# The Dynamic Location/Scale Model: with applications to intra-day financial data

Philipp Andres and Andrew Harvey

September 2012

CWPE 1240

# The Dynamic Location/Scale Model: with applications to intra-day financial data

Philipp Andres and Andrew Harvey
Faculty of Economics, University of Cambridge

September, 2012

**Abstract**

In dynamic conditional score models, the innovation term of the dynamic specification is the score of the conditional distribution. These models are investigated for non-negative variables, using distributions from the generalized beta and generalized gamma families. The log-normal distribution is also considered. Applications to the daily range of stock market indices are reported and models are fitted to duration data.

**KEYWORDS**: Burr distribution; Durations; Range; Score; Unobserved components; Weibull distribution.

JEL Classification: C22, G10

## 1 Introduction

Many variables, particularly those associated with intra-day financial data, are intrinsically non-negative. Examples include the time between trades, the range of a price over a day and realized volatility; see Brownlees and Gallo (2010). Distributions appropriate for non-negative variables include the gamma, Weibull, Burr and $F$. As a rule, the location and scale for such distributions are closely connected, usually depending on the same parameter. If the location/scale is to change over time, the use of an exponential link function ensures that it remains positive. The unobserved components model is then

$$y_t = \varepsilon_t \exp(\lambda_t), \qquad 0 \leq y_t < \infty, \quad t = 1, ...., T,$$

where $\lambda_t = \ln \mu_t$ depends on a disturbance term, $\eta_t$, which may or may not be correlated with the standardized IID variable, $\varepsilon_t$. In the first-order model

$$\lambda_{t+1} = \delta + \phi\lambda_t + \eta_t, \quad \eta_t \sim NID\left(0, \sigma_\eta^2\right); \tag{1}$$

see, for example, Bauwens and Veredas (2004) and Bauwens and Hautsch (2009, p 964-5). Taking logarithms, that is

$$\ln y_t = \lambda_t + \ln \varepsilon_t, \quad t = 1, ...., T, \tag{2}$$

gives a linear state space form. For some variables, like the logarithm of range, quasi-maximum likelihood (QML) estimation using the Kalman filter may reasonably good because $\ln \varepsilon_t$ is often close to a normal distribution. Nevertheless efficient estimation usually requires the use of simulation-based methods.

Multiplicative error models (MEMs) provide an observation-driven approach for dynamic non-negative variables; see Russell and Engle (2010) for a recent survey. In these models, the conditional mean, $\mu_{t|t-1}$, and hence the conditional scale, is a linear function of past observations. The model can be written

$$\begin{aligned} y_t &= \varepsilon_t \mu_{t|t-1}, \qquad 0 \leq y_t < \infty, \quad t = 1, ...., T, & (3)\\ \mu_{t+1|t} &= \delta + \beta\mu_{t|t-1} + \alpha y_t, \qquad \delta, \alpha, \beta > 0 & (4)\end{aligned}$$

where $\varepsilon_t$ has a distribution with mean one. The emphasis in early work was on the gamma and Weibull distributions, both of which include the exponential distribution as a special case.

An exponential link function, $\mu_{t|t-1} = \exp(\lambda_{t|t-1})$ ensures that $\mu_{t|t-1}$ is positive. Exponential link functions have been studied and applied by Brandt and Jones (2006) and Bauwens and Giot (1997). However, it is the combination of the exponential link function with the conditional score that facilitates the development of an asymptotic distribution theory and enables comprehensive expressions for the moments, autocorrelations and forecasts to be derived. The practical implication is that the conditional score for a heavy-tailed distribution will give extreme observations less weight than they would receive in the standard MEM framework.

It is not always convenient to define $\varepsilon_t$ so that its mean is one. For many purposes it is better to work with a measure of scale and to set its logarithm equal to $\lambda_{t|t-1}$. Since scale and location only differ by a factor of proportionality, the statistical properties of parameters estimated with an exponential link function are essentially unchanged. The model can be written

$$y_t = \varepsilon_t \exp(\lambda_{t|t-1}), \quad t = 1, ...., T, \tag{5}$$

$$\lambda_{t+1|t} = (1-\phi)\omega + \phi\lambda_{t|t-1} + \kappa u_t, \qquad |\phi| < 1, \tag{6}$$

where $\omega$ is the unconditional mean of $\lambda_{t|t-1}$ and $\exp(\lambda_{t|t-1})$ is equal to a measure of scale,with the distribution of $\varepsilon_t$ standardized accordingly. The dynamics are driven by the conditional score, $u_t$,that is the first derivative of the logarithm of the conditional probability density function of $y_t$. We call such models dynamic conditional score (DCS) models[1].

The statistical theory of DCS models for non-negative variables is simplified by the fact that for, the gamma and Weibull distributions, the score and its derivatives are dependent on a gamma variate, while for the Burr and F-distributions the dependence is on a beta variate. The log-logistic distribution is a special case of the Burr and hence is also dependent on a beta variable. In fact the theory can be rationalized by regarding gamma and Weibull as special cases of the generalized gamma (GG) distribution, while the Burr and log-logistic distributions are special cases of the generalized beta (GB) distribution. The F-distribution is related to the GB distribution in that the special case when the degrees of freedom are the same is equivalent to a special case of GB. The generalized gamma and the generalized beta are both described in Kotz and Kleiber (2003). The distributions in the GB class are particularly useful in situations where there is evidence of heavy tails.

The first section below sets out some generic properties of DCS models for non-negative distributions and this is followed by a detailed treatment of generalized gamma, generalized beta and F-distributions. Section 5 discusses the lognormal model and it is noted that, because $\ln y_t$ is the sum of volatility and noise, the model can be treated as parameter driven or observation driven. After a section on tests and model selection, models are fitted to intra-day data on range and duration.

[1]Rather than the term dynamic conditional score models, which we use here, Creal, Koopman and Lucas ( 2011) prefer the name generalized autoregressive score (GAS) . However, despite the attraction of the acronym, the term 'autoregressive ' conveys a more limited dynamic structure than is actually the case.

# 2 General properties

The gamma distribution plays a similar role to the role of the Gaussian distribution in modeling location and scale in two-sided distributions. The ML estimators of mean and variance in samples of IID observations from a Gaussian distribution are linear combinations of the observations and their squares respectively. The ML estimator of location/scale for a gamma distribution is likewise a linear combination of the observations. For more complex dynamic models, estimation procedures associated with the gamma distribution provide a simple benchmark against which to assess other methods, in the same way as do methods associated with the normal. However, Gaussian models are vulnerable to outliers and the same is true of a MEM model or any other model, including a DCS model, which assumes a gamma distribution; see, for example, Victoria-Feser and Ronchetti (1994) and Kotz and Kleiber (2003, p 165). Using a more heavy-tailed conditional distribution does not solve the problem within the MEM framework, but it does when the DCS approach is adopted.

This section sets out the general approach to formulating DCS models for non-negative variables and explains how to derive their properties. The use of an exponential link function for each of the distributions considered yields a score that is independent of $\lambda_{t|t-1}$ and is a linear function of IID variables, which in turn depend on the standardized variables, $\varepsilon_t$.

## 2.1 Heavy tails

Although kurtosis is a good measure of heavy-tail behaviour when there are two tails, it is not satisfactory for non-negative variables. The coefficient of variation (CV), which is defined as the ratio of the standard deviation to mean, is a useful measure for characterizing distributions of non-negative variables; see, for example, Bauwens et al (2004, table 2). A distribution is said to exhibit *overdispersion* if the CV exceeds one. However, as will become apparent later, overdispersion is neither necessary nor sufficient for a distribution to be heavy-tailed and so the CV is of limited value in this respect.

The most widely accepted criterion for classifying a distribution as heavy-tailed is, like overdispersion, by reference to the exponential distribution; see Asmussen (2003).

