistic link function

$$\xi_{t|t-1} = \frac{\exp(\gamma_{t|t-1})}{1+\exp(\gamma_{t|t-1})}, \qquad -\infty < \gamma_{t|t-1} < \infty, \tag{35}$$

confines $\xi_{t|t-1}$ to the range $0 < \xi_{t|t-1} < 1$. The score with respect to $\gamma_{t|t-1}$, but written in terms of $\xi_{t|t-,1}$, is

$$u_t = \frac{\partial \ln f_{t|t-1}}{\partial \gamma_{t|t-1}} = \frac{f_{1,t|t-1} + (1-\xi_{t|t-1})f_{2,t|t-1}}{f_{t|t-1}} \xi_{t|t-1}(1-\xi_{t|t-1}) \tag{36}$$

and this drives a dynamic equation. The filter in the Markov chain switching model depends on similar variables to those in u_t.

The score for a dynamic parameter within a regime of a DAMM is

$$\frac{\partial \ln f_{t|t-1}}{\partial \theta_i} = \frac{\partial \ln f_{t|t-1}}{\partial f_{it}} \frac{\partial f_{i,t|t-1}}{\partial \theta_i} = \xi_{i,t|t-1} \frac{f_{it}}{f_{t|t-1}} \frac{\partial \ln f_{it}}{\partial \theta_i} = \xi_{i,t} \frac{\partial \ln f_{i,t|t-1}}{\partial \theta_i}, \quad i = 1, 2,$$

where $\xi_{1,t|t-1} = \xi_{t|t-1}$ and $\xi_{2,t|t-1} = 1 - \xi_{t|t-1}$. When $\xi_{i,t}$, the probability of being in a given regime, is small, the contribution of the observation to the score is downweighted; there is no such weighting in Markov-switching models.

The DAMM can be combined with other score-driven models. For example, Harvey and Palumbo (2021) set up a bivariate model for wind speed and direction with EGARCH effects. The aim is to capture the switching between two prevailing winds at a site in North-West Spain. The filtered probability, $\xi_{t|t-1}$, of being in the higher state, that is around four radians, is shown in Figure 5. The data lie between 0 and 2π radians, with the circularity meaning that observations near the top of the graph are close to those at the bottom.

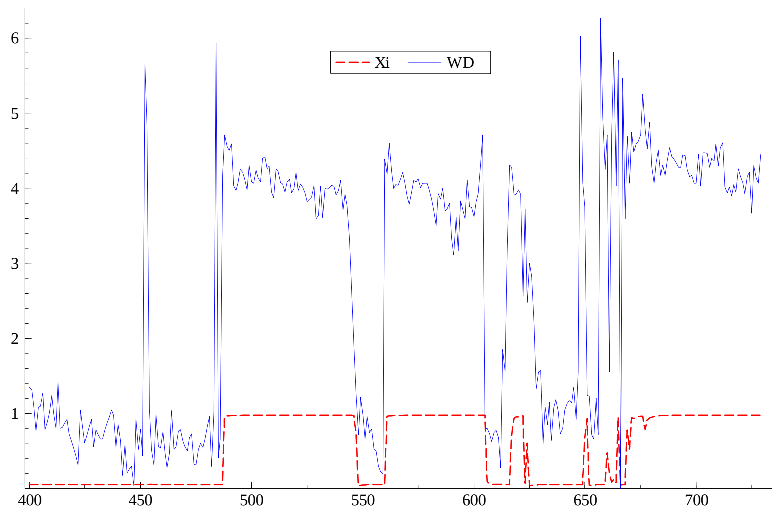

Figure 5: Filtered regime probability for wind direction from a two-regime heteroscedastic von Mises model. Vertical axis is direction in radians and regime probability from 0 to 1. The horizontal axis is for hours at the end of January 2004.