**Definition 1** *A distribution is said to be heavy-tailed if*

$$\lim_{y\to\infty} \exp(y/\alpha)\overline{F}(y) = \infty \qquad \textit{for all } \alpha > 0, \tag{7}$$

*where* $\overline{F}(y) = \Pr(Y > y) = 1 - F(y)$ *is the survival function.*

When $y$ has an exponential distribution, $\overline{F}(y) = \exp(-y/\alpha)$, so $\exp(y/\alpha)\overline{F}(y) = 1$ for all $y$. Overdispersion arises when the CV exceeds the CV for an exponential distribution, that is unity, but overdispersion is neither necessary not sufficient for (7) to hold.

**Definition 2** *A distribution is said to be long-tailed if, for a fixed value of* $x$,

$$\lim_{y\to\infty} \Pr(Y > y + x \mid y) = 1. \tag{8}$$

When a distribution is long-tailed, the probability of an observation being bigger than a value at some point beyond $y$, given that it is known to be at least $y$, is close to one. In other words for large $y$, $\overline{F}(y + x) \simeq \overline{F}(y)$. All long-tailed distributions are heavy-tailed, but the converse is not true.

The above criteria are related to the behaviour of the conditional score and whether or not it discounts large observations.

## 2.2 Moments

When $\lambda_{t|t-1}$ is generated by a stationary process with mean $\omega$, that is

$$\lambda_{t|t-1} = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \tag{9}$$

with $\psi_j, j = 1, 2, .., $ fixed,

$$E\left(y_t^m\right) = E\left(\varepsilon_t^m\right) E(e^{m\lambda_{t|t-1}}) = E\left(\varepsilon_t^m\right) e^{m\omega} \prod_{j=1}^{\infty} E(e^{m\psi_j u_{t-j}}), \qquad m = 1, 2, ... \tag{10}$$

For all the models considered here, the $u_t's$ are linear functions of independent gamma, beta or normal variates, so the terms $E(\exp m\psi_j u_{t-j})$, $j = 1, 2, ..$, are all moment generating functions (MGFs) with a known analytic form. When

$u_t$ is beta distributed, the existence of moments of $y_t$ depends solely on the existence of moments for the conditional distribution, the reason being that the beta variable is bounded. For models in which $u_t$ is gamma distributed, conditions need to be placed on the parameters of the dynamic scale process.

The effect of volatility is to inflate the moments. In other words, the unconditional moments are greater than , or equal to, the conditional ones. Specifically, $E\left(y_t^m\right) \geq E\left(\varepsilon_t^m\right)$ because, from Jensen's inequality, $E\left(e^{m\lambda_{t|t-1}}\right) \geq \exp E\left(m\lambda_{t|t-1}\right).$

Expressions for the autocorrelations of the observations to a positive power may also be derived by making use of the formulae for moment generating functions, but the details are beyond the scope of this paper.

## 2.3 Forecasts

When $\lambda_{t|t-1}$ has a moving average representation, as in (9), the optimal estimator of $\lambda_{T+\ell|T+\ell-1}$ is its conditional expectation

$$\lambda_{T+\ell|T} = \omega + \sum_{k=0}^{\infty} \psi_{\ell+k} u_{T-k}, \quad \ell = 2, 3, ..$$

The prediction error is $\sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j}$ and the conditional moments can be found in a similar way to the unconditional moments, leading to the following results. Thus

$$E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right) = e^{\lambda_{T+\ell|T}} \prod_{j=1}^{\ell-1} E(e^{\psi_j u_{t-j}}), \quad \ell = 2, 3, ..$$

The volatility of the volatility is, for $\psi_j < \gamma/2$, $j = 1, 2, ...,$

$$\begin{aligned} VoV(\ell) &= E_T\left(e^{2\lambda_{T+\ell|T+\ell-1}}\right) - \left(E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right)\right)^2, \quad \ell = 2, 3, .. \\ &= e^{2\lambda_{T+\ell|T}} \left[ \prod_{j=1}^{\ell-1} E(e^{2\psi_j u_{t-j}}) - \left( \prod_{j=1}^{\ell-1} E(e^{\psi_j u_{t-j}}) \right)^2 \right]. \end{aligned}$$

Expressions for the optimal (MMSE) predictor of the observation at $T+\ell$, that is $E_T\left(y_{T+\ell}\right)$, and the corresponding predictor of the variance follow.

Furthermore it is straightforward to simulate the multi-step predictive distributions because

$$y_{T+\ell} = \varepsilon_{T+\ell} \exp(\sum_{j=1}^{\ell-1} \psi_j u_{T+\ell-j})[\exp(\lambda_{T+\ell|T})]. \tag{11}$$

and, as already, noted $u_t$ is a linear function of gamma or beta variates.

## 2.4 Asymptotic distribution of the maximum likelihood estimator

For the first-order model, (6), define

$$\begin{aligned} a &= \phi + \kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) \\ b &= \phi^2 + 2\phi\kappa E\left(\frac{\partial u_t}{\partial \lambda}\right) + \kappa^2 E\left(\frac{\partial u_t}{\partial \lambda}\right)^2 \geq 0 \\ c &= \kappa E\left(u_t \frac{\partial u_t}{\partial \lambda}\right), \end{aligned} \tag{12}$$

Let $\lambda_1$ be the time-varying parameter, which is a function of $\kappa, \phi$ and $\omega$,and let $\boldsymbol{\lambda}_2$ denote the fixed parameters. The static information matrix for the distribution in question is combined with the matrix

$$\mathbf{D}(\boldsymbol{\psi}) = \mathbf{D}\begin{pmatrix} \kappa \\ \phi \\ \omega \end{pmatrix} = \frac{1}{1-b}\begin{bmatrix} A & D & E \\ D & B & F \\ E & F & C \end{bmatrix} \tag{13}$$

where

$$\begin{aligned} A &= \sigma_u^2, \qquad B = \frac{\kappa^2\sigma_u^2(1+a\phi)}{(1-\phi^2)(1-a\phi)}, \qquad C = \frac{(1-\phi)^2(1+a)}{1-a}, \\ D &= \frac{a\kappa\sigma_u^2}{1-a\phi}, \qquad E = \frac{c(1-\phi)}{1-a} \qquad \textit{and} \qquad F = \frac{ac\kappa(1-\phi)}{(1-a)(1-a\phi)}, \end{aligned}$$

to give the full information matrix

$$\mathbf{I}(\boldsymbol{\psi}, \boldsymbol{\lambda}_2) = \begin{bmatrix} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\right)^2 \mathbf{D}(\boldsymbol{\psi}) & \begin{pmatrix} 0 \\ 0 \\ \frac{1-\phi}{1-a} \end{pmatrix} E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2'}\right) \\ E\left(\frac{\partial \ln L_t}{\partial \lambda_1}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2}\right)\begin{pmatrix} 0 & 0 & \frac{1-\phi}{1-a} \end{pmatrix} & E\left(\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2}\frac{\partial \ln L_t}{\partial \boldsymbol{\lambda}_2'}\right) \end{bmatrix}. \tag{14}$$

Proof of consistency and asymptotic normality follows from Harvey (2011); see appendix for more details. The condition $b < 1$ needs to be satisfied and this imposes constraints on $\kappa$. Provided these constraints are satisfied and $\kappa \neq 0$, the ML estimators of $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}_2$, denoted $\widetilde{\boldsymbol{\psi}}$ and $\widetilde{\boldsymbol{\lambda}}_2$, are consistent and the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\boldsymbol{\lambda}}_2 - \boldsymbol{\lambda}_2)'$ is multivariate normal with mean vector zero and covariance matrix

$$Var(\widetilde{\boldsymbol{\psi}}, \widetilde{\boldsymbol{\lambda}}_2) = \mathbf{I}^{-1}(\boldsymbol{\psi}, \boldsymbol{\lambda}_2). \tag{15}$$

# 3 Generalized gamma distribution

The pdf of a gamma variable, $gamma(\alpha, \gamma)$, is

$$f(g) = \alpha^{-\gamma} g^{\gamma-1} e^{-g/\alpha} / \Gamma(\gamma), \quad 0 \leq g < \infty, \quad \alpha, \gamma > 0, \tag{16}$$

where $\alpha$ is the scale parameter. When the conditional distribution of $y_t$ in (5) is $gamma(\alpha, \gamma)$, with $\varepsilon_t$ standardized by setting its scale equal to one, the exponential link function $\alpha_{t|t-1} = \exp(\lambda_{t|t-1}) = \mu_{t|t-1}/\gamma$, $t = 1, .., T$, yields the log-density for the $t-th$ observation as

$$\ln f_t(\boldsymbol{\psi}, \gamma) = -\gamma\lambda_{t|t-1} + (\gamma - 1)\ln y_t - y_t \exp(-\lambda_{t|t-1}) - \ln\Gamma(\gamma), \quad t = 1, ...., T.$$

The conditional score is then

$$u_t = y_t / \exp(\lambda_{t|t-1}) - \gamma = \varepsilon_t - \gamma, \quad t = 1, ...., T,$$

with $\sigma_u^2 = \gamma$. Thus the score variables are just the centered IID standardized gamma variables of (5).

The pdf of of the Weibull distribution is

$$f(y; \alpha, \upsilon) = \frac{\upsilon}{\alpha}\left(\frac{y}{\alpha}\right)^{\upsilon-1} \exp\left(-(y/\alpha)^{\upsilon}\right), \quad 0 \leq y < \infty, \quad \alpha, \upsilon > 0.$$

where $\alpha$ is the scale and $\upsilon$ is the shape parameter. The mean is $\mu = \alpha\Gamma(1 + 1/\upsilon)$ and the variance is $\alpha^2\Gamma(1 + 2/\upsilon) - \mu^2$. Since $\overline{F}(y) = \exp[-(y/\alpha)^{\upsilon}]$, the expression in Definition 1 is $\exp[y/\alpha - (y/\alpha)^{\upsilon}]$. Hence the Weibull distribution has a heavy tail when $\upsilon < 1$ and in this case a plot of the score shows that large observations are discounted.

The properties of the gamma model are relatively easy to derive. However, most of them are given as a special case of the generalized gamma distribution, as are the properties of the Weibull distribution. The generalized gamma (GG) distribution is

$$f(y;\alpha,\gamma,\upsilon)=\frac{\upsilon}{\alpha\Gamma(\gamma)}\left(\frac{y}{\alpha}\right)^{\upsilon\gamma-1}\exp\left(-(y/\alpha)^{\upsilon}\right),\quad 0\leq y<\infty,\quad \alpha,\gamma,\upsilon>0,$$

The mean is $\alpha\Gamma(\gamma+1/\upsilon)/\Gamma(\gamma)$. A full description is given in Kleiber and Kotz (2003, ch 5). The gamma distribution is obtained for $\upsilon=1$, while setting $\gamma=1$ with $\upsilon>0$ yields the Weibull distribution. Setting both parameters equal to one gives the exponential distrbution.

**Remark 3** *For a gamma distribution, the coefficient of variation is $1/\gamma$. Thus there is overdispersion if $\gamma<1$. Similarly the Weibull distribution displays overdispersion when $\upsilon<1$. As shown earlier, a Weibull distribution has a heavy tail when $\upsilon<1$. However, a gamma distribution never has a heavy tail and this feature is consistent with its linear score function.*

The log-density for the $t-th$ observation when $\alpha_{t|t-1}$ is time-varying with an exponential link function is

$$\ln f_t(\lambda,\gamma,\upsilon)=\ln\upsilon-\lambda_{t|t-1}+(\upsilon\gamma-1)\ln(y_te^{-\lambda_{t|t-1}})-(y_te^{-\lambda_{t|t-1}})^{\upsilon}-\ln\Gamma(\gamma)$$

giving a score of

$$\frac{\partial\ln f_t}{\partial\lambda_{t|t-1}}=u_t=\upsilon(y_te^{-\lambda_{t|t-1}})^{\upsilon}-\upsilon\gamma=\upsilon(\varepsilon_t^{\upsilon}-\gamma) \tag{17}$$

The following result, which can be proved directly by change of variable, enables the properties of the model to be obtained relatively easily.

**Proposition 4** *For the generalized gamma distribution, the variables $g_t=\varepsilon_t^{\upsilon}=(y_te^{-\lambda_{t|t-1}})^{\upsilon}$ are IID as gamma$(1,\gamma)$ at the true parameter values.*

## 3.1 Moments

**Proposition 5** *For the generalized gamma model defined by (5) and (9), and $u_t$ as in (17), the $m-th$ moment exists if and only if $\psi_j<1/\upsilon m$, for all $j=1,2,...$, and is given by the expression*

$$E\left(y_t^m\right)=\frac{\Gamma(\gamma+m/\upsilon)}{\Gamma(\gamma)}e^{m(\omega-\gamma\upsilon\Sigma\psi_j)}\prod_{j=1}^{\infty}(1-\upsilon m\psi_j)^{-\gamma},\quad \psi_j<1/\upsilon m,\quad m>0. \tag{18}$$

**Proof.** In (10) $u_t = \upsilon(\varepsilon_t^{\upsilon} - \gamma)$ and the formula for the moment generating function of a standardized gamma variate is $\gamma_{\gamma}(c) = E(e^{cg}) = (1-c)^{-1/\gamma}$. Since $c = \upsilon m \psi_j$, the result follows. ■

Conditional moments of forecasts can be computed in a similar way. It is easy to simulate the multi-step predictive distribution simply by generating independent gamma variates.

## 3.2 Asymptotic distribution of ML estimators

Since

$$\frac{\partial^2 \ln f_t}{\partial \lambda^2} = -\upsilon^2 (y_t e^{-\lambda})^{\upsilon} = -\upsilon^2 g_t, \qquad t = 1, ..., T,$$

both the score and its derivative depend on $g_t$. Hence $a, b$ and $c$ are easily evaluated.

The information matrix is given in Kleiber and Kotz (2003, p 157). When the scale parameter is replaced by its logarithm, $\lambda$, the information matrix is independent of $\lambda$ and given by

$$\mathbf{I}\begin{pmatrix} \lambda \\ \gamma \\ \upsilon \end{pmatrix} = \begin{bmatrix} \upsilon^2\gamma & \upsilon & -1-\gamma\psi(\gamma) \\ \upsilon & \psi'(\gamma) & -\psi(\gamma)/\upsilon \\ -1-\gamma\psi(\gamma) & -\psi(\gamma)/\upsilon & \upsilon^{-2}\{1+\psi(\gamma)[2+\psi(\gamma)]+\gamma\psi'(\gamma)\} \end{bmatrix}$$

where $\psi(\gamma)$ and $\psi'(\gamma)$ are the digamma and trigamma functions respectively.

**Proposition 6** *Consider the first-order model, (6), with unknown parameters* $\boldsymbol{\psi} = (\kappa, \phi, \omega)'$. *Provided that* $|\phi| < 1$ *and* $b < 1$, *the limiting distribution of* $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\gamma} - \gamma, \widetilde{\upsilon} - \upsilon)'$ *is multivariate normal with zero mean and covariance matrix given by the inverse of (14) with* $\boldsymbol{\lambda}_2 = (\gamma, \upsilon)'$ *and* $\mathbf{D}(\boldsymbol{\psi})$ *is as in (13) with*

$$\begin{aligned} a &= \phi - \gamma\upsilon^2\kappa \\ b &= \phi^2 - 2\phi\kappa\upsilon^2\gamma + \kappa^2\upsilon^4(1+\gamma)\gamma \\ c &= -\kappa\upsilon^3\gamma, \end{aligned}$$

*and* $\sigma_u^2 = \upsilon^2\gamma$.