8.4 Dynamic shape parameters, adaptive models and missing observations

8.4.1 Changing degrees of freedom

Dynamic models for shape parameters may be formulated using the score-driven approach. For example the degrees of freedom, ν, in a t-distribution may change over time. The score for $\overline{\nu} = -\ln \nu$ is

$$\frac{\partial \ln f_t}{\partial \overline{\nu}} = \frac{\nu}{2}\psi\left(\nu/2\right) - \frac{\nu}{2}\psi\left((\nu+1)/2\right) + \frac{1}{2} - \frac{\nu+1}{2}b_t - \frac{\nu}{2}\ln(1-b_t), \quad (37)$$

where b_t is as in the score for scale, (20), and $\psi\left(.\right)$ is the digamma function. Figure 6 plots this score against the standardized y for $\nu = 5$. As y moves towards the tails, $\overline{\nu}$ increases and so the degrees of freedom falls. This behaviour contrasts with that of the score for the logarithm of scale, λ, which is bounded as $y \to \pm\infty$. It is interesting that the sum of the third and fourth terms in (37) is equal to the score of λ multiplied by $-1/2$. As in other instances, the behaviour of the score makes perfect sense. The only difficulty in implementing shape parameter filters like this one is that a large sample is needed to obtain reliable estimates. Most of the applications so far have been for financial time series with several thousand observations.

8.4.2 Adaptive models

Nonstationary time series are sometimes subject to sudden upward or downward shifts. A score-driven filter will adapt to such breaks. However, it may be possible to speed up the adjustment by introducing a second layer of dynamics into the model. Thus in the location model, (9), κ becomes $\kappa_{t+1|t}$ and this evolves according to a dynamic equation in which the forcing variable is the conditional score $\partial \ln f_t/\partial\kappa_{t|t-1} = u_t u_{t-1}$; see Blasques et al (2019). Adaptive models are also discussed in Delle Monache and Petrella (2017).

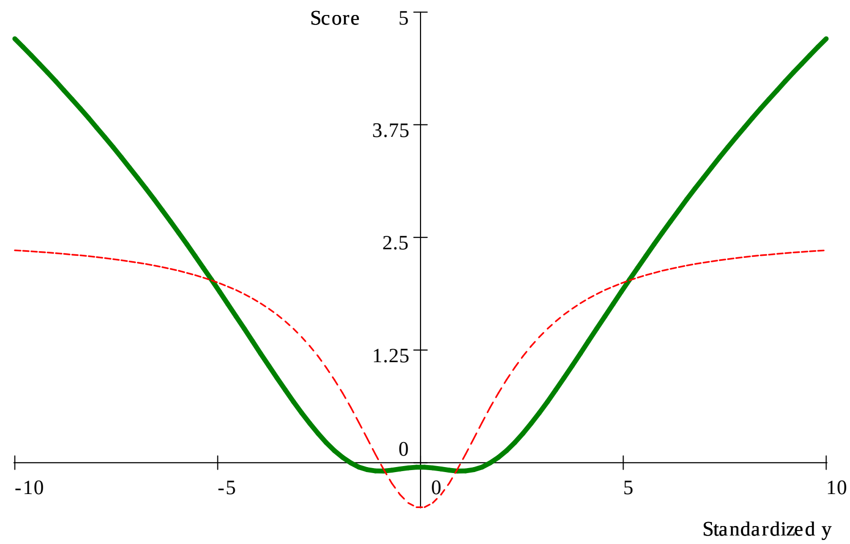

Figure 6: Score for $\overline{\nu}$ (bold) and score for logarithm of scale, λ, against standardized y.

8.4.3 Missing observations

A practical way of dealing with a missing observation is to set $u_t = 0$ and to drop that time period from the likelihood function. Thus the filter makes no adjustment for the increased variance, as in the linear Gaussian model where the solution is exact. Furthermore the conditional distribution is assumed to be the same as that of the one-step ahead conditional distribution. Blasques, Gorgi and Koopman (2021) provide a more theoretically sound solution to the problem of missing observations by using indirect inference.

9 Beyond the score

The function to be maximized need not be a likelihood. For example it may be a sum of squares or absolute values, a quasi-likelihood or a robust function, such as an M-estimator as in Maronna et al (2006). Dynamic quantiles and kernels may be obtained in this way and the dynamic equations for them may be constructied by a natural extension to the score-driven framework.