**Proof.** Since $g_t \sim gamma(1, \gamma)$,

$$E\left[\left(\frac{\partial u_t}{\partial \lambda}\right)^k\right] = (-1)^k \upsilon^{2k} \frac{\Gamma(k+\gamma)}{\Gamma(\gamma)}, \quad k = 1, 2, .,$$

To find $a$ and $b$, note that, from the formula above, $E(u_t') = -\upsilon^2\gamma$ and $E(u_t'^2) = \upsilon^4(1+\gamma)\gamma$, while for $c$

$$E\left[u_t\left(\frac{\partial u_t}{\partial \lambda}\right)\right] = \upsilon^3 E\left[g_t^2 - \gamma g_t\right] = -\upsilon^3\gamma.$$

■

For the exponential distribution, given by setting $\gamma = \upsilon = 1$, $Var\left(\widetilde{\boldsymbol{\psi}}\right) = \mathbf{D}^{-1}(\boldsymbol{\psi})$, while the expression for $b$ is $b = \phi^2 - 2\phi\kappa + 2\kappa^2$. The constraint that $b < 1$ permits a wide range of admissible values of $\kappa$ and the same is true for gamma and Weibull distributions. As will be seen later, estimates obtained in practice are typical quite small.

# 4 Generalized beta distribution

The generalized beta distribution (of the second kind)

$$f(y) = \frac{\nu(y/\alpha)^{\nu\xi-1}}{\alpha B(\xi,\varsigma)\left[(y/\alpha)^\nu + 1\right]^{\xi+\varsigma}}, \qquad \alpha, \nu, \xi, \varsigma > 0 \tag{19}$$

where $\alpha$ is the scale parameter, $\nu, \xi$ and $\varsigma$ are shape parameters and $B(\xi, \varsigma)$ is the beta function.

If all three shape parameters are unrestricted, the model is difficult to estimate. However, the GG distribution is extremely useful from the theoretical point of view as it contains many important distributions as special cases. These include the log-logistic and Burr distributions. The F-distribution is closely related. Many of these distributions have been found useful as models of income and wealth and losses in insurance; see Kleiber and Kotz (2003, ch. 6). All have potential in financial econometrics,

The log-logistic distribution contains only one shape parameter and so, like the gamma and Weibull distributions, it is relatively easy to handle. We therefore begin by examinating this distribution before embarking on the general case.

## 4.1 Log-logistic distribution

The pdf for the log-logistic distribution is

$$f(y) = (\nu/\alpha)(y/\alpha)^{\nu-1}(1 + (y/\alpha)^{\nu})^{-2}, \qquad \nu, \alpha > 0.$$

A time-varying scale with an exponential link function gives a log-density of

$$\ln f_t(\boldsymbol{\psi}, \nu) = \ln \nu - \nu \lambda_{t|t-1} + (\nu - 1) \ln y_t - 2 \ln(1 + (y_t e^{-\lambda_{t|t-1}})^{\nu}), \qquad t = 1, ..., T,$$

and so

$$\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = u_t = \frac{2\nu(y_t e^{-\lambda_{t|t-1}})^{\nu}}{1 + (y_t e^{-\lambda_{t|t-1}})^{\nu}} - \nu = 2\nu b_t(1, 1) - \nu, \tag{20}$$

where

$$b_t(1, 1) = \frac{(y_t e^{-\lambda_{t|t-1}})^{\nu}}{1 + (y_t e^{-\lambda_{t|t-1}})^{\nu}}$$

is distributed as $beta(1, 1)$. This result may be shown directly from the change of variable. Since a $beta(1, 1)$ distribution is a standard uniform distribution, it is immediately apparent that the expectation of $u_t$ is zero. Furthermore $\sigma_u^2 = \nu^2/3$.

The asymptotic theory for the log-logistic distribution is not complicated. Differentiating the score gives

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = -2\nu^2 b_t(1 - b_t).$$

Provided that $b < 1$, the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\nu} - \nu)'$ for a stationary first-order dynamic equation, (6), is multivariate normal with zero mean and covariance matrix

$$Var\begin{pmatrix} \widetilde{\boldsymbol{\psi}} \\ \widetilde{\nu} \end{pmatrix} = \begin{bmatrix} (3/\nu^2)\mathbf{D}^{-1}(\boldsymbol{\psi}) & 0 \\ 0 & \nu^2/1.430 \end{bmatrix}, \tag{21}$$

where $\mathbf{D}(\boldsymbol{\psi})$ is as in (13) with $\sigma_u^2 = \nu^2/3$, $a = \phi - \kappa\nu^2/3$, $b = \phi^2 - (2/3)\nu^2\phi\kappa + 2\kappa^2\nu^4/15$ and $c = 0$.

## 4.2 Asymptotic theory for the generalized beta distribution

The log-density for the generalized beta (GB2) distribution with an exponential link function for the scale is

$$\ln f_t(\nu, \xi, \varsigma) = \ln \nu - \nu\xi\lambda_{t|t-1} + (\nu\xi - 1) \ln y_t - (\xi + \varsigma) \ln((y_t e^{-\lambda_{t|t-1}})^{\nu}) + 1) - \ln B(\xi, \varsigma),$$

and so

$$\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = u_t = \nu(\xi+\varsigma)\frac{(y_t e^{-\lambda_{t|t-1}})^{\nu}}{(y_t e^{-\lambda_{t|t-1}})^{\nu}+1} - \nu\xi \tag{22}$$
$$= \nu(\xi+\varsigma)b_t(\xi,\varsigma) - \nu\xi,$$

where

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

Since $0 \leq b_t(\xi,\varsigma) \leq 1$, it follows that as $y \to \infty$, the score approaches an upper bound of $\nu\varsigma$.

The following result can be proved directly.

**Proposition 7** *At the true parameter values, the variable $b_t(\xi,\varsigma)$ is IID with a beta$(\xi,\varsigma)$ distribution.*

The static information matrix can be found in Kleiber and Kotz ( 2003, p194). Adapting it to the exponential link function gives

$$\mathbf{I}\begin{pmatrix} \lambda \\ \nu \\ \xi \\ \varsigma \end{pmatrix} = \begin{bmatrix} \frac{\nu^2\xi\varsigma}{1+\xi+\varsigma} & I_{12} & \frac{\nu\varsigma}{\xi+\varsigma} & \frac{-\nu\xi}{\xi+\varsigma} \\ I_{21} & I_{22} & I_{23} & I_{24} \\ \frac{\nu\varsigma}{\xi+\varsigma} & I_{23} & \psi'(\xi)-\psi'(\xi+\varsigma) & -\psi'(\xi+\varsigma) \\ \frac{-\nu\xi}{\xi+\varsigma} & I_{24} & -\psi'(\xi+\varsigma) & \psi'(\varsigma)-\psi'(\xi+\varsigma) \end{bmatrix}, \tag{23}$$

where

$$I_{21} = I_{12} = \frac{\xi-\varsigma-\xi\varsigma\left(\psi(\xi)-\psi(\varsigma)\right)}{1+\xi+\varsigma}$$
$$I_{23} = I_{32} = -\frac{\varsigma\left(\psi(\xi)-\psi(\varsigma)\right)-1}{\nu(\xi+\varsigma)}$$
$$I_{24} = I_{42} = -\frac{\xi\left(\psi(\varsigma)-\psi(\xi)\right)-1}{\nu(\xi+\varsigma)}$$

and

$$I_{22} = \frac{1}{\nu^2(1+\xi+\varsigma)}\left[1+\xi+\varsigma+\xi\varsigma\left(\psi'(\xi)+\psi'(\varsigma)\right)+\left(\psi(\varsigma)-\psi(\xi)+\frac{\xi-\varsigma}{\xi\varsigma}\right)^2-\left(\frac{\xi^2+\varsigma^2}{\xi^2\varsigma^2}\right)\right].$$

The proposition below follows from proposition 7 on the score and the fact that its derivative is

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = \frac{-\nu^2(\xi+\varsigma)(y_t e^{-\lambda_{t|t-1}})^\nu}{((y_t e^{-\lambda_{t|t-1}})^\nu+1)^2} = -\nu^2(\xi+\varsigma)b_t(1-b_t). \tag{24}$$

The following result is used for a *beta* $(\alpha,\beta)$ variable:

$$E(b^h(1-b)^k) = \frac{B(\alpha+h,\beta+k)}{B(\alpha,\beta)}, \quad h>-\alpha, k>-\beta. \tag{25}$$

A proof of consistency and asymptotic normality of the ML estimator poses no serious problems because of the boundedness of the score and its derivatives.

**Proposition 8** *For a conditional GB2 distribution with a first-order stationary dynamic model with $b<1$, the limiting distribution of $\sqrt{T}(\widetilde{\boldsymbol{\psi}}' - \boldsymbol{\psi}', \widetilde{\nu}-\nu, \widetilde{\xi}-\xi, \widetilde{\varsigma}-\varsigma)'$ is multivariate normal with covariance matrix given by the inverse of (14) with the static information matrix given by (23) and $a, b$ and $c$ obtained using (25) to evaluate*

$$E\left[\frac{\partial u_t}{\partial \lambda}\right] = -\nu^2(\xi+\varsigma)E(b_t(1-b_t)) = \frac{-\nu^2\xi\varsigma}{\xi+\varsigma+1},$$

$$E\left[\frac{\partial u_t}{\partial \lambda}\right]^2 = \nu^4(\xi+\varsigma)^2E[b_t^2(1-b_t)^2] = \frac{\nu^4(\xi+\varsigma)\xi\varsigma\,(\varsigma+1)\,(\xi+1)}{(\varsigma+\xi+3)(\varsigma+\xi+2)(\varsigma+\xi+1)}$$

*and*

$$\begin{aligned} E\left[u_t\frac{\partial u_t}{\partial \lambda}\right] &= \nu^3(\xi+\varsigma)^2E[b_t^2(1-b_t)] - \nu^3(\xi+\varsigma)E[b_t(1-b_t)] \\ &= \frac{\nu^3\xi\varsigma(\xi+\varsigma)\,(\xi+1)}{(\varsigma+\xi+2)(\varsigma+\xi+1)} - \frac{\nu^3\xi\varsigma}{(\xi+\varsigma+1)}. \end{aligned}$$

**Corollary 9** *For the log-logistic distribution, $\xi=\varsigma=1$, and the result in (21) follows because*

$$I_{11} = \nu^2/3, \qquad I_{12}=0 \qquad \textit{and}$$

$$I_{22} = \frac{1}{3\nu^2}\left[3+\psi'(1)+\psi'(1)-2\right] = \frac{1}{3\nu^2}\left[\frac{\pi^2}{3}+1\right] = 1.4300/\nu^2.$$

## 4.3 Moments and forecasts for the generalized beta distribution

Expressions for unconditional moments and ACFs of observations raised to any positive power can be derived, as can the conditional expectations needed for forecasts.The unconditional expectation of $\exp(m\lambda_{t|t-1})$ is needed but this can be found from the MGF of a beta variate. Suppose $b$ has a $beta(\alpha,\beta)$ distribution and $c$ is a finite number. Then

$$M_{\beta}(c;\alpha,\beta)=E(e^{cb})=1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{c^k}{k!},\quad \alpha,\beta>0 \tag{26}$$

The above expression is Kummer's (confluent hypergeometric) function, ${}_1F_1(\alpha;\beta;c)$, and it is available as a standard routine in many packages.

**Proposition 10** *For the generalized beta model defined by (5) and (9), and $u_t$ as in (22), the $m-th$ moment exists if and only if $\psi_j<1/\nu m$, for all $j=1,2,...$, and is given by the expression*

$$E(y_t^m)=E(\varepsilon_t^m)E\exp(m\lambda_{t|t-1}),\quad -\nu\xi<m<\nu\varsigma,$$

*where the moments of the standardized variable are*

$$E(\varepsilon_t^m)=\frac{\Gamma(\xi+m/\nu)\Gamma(\varsigma-m/\nu)}{\Gamma(\xi)\Gamma(\varsigma)},\quad -\nu\xi<m<\nu\varsigma,$$

*and*

$$E\exp(m\lambda_{t|t-1})=e^{m(\omega-\nu\xi\Sigma\psi_j)}\prod_{j=1}^{\infty}M_{\beta}(\nu(\xi+\varsigma)m\psi_j;\xi,\varsigma). \tag{27}$$

*with $M_{\beta}(c;\alpha,\beta)$ denoting the MGF of a standardized beta variate, (26).*

**Proof.** In (10) $u_t=\nu(\xi+\varsigma)b_t-\nu\xi$ and from (26) by setting $c=\nu(\xi+\varsigma)m\psi_j$. ■

The optimal predictor of the scale is

$$\begin{aligned}\alpha_{T+\ell|T} &= E_T\left(e^{\lambda_{T+\ell|T+\ell-1}}\right)\\ &= e^{\lambda_{T+\ell|T}}e^{m(\omega-\nu(\xi+\varsigma)m\Sigma\psi_j)}\prod_{j=1}^{\infty}e^{\nu\xi m\Sigma\psi_j}M_{\beta}(\nu(\xi+\varsigma)m\psi_j;\xi,\varsigma),\end{aligned}$$

while the predictor of level is

$$\mu_{T+\ell|T} = \frac{\Gamma(\xi + 1/\nu)\Gamma(\varsigma - 1/\nu)}{\Gamma(\xi)\Gamma(\varsigma)} \alpha_{T+\ell|T}, \quad \ell = 2, 3, .., \qquad \nu\varsigma > 1.$$

The volatility of the volatility can be similarly found as can the predictor of the variance of $y_{T+\ell}$. It is easy to simulate the multi-step predictive distribution by generating independent beta variates in (11).

### 4.4 Burr distribution

The generalized (Type XII) Burr distribution, also known as the Singh-Maddala distribution, is obtained by setting $\xi = 1$. There are a number of different parameterizations; see, for example, Tadikamalla (1980) and Grammig and Maurier (2000). But the one based on the GB2 distribution is most convenient. The Weibull distribution can be obtained by letting $\varsigma \to \infty$. Indeed the Burr distribution is sometimes called the compound Weibull.

The log-logistic distribution is a special case of the Burr distribution obtained by setting $\varsigma = 1$. When a Burr distribution is fitted, a Wald test of $\varsigma = 1$ is straightforward.

The CDF, and hence the PIT, of a Burr variate has, as Kleiber and Kotz (2003, p198) put it, the 'pleasantly simple form'

$$F(y) = 1 - \left[ \left( ye^{-\lambda_{t|t-1}} \right)^{\nu} + 1 \right]^{-\varsigma}. \tag{28}$$

The quantile function is equally simple, namely

$$F^{-1}(\tau) = \exp \lambda_{t|t-1} [(1-\tau)^{-1/\varsigma} - 1]^{1/\nu}, \qquad 0 < \tau < 1.$$

The PIT for the unrestricted GB2 distribution is more complicated; see Kleiber and Kotz (2003, p188).

The fact that the survival function, $\overline{F}(y)$, is $[(y/\alpha)^{\nu} + 1]^{-\varsigma}$ makes it easy to see that the Burr distribution is long-tailed, and therefore heavy-tailed, since in Definition 2

$$\frac{\overline{F}(y+x)}{\overline{F}(y)} = \left[ \frac{(y/\alpha)^{\nu} + 1}{((y+x)/\alpha)^{\nu} + 1} \right]^{\varsigma} \to 1 \qquad as \quad y \to \infty.$$

As noted below (22), the score for all GB2 distributions discounts large observations as it approaches an upper bound of $\nu\varsigma$ as $y \to \infty$.

The coefficient of variation can indicate overdispersion or underdispersion. For a Burr distribution the variance, and hence the CV, goes to zero as $\nu\varsigma \rightarrow \infty$. It only exceeds one if $\nu\varsigma$ is sufficiently small, but recall that when $\nu\varsigma \leq 2$ the variance does not exist as $\Gamma(\varsigma - 2/\nu) \rightarrow \infty$ as $\nu\varsigma \rightarrow 2$.