A sample quantile, $\widetilde{\xi}(\tau)$, $0 < \tau < 1$, can be obtained as the solution to the minimization of

$$S_\tau(\xi) = \sum_{t=1}^{T} \rho_\tau(y_t - \xi) = \sum_{y_t < \xi} (\tau - 1)(y_t - \xi) + \sum_{y_t \geq \xi} \tau(y_t - \xi), \qquad 0 < \tau < 1, \quad (38)$$

with respect to $\xi = \xi(\tau)$, where $\rho_\tau(.)$ is the *check function*. The derivative of this criterion function is the quantile indicator function

$$\mathrm{IQ}(y_t - \xi_t(\tau)) = \begin{cases} \tau - 1, & y_t < \xi_t(\tau), \\ \tau, & y_t > \xi_t(\tau), \end{cases} \qquad t = 1, ..., T, \qquad (39)$$

where IQ(0) is not determined, but can be set to zero; see De Rossi and Harvey (2009). This indicator provides the forcing variable, $u_t(\tau)$, in the quantile filter

$$\xi_{t+1|t}(\tau) = \phi \xi_{t|t-1}(\tau) + \kappa u_t(\tau). \qquad (40)$$

The quantile indicator plays a similar role to that of the conditional score. Indeed it is the score for an asymmetric Laplace distribution.

The filter in (40) belongs to the class of CAViaR models proposed by Engle and Manganelli (2004) in the context of tracking value at risk (VaR). In CAViaR, the filter is driven by a function of y_t, but includes an adaptive model, which in a limiting case has the same form as (40). Other CAViaR specifications, which are based on actual values, rather than indicators, are not only inconsistent with the quantile framework but may suffer from a lack of robustness to outliers.

Patton et al (2019) show that it is possible to set up a joint dynamic model for VaR, that is $\xi_{t|t-1}(\tau)$, and expected shortfall, $E_{t-1}(y_t \mid \xi_{t|t-1}(\tau))$. The forcing variables depend on conditional quantiles and expectiles.

Harvey and Oryshchenko (2011) construct a dynamic kernel estimator for the PDF of a time series. The criterion function for the observation at time t, is

$$\rho(y_t|f_{t|t-1}(y)) = -\frac{1}{2}\left[\frac{1}{h}K\left(\frac{y_t - y}{h}\right) - f_{t|t-1}(y)\right]^2, \qquad -\infty < y, y_t < \infty, \quad t = 1, ..., T,$$

where $K(.)$ is a kernel, and differentiating with respect to $f_{t|t-1}(y)$ gives the forcing variable

$$u_t(f_{t|t-1}(y)) = \frac{1}{h}K\left(\frac{y_t - y}{h}\right) - f_{t|t-1}(y), \qquad t = 1, ..., T. \tag{41}$$

The updating filter in the basic case is then

$$f_{t+1|t}(y) = (1-\phi)\overline{f}_T(y) + \phi f_{t|t-1}(y) + \kappa u_t, \quad t = 1, .., T,$$

where $u_t = u_t(f_{t|t-1}(y))$. The application in Harvey and Oryshchenko (2011) has ϕ set to one.

10 Conclusion

Modelling the dynamics in nonlinear time series by the score of the conditional distribution provides a comprehensive and unified solution to a range of problems. Estimation is by maximum likelihood and is usually straightforward. Tests, including diagnostics based on the Lagrange multiplier approach, can be formulated.

It might be thought that assuming a particular parametric distribution makes the resulting filter vulnerable to misspecification. On the contrary, basing a model on a heavy-tailed distribution makes it far more robust than methods, such as quasi-maximum likelihood, that are usually motivated by analogies with Gaussian models.

Acknowledgements

I am grateful to Francisco Blasques, Leopoldo Catania, Christopher Haffner, Jonas Knecht, Siem Jan Koopman, Rutger-Jan Lange, Rutger Lit and an editor for helpful comments and suggestions; of course they bear no responsibility for opinions expressed or mistakes made.