## 4.5 F-distribution

The $F(\nu_1, \nu_2)$-distribution with $\nu_1 = \nu_2$ is a special case of the GB2 distribution obtained by setting $\xi = \varsigma = \nu_1/2 = \nu_2/2$ and $\nu = 1$. Even though $F$-distributions with different degrees of freedom do not fall within the GB2 class, the score has a beta distribution and the properties of the model may be derived along similar lines.

When $\varepsilon_t$ in (5) is from an $F(\nu_1, \nu_2)$ distribution, the logarithm of the pdf for the conditional distribution of the $t$-th observation in (5) is

$$\begin{aligned}\ln f_t(\boldsymbol{\psi},\nu_1,\nu_2) \;=\; & \frac{\nu_1}{2}\ln \nu_1 y_t e^{-\lambda_{t|t-1}} + \frac{\nu_2}{2}\ln \nu_2 \;-\; \frac{\nu_1+\nu_2}{2}\ln(\nu_1 y_t e^{-\lambda_{t|t-1}} + \nu_2) \\ & - \ln y_t - \ln B(\nu_1/2, \nu_2/2).\end{aligned}$$

Hence the score is

$$\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = \frac{\nu_1+\nu_2}{2} b_t(\nu_1/2, \nu_2/2) - \frac{\nu_1}{2},$$

where

$$b_t(\nu_1/2, \nu_2/2) = \frac{\nu_1 y_t e^{-\lambda_{t|t-1}}/\nu_2}{1 + \nu_1 y_t e^{-\lambda_{t|t-1}}/\nu_2} = \frac{\nu_1 \varepsilon_t/\nu_2}{1 + \nu_1 \varepsilon_t/\nu_2}.$$

Since $\varepsilon_t$ depends on the ratio of independent chi-square variables, $b_t(\nu_1/2, \nu_2/2)$ is distributed as $beta(\nu_1/2, \nu_2/2)$. Taking expectations confirms that the score has zero mean since $E(b_t(\nu_1/2, \nu_2/2)) = \nu_1/(\nu_1 + \nu_2)$.

The moments of the dynamic $F$-distribution can be found from the properties of the beta distribution. As regards the asymptotic distribution, differentiating the score gives

$$\frac{\partial u_t}{\partial \lambda_{t|t-1}} = -\frac{\nu_1+\nu_2}{2} b_t(1 - b_t).$$

Hence

$$E_{t-1}\left[\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right]^2 = \left(\frac{\nu_1+\nu_2}{2}\right)^2 E(b_t^2(1-b_t)^2)$$

and

$$E_{t-1}\left[u_t \frac{\partial u_t}{\partial \lambda_{t|t-1}}\right] = -\left(\frac{\nu_1+\nu_2}{2}\right) E(b_t^2(1-b_t)) + \frac{\nu_1}{2} E(b_t(1-b_t)).$$

are easily found. The formulae for $a, b$ and $c$ are similar to those for the Burr distribution.

# 5 Lognormal

If, in the UC model (2) and (1), the disturbance term, $\ln \varepsilon_t$, and the disturbance driving $\lambda_t$ are both Gaussian, the model is linear and can be handled efficiently by the Kalman filter. The log-likelihood function is constructed from the prediction error decomposition. The log-density is

$$\begin{aligned} \ln f_t &= -(1/2)\ln 2\pi - \frac{1}{2}\ln \sigma_t^2 - \frac{1}{2\sigma_t^2}(\ln y_t - \lambda_{t|t-1})^2, \qquad t = 1, ..., T, \\ &= -(1/2)\ln 2\pi - \frac{1}{2}\ln \sigma_t^2 - \frac{1}{2\sigma_t^2}(\ln y_t e^{-\lambda_{t|t-1}})^2 \end{aligned}$$

where $\sigma_t^2$ is the prediction error variance. Here $\lambda_{t|t-1}$ is a time-varying conditional mean for $\ln y_t$ but the logarithm of a conditional scale for $y_t$. The parameters are $Var(\ln \varepsilon_t)$, $\delta$, $\phi$ and $\sigma_\eta^2$ ( or the signal-noise ratio $q = \sigma_\eta^2 / Var(\ln \varepsilon_t)$. However, the parameters could also be taken to be as in the steady-state Kalman filter, so they become[2] $\sigma^2$, $\delta$ (or $\omega$), $\phi$ and $\kappa$. The model is then within the DCS class, with conditional score

$$\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} = \frac{\ln(y_t \exp(-\lambda_{t|t-1}))}{\sigma^2}$$

Dividing by the information quantity gives

$$u_t = \ln(y_t \exp(-\lambda_{t|t-1})) = \ln \varepsilon_t, \qquad t = 1, ..., T.$$

The distribution of $y_t$ is lognormal, and from this point of view the model is seen to be within the DCS class for non-negative variables. Using obvious notation, the likelihood function for the $y_t's$ is

$$\ln L(y) = \ln L(\ln y) - \sum \ln y_t,$$

[2] When the KF is run, $\sigma_t^2$ will normally be time-varying because of the initialization, but will tend towards a constant as $t \to \infty$.

thereby allowing a direct comparison of the fit achieved with other distributions.

# 6 Monte Carlo experiments

A set of Monte Carlo experiments was carried out to determine the small sample properties of ML estimators. For samples of size 1000 and 10,000, the means and standard deviations of the estimates from 5,000 replications were obtained for gamma ($\gamma = 6$), Weibull ($\upsilon = 2$), log-logistic ($\nu = 4$) and Burr ($\nu = 4, \varsigma = 0.75$). The dynamic parameters were set to $\phi = 0.98$ and $\kappa = 0.1$, both of which are typical of what might be expected in practice. According to the asymptotic theory the large sample distribution of all estimators is independent of $\omega$, which was therefore set to zero.

The averages of the ML estimates were close to the true values. The standard deviations for $\omega$, $\phi$ and $\kappa$ shown in Table 1 tend to be slightly bigger than the asymptotic standard errors ($ASEs$). In almost all cases the figures are closer for $T = 10,000$ than for $T = 1,000$. No problems were encountered in maximizing the log-likelihood functions using the FMINCON routine in Matlab.

| Parameters | Gamma Model 1,000 | $ASE$ | 10,000 | $ASE$ | Weibull Model 1,000 | $ASE$ | 10,000 | $ASE$ |
|---|---|---|---|---|---|---|---|---|
| $\omega$ | 2.51 | 1.91 | 0.58 | 0.60 | 6.21 | 4.24 | 1.40 | 1.34 |
| $\phi$ | 2.51 | 1.89 | 0.57 | 0.60 | 1.24 | 0.85 | 0.28 | 0.27 |
| $\kappa$ | 2.00 | 1.86 | 0.59 | 0.59 | 0.83 | 0.81 | 0.25 | 0.26 |
| $\nu$ or $\upsilon$ | 10.56 | 10.53 | 3.27 | 3.33 | 5.04 | 6.01 | 1.50 | 1.90 |

| Parameters | Lognormal Model 1,000 | $ASE$ | 10,000 | $ASE$ | Log-Logistic Model 1,000 | $ASE$ | 10,000 | $ASE$ |
|---|---|---|---|---|---|---|---|---|
| $\omega$ | 2.69 | 2.01 | 0.66 | 0.63 | 2.69 | 2.01 | 0.66 | 0.63 |
| $\phi$ | 1.10 | 0.80 | 0.27 | 0.25 | 1.10 | 0.80 | 0.27 | 0.25 |
| $\kappa$ | 1.31 | 1.32 | 0.43 | 0.42 | 1.31 | 1.32 | 0.43 | 0.42 |
| $\sigma$ or $\nu$ | 10.57 | 10.58 | 3.43 | 3.35 | 10.57 | 10.58 | 3.43 | 3.35 |

| Parameters | Burr Model 1,000 | $ASE$ | 10,000 | $ASE$ | F Model 1,000 | $ASE$ | 10,000 | $ASE$ |
|---|---|---|---|---|---|---|---|---|
| $\omega$ | 3.10 | 2.21 | 0.68 | 0.70 | 2.84 | 2.07 | 0.69 | 0.66 |
| $\phi$ | 1.26 | 0.84 | 0.31 | 0.27 | 1.17 | 0.81 | 0.27 | 0.26 |
| $\kappa$ | 1.50 | 1.60 | 0.58 | 0.50 | 1.56 | 1.38 | 0.46 | 0.44 |
| $\nu$ or $\nu_1$ | 8.88 | 9.86 | 2.36 | 3.12 | 28.24 | 13.93 | 9.00 | 4.40 |
| $\varsigma$ or $\nu_2$ | 20.96 | 22.31 | 6.22 | 7.05 | 65.18 | 32.64 | 27.50 | 10.32 |

Table 1 Estimated standard deviations of ML estimates ($\times 100$) for $T = 1000$ and $T = 10,000$ from 5,000 replications

# 7 Leverage and components

Leverage effects are likely to play a prominent role for variables associated with stock returns. They may be included in the DCS models by adding the variable $sgn(return_t)(u_t + E(u_t))$ to the dynamic equation. An alternative, but equivalent formulation takes the additional leverage variable to be $I(return_t < 0)(u_t + E(u_t))$, where $I(return_t < 0)$ is an indicator taking the value one when the return is negative and zero otherwise.

Alizadeh, Brandt and Diebold (2002, p 1088) argue strongly for two component (or two factor) stochastic volatility dynamics, in both equity and foreign exchange. They model such components using a SV framework while Engle and Lee (1999) proposed a two component GARCH model. In both papers, volatility is modeled with a long-run and a short-run component, the main role of the short-run component being to pick up the temporary increase in volatility after a large shock. Such a model can mimic long memory behaviour.

The DCS two-component model is

$$\lambda_{t|t-1} = \omega + \lambda_{1,t|t-1} + \lambda_{2,t|t-1}, \qquad with \qquad \lambda_{i,t+1|t} = \phi_i \lambda_{i,t|t-1} + \kappa_i u_t, \quad i = 1, 2,$$

Note that the score appears in the equations for both components. The formulae for moments and ACFs can be obtained as before since the moving average representation is

$$\lambda_{t|t-1} = \omega + \sum_{k=1}^{\infty} \psi_{1,k} u_{t-k} + \sum_{k=1}^{\infty} \psi_{2,k} u_{t-k} = \omega + \sum_{k=1}^{\infty} \psi_k u_{t-k}$$

where $\psi_k = \psi_{1k} + \psi_{2k}$, $k = 1, 2, ...$

The long-term component, $\lambda_{1,t|t-1}$, will usually have $\phi_1$ close to one, or even set equal to one. The short-term component, $\lambda_{2,t|t-1}$, will have higher $\kappa$ combined with lower $\phi$. Hence the constraint $0 < \phi_2 < \phi_1 < 1$ is typically imposed to ensure identifiability (and stationarity) and $\kappa_1$ may be forced to be less than or equal to $\kappa_2$. In the context of GARCH models, Engle and Lee (1999, p 487) find that the leverage effect is mainly restricted to the short-run component. Similar results tend to be found when DCS model are fitted.

# 8 Tests and model selection

Model selection requires decisions to be made about the distribution and the form of the dynamic equation for the scale. The starting point is testing against serial correlation using a portmanteau or Box-Ljung statistic constructed from the sample autocorrelations of the observations. But just as squared observations may be unduly influenced by outliers in returns, so the observations themselves may not be robust here. A square root or logarithmic transformation may be better, but since the observations are independent under the null hypothesis, any transformation can be used.

If a distribution is specified at the outset, the Lagrange multiplier (LM) principle suggests tests based on the score. When the distribution is gamma, the LM test simply uses the observations, the $y_t' s$. For distributions from the generalized gamma family, the LM test will use the autocorrelations for $y_t^{\upsilon}$, where the shape parameter, $\upsilon$, is estimated.

When there is no prior guidance as to the form of the conditional distribution, it may be useful to fit an unobserved components dynamic model to the logarithm of the observations. The measurement equation is (2) and this is combined with a suitable transition equation. The simplicity of the estimation procedure, which assumes a linear model, offers the opportunity to experiment with various dynamic specifications. Examining the distribution of the residuals may suggest which distributions are likely to prove most useful. Indeed a test of normality may indicate that the lognormal model fits the data in which case there is no need to proceed any further.

When a DCS model is fitted, diagnostic tests of serial correlation and distribution can be based on the scores, the residuals, that is $y_t \exp(-\lambda_{t|t-1})$, the PITs of the residuals and the normalized PITs. As with the exponential distribution, the PITs for a Weibull distribution are given by a simple formula, namely

$$PIT(y_t) = 1 - \exp[-(y_t e^{-\lambda_{t|t-1}})^{\upsilon}], \qquad t = 1, ...., T.$$

Computing the PIT for an observation from a gamma or generalized gamma distribution is a little more complicated in that it requires the evaluation of an incomplete gamma function. Similarly for some members of the generalized beta family, including the $F-$distribution, finding the PIT requires a routine for computing a regularized incomplete beta function. However, there is a closed form expression, (28), for the PIT of a Burr distribution, and hence for the log-logistic as well.

The Lagrange multiplier test principle suggest that the scores be used to test against serial correlation. However, a test based on the residuals may also be informative. An attraction of making the probability integral transformation to the residuals is that it may yield serial correlation tests which are more robust. Furthermore the PITs are comparable for different conditional distributions and their histograms are very useful for assessing goodness of fit. Figure 1 shows the PIT from fitting a DCS gamma model to the range data for the DOW Jones described in Section 9. The gamma distribution is clearly unsatisfactory: the fit near the origin is poor and the high values close to one indicate that a heavy-tail is not being captured. The parallel lines on the graph are such that, if the PITs were independent and

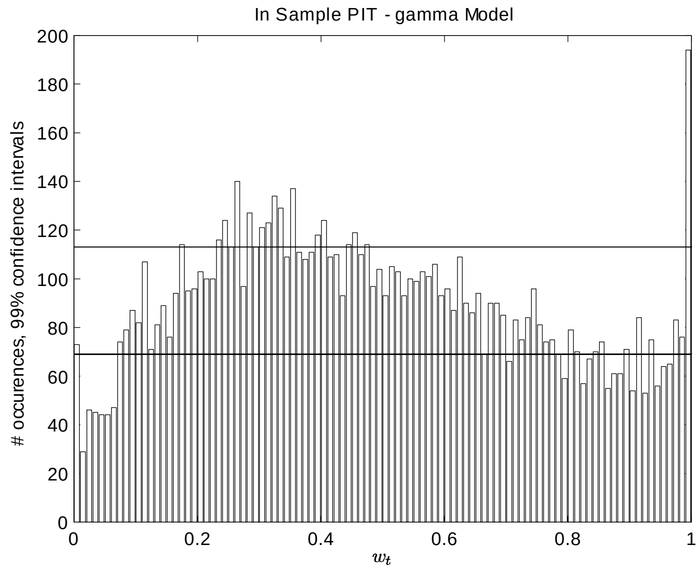



Figure 1: PITs from fitting a DCS gamma model to trade duration data for Boeing

uniformly distributed[3], only 1% of them would lie outside the range.

While an inspection of the histogram of PITs or normalized PITs is often sufficient to eliminate a distribution from further consideration, the choice between competing candidates is best made by goodness of fit criteria. The AIC or BIC may be used within the sample, while outside the sample, the predictive likelihood (sometimes called the log-score) is simple and effective. Looking at the post sample residuals, scores and PITs may also provide valuable information. Mitchell and Wallis (2011) provide a recent discussion of the issues involved.

[3]Even if the model is correct, this assumption does not hold when parameters are estimated. Thus the lines serve to indicate goodness of fit rather than providing the basis for a formal test.