REFERENCES

Adrian T, Rosenberg J 2008. Stock returns and volatility: pricing short-run and long-run components of market risk. *Journal of Finance* 63: 2997-3030

Alizadeh S, Brandt MW, Diebold FX 2002. Range-based estimation of stochastic volatility models. *Journal of Finance* 62:1047-91

Ardia D, Boudt K, Catania L 2019. Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software* 88 doi: 10.18637/jss.v088.i06

Bernardi M, Catania L 2019. Switching generalized autoregressive score copula models with application to systemic risk. *Journal of Applied Econometrics* 34: 43-65

Billé AG, Catania L 2017. Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics* 32:1178-96

Blasques F, Koopman SJ, Lucas A 2015. Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika* 102:325-343

Blasques F, Gorgi P, Koopman SJ 2019. Accelerating score-driven time series models. *Journal of Econometrics* 212:359-76.

Blasques F, Koopman SJ, Lucas A, Schaumburg J 2016. Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics* 195:211-23

Blasques F, Gorgi P, Koopman SJ, Wintenberger O 2018. Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electronic Journal of Statistics* 12:1019-52

Blasques F, Gorgi P, Koopman SJ 2021. Missing observations in observation-driven time series models. *Journal of Econometrics* 221:542-68

Bollerslev T 1987. A conditionally heteroskedastic time series model for security prices and rates of return data. *Review of Economics and Statistics* 59:542-547

Bollerslev T, Engle RF, Nelson DB 1994. ARCH models, in *Handbook of Econometrics, Volume 4*, 2959-3038. Engle RF and McFadden DL (eds). Amsterdam: North-Holland.

Breckling J 1989. *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*. Berlin: Springer

Caivano M, Harvey AC 2014. Time series models with an EGB2 conditional distribution. *Journal of Time Series Analysis* 34: 558-71.

Caivano M, Harvey AC, Luati A 2016. Robust time series models with trend and seasonal components. *SERIEs* 7: 99-120.

Calvori F, Creal D, Koopman SJ, Lucas, A 2017. Testing for parameter instability in competing modeling frameworks. *Journal of Financial Econometrics* 15: 223-246.

Catania, L 2019. Dynamic adaptive mixture models with an application to volatility and risk. *Journal of Financial Econometrics* 17:1-34.

Catania L, Nonejad N 2019. Density forecasts and the leverage effect: evidence from observation and parameter-driven volatility models. *European Journal of Finance* https://doi.org/10.1080/1351847X.2019.1586744

Cox D R 1981. Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8:93-105.

Creal D, Koopman SJ, Lucas A 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* 29:552-63

Creal D, Koopman SJ, Lucas A 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28:777-95

De Lira Salvatierra I, Patton AJ 2015. Dynamic copula models and high frequency data. *Journal of Empirical Finance*, 30:120-135

Delle Monache D, Petrella I 2017. Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting* 33:482-501

De Rossi G, Harvey AC 2009. Quantiles, expectiles and splines. *Journal of Econometrics* 152:179-85.

Embrechts P, Kluppelberg C, Mikosch T 1997. *Modelling Extremal Events* Berlin: Springer Verlag.

Engle RF, Lilien DM, Robins, RP 1987. Estimating time varying risk premia in the term structure: the Arch-M model. *Econometrica* 55:391-407

Engle RF Manganelli S 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22:367–381.

Fisher NI, Lee AJ 1994. Time series analysis of circular data. *J R Stat Soc B* 70:327–332

Glosten LR, Jagannathan R, Runkle DE 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779-801.

Gorgi P, Hansen PR, Janus P, Koopman SJ 2019. Realized Wishart-GARCH: a score-driven multi-asset volatility model. *Journal of Financial Econometrics* 17:1-32.

Hamilton J 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384

Harvey AC 2013. *Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series*. Econometric Society Monograph. New York: Cambridge University Press

Harvey AC, Hurn S, Thiele S 2019. Modeling directional (circular) time series. *Cambridge Working Papers in Economics*, number 1971. Cambridge: Faculty of Economics.

Harvey AC, Luati A 2014. Filtering with heavy tails, *Journal of the American Statistical Association* 109:1112-22.

Harvey AC, Thiele S 2016. Testing against changing correlation. *Journal of Empirical Finance* 38:575-89

Harvey AC, Lange R-J 2017. Volatility modelling with a generalized t-distribution, *Journal of Time Series Analysis* 38:175-90.

Harvey AC, Lange R-J 2018. Modelling the interactions between volatility and returns. *Journal of Time Series Analysis* 39:909–919

Harvey AC, Ito R 2019. Modeling time series when some observations are zero. *Journal of Econometrics* 214:33-45.

Harvey AC and Liao Y 2019. Dynamic Tobit models. *Cambridge Working Papers in Economics*, number 1950, Cambridge: Faculty of Economics

Harvey A, Kattuman P 2020. Time series models based on growth curves with applications to forecasting coronavirus. *Harvard Data Science Review.* Special issue 1 - COVID -19. https://hdsr.mitpress.mit.edu/pub/ozgjx0yn

Harvey AC and Oryshchenko V 2011. Kernel density estimation for time series data. *International Journal of Forecasting* 28:3-14

Harvey A, Palumbo D 2021. Regime switching models for directional and linear observations. *Cambridge Working Papers in Economics*, number 2123. Cambridge: Faculty of Economics

Janus P, Koopman SJ, Lucas A 2014. Long memory dynamics for multivariate dependence under heavy tails *Journal of Empirical Finance* 29:187-206

Kleiber C, Kotz S 2003. *Statistical Size Distributions in Economics and Actuarial Sciences.* New York: Wiley.

Koopman SJ, Lucas A, Scharth M 2016. Predicting time-varying parameters with parameter-driven and observation-driven models, *Review of Economics and Statistics*, 98:97-110

Koopman SJ, Lit R 2019. Forecasting football match results in national league competitions using score-driven time series models *International Jour-*

nal of Forecasting 35:797–809

Koopman SJ, Lit R, Lucas A 2017. Intraday stochastic volatility in discrete price changes: the dynamic Skellam model. *Journal of the American Statistical Association* 112:1490–1503.

Lange, KL, Little RA, Taylor, JMG 1989. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* 84:881-896

Lewis RA, McDonald JB 2014. Partially adaptive estimation of censored regression models. *Econometric Reviews* 33:732-50

Lit R, Koopman SJ, Harvey AC 2020. Time Series Lab: https://timeserieslab.com

Mardia, KV, Jupp, PE 2000. *Directional statistics.* Chichester: John Wiley & Sons Ltd

Maronna R, Martin D, Yohai V 2006. *Robust Statistics: Theory and Methods.* Chichester: John Wiley & Sons Ltd.

McDonald JB, Newey WK 1988. Partially adaptive estimation of regression models via the generalized t distribution, *Econometric Theory* 4:428-457

McNeil SJ, Frey R, Embrechts P 2005. Quantitative Risk Management. Princeton: Princeton University Press

Oh DH, and Patton A 2018. Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads *Journal of Business and Economic Statistics* 36:181-95

Opschoor A, Janus P, Lucas A, van Dijk DJ 2018. New HEAVY models for fat-tailed realized covariances and returns, *Journal of Business and Economic Statistics* 36:643-7

Patton AJ 2013. Copula Methods for Forecasting Multivariate Time Series, 899-960. *Handbook of Economic Forecasting* vol 2, Elliot G, Timmermann A (eds) Amsterdam: North-Holland

Patton AJ, Ziegel JF, Chen R 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk) *Journal of Econometrics,* 21:388-413.

Zeger S L, Brookmeyer R 1986. Regression analysis with censored autocorrelated data. *Journal of the American Statistical Association* 81:722–9

Zucchini W, MacDonald IL, Langrock R 2016. Hidden Markov Models for Time Series: An Introduction using R. Boca Raton, FL: CRC Press