# 9 Estimating volatility from the range for CAC and Dow-Jones

Although it requires monitoring the price throughout the day, the range, that is the difference between the highest and lowest log prices in a day, is very easy to obtain and has long been reported. When price movements within the day can be represented by Brownian motion, the distribution of the logarithm of the range can be shown to be approximately normal. The mean of the logarithm of the range is a linear function of volatility - as measured by scale - and so when volatility changes over time it can be extracted using the Kalman filter. Alizadeh, Brandt and Diebold (2002) study this estimator in some detail and make a convincing case for its use. To be more specific

$$\ln R_t = \lambda_t + v_t, \quad t = 1, ..., T, \tag{29}$$

where $R_t$ is the range in day $t$ and $\lambda_t$ follows an unobserved components volatility model, such as (6).

There is a weakness in the case for using a Gaussian UC model to extract volatility from the range and this stems from the fact that intra-day price movements may not be well-approximated by Brownian motion because of occasional jumps. Consequently the Gaussian approximation to the logarithm of range may not always be satisfactory. The favorable evidence for Gaussianity reported in Alizadeh, Brandt and Diebold (2002) is for exchange rates. For equities it is arguably less convincing. The DCS approach offers a wide variety of distributional options for modeling the range itself (rather than its logarithm). Just as there is no theory leading from intra-day models of price movements to $t$-distributions for daily returns, so there is no theory for suggesting what distributions for the range might arise when intra-day movements are not fully described by Brownian motion. Thus the choice between candidate distributions such as gamma, Weibull, lognormal, Burr and F is a matter for empirical investigation.

Support for using the range instead of the logarithm of the range comes from Chou (2005). He develops the Conditional Autoregressive Range (CARR) model based on the MEM approach. However, he only investigate conditional exponential, gamma and Weibull distributions. In a related study, Chou and Wang (2007) show the effectiveness of the CARR model for FTSE data.

Two datasets were used to estimate DCS models for the range. These are the Paris CAC 40 index and the Dow-Jones[4]. The slow decline in the correlograms of the raw data on range and its logarithm for both CAC and Dow-Jones indicates long memory effects that may be best captured by a two component model. The two components interpretation is that after a very large movement, there will typically be after-shocks for a few periods.

One and two component models were fitted to gamma, Weibull, lognormal, Burr, log-logistic and F-distributions. Both numerical (based on the inverse of the Hessian matrix) and analytic expressions for the standard errors of the estimates were calculated. It was found that the numerical standard errors were not always reliable and could be very dependent on starting values. As was apparent from Table 1, the analytic standard errors seem to be rather accurate for moderate sample sizes.

Table 2 shows the estimates for both indices from fitting various distributions. The parameters labeled (a) are $\gamma, \nu, \nu_1$ or $\sigma^2$, while those labeled (b) are $\varsigma$ or $\nu_2$. The last column shows three measures of goodness of fit: the maximized log-likelihood, the AIC and the BIC. In the case of the CAC, the logarithm of the range is close to being normal and hence the lognormal DCS model fits well[5]. On the other hand, as already shown, the logarithm of the range is far from being normal for Dow-Jones and the Burr, log-logistic and $F$ all fit better. The fit with the Weibull distribution was particularly bad. The reason is that in order to have a long tail, the Weibull must have considerable mass near zero and the range has near zero mass at the origin.

As might be expected from the correlograms of the raw data, fitting two first-order components, with the short-run component containing a leverage term, gave a much better fit than a simple one component model. The results are shown in Table 4.

---

[4] The data were taken from the Yahoo finance webpage. First, the range was constructed for the CAC40 index between 1st March 1990 and 17th August 2011 (approximately 5400 observations). The in-sample estimation period runs until first of January 2008. Similarly, the daily high-low range is constructed for the Dow Jones index between the 1st of October 1975 and the 17th of August 2011. The out-of-sample period starts on the 1st of January 2009.

[5] The logarithm of a log-logistic variable has a logistic distribution. This distribution is symmetric with an excess kurtosis of 6/5.

| | $\omega$ | $\phi$ | $\kappa$ | a | b | Fit |
|---|---|---|---|---|---|---|
| Gamma | -0.070 | 0.984 | 0.144 | 6.785 | - | 17279 |
| | (0.014) | (0.003) | (0.008) | (0.140) | - | -34550 |
| | (0.022) | (0.004) | (0.006) | (0.158) | - | -34555 |
| Weibull | -0.130 | 0.969 | 0.046 | 2.379 | - | 16802 |
| | (0.023) | (0.006) | (0.004) | (0.039) | - | -33597 |
| | (0.398) | (0.004) | (0.125) | (0.060) | - | -33602 |
| Lognormal | -0.061 | 0.986 | 0.152 | 0.148 | - | 17382 |
| | (0.013) | ( 0.003) | ( 0.008) | (0.000) | - | -34757 |
| | (0.038) | (0.003) | (0.012) | (0.031) | - | -34763 |
| Log-Logistic | -0.059 | 0.987 | 0.095 | 4.519 | - | 17324 |
| | (0.012) | (0.003) | (0.005) | (0.056) | - | -34641 |
| | (0.010) | (0.005) | (0.018) | (0.066) | - | -34646 |
| Burr | -0.059 | 0.987 | 0.095 | 1.000 | 4.519 | 17324 |
| | (0.013) | (0.003) | (0.006) | (0.064) | (0.116) | -34639 |
| | (0.008) | (0.001) | (0.021) | (0.011) | (0.010) | -34646 |
| F | -0.059 | 0.987 | 0.022 | 36.503 | 22.773 | 17381 |
| | (0.013) | (0.003) | ( 0.001) | (532.788) | (332.386) | -34753 |
| | (18.677) | (0.019) | (26.158) | (11.802) | ( 43.695) | -34759 |

Table 2a ML estimates for DCS models fitted to CAC data. Numbers in first brackets give analytic standard errors, figure below is the corresponding numerical SE. 'Fit' lists $\ln L$, $AIC$ and $BIC$

| | $\omega$ | $\phi$ | $\kappa$ | a | b | Fit |
|---|---|---|---|---|---|---|
| Gamma | -0.078 | 0.980 | 0.237 | 22.091 | - | 34160 |
| | (0.010) | (0.002) | (0.007) | (0.339) | - | -68313 |
| | (0.010) | (0.001) | (0.005) | (0.058) | - | -68318 |
| Weibull | -0.194 | 0.950 | 0.035 | 3.261 | - | 31724 |
| | (0.023) | (0.006) | (0.001) | (0.000) | - | -63440 |
| | (0.051) | (0.001) | (0.016) | (0.006) | - | -63445 |
| Lognormal | -0.067 | 0.983 | 0.244 | 0.045 | - | 34259 |
| | (0.009) | (0.002) | (0.008) | (0.001) | - | -68511 |
| | (0.084) | (0.001) | ( 0.134) | (0.003) | - | -68517 |
| Log-Logistic | -0.063 | 0.984 | 0.088 | 8.844 | - | 34624 |
| | (0.008) | (0.002) | (0.003) | (0.081) | - | -69240 |
| | (0.012) | (0.002) | (0.141) | (0.007) | - | -69245 |
| Burr | -0.060 | 0.985 | 0.096 | 0.757 | 9.880 | 34655 |
| | (0.010) | (0.002) | (0.007) | (0.078) | (0.423) | -69300 |
| | (0.004) | (0.002) | (0.100) | (0.005) | (0.918) | -69307 |
| F | -0.068 | 0.983 | 0.011 | 86.092 | 96.190 | 34292 |
| | (0.009) | (0.002) | (0.000) | (0.000) | (0.000) | -68574 |
| | (0.023) | (0.002) | (0.045) | (10.208) | (9.833) | -68581 |

Table 2b ML estimates for DCS models fitted to Dow-Jones data. Numbers in first brackets give analytic standard errors, figure below is the corresponding numerical SE. 'Fit' lists $\ln L$, $AIC$ and $BIC$

| | Scores | | PITs | | $\varepsilon_t$ | | |
|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE |
| | | | | CAC Range | | | |
| Gamma | 13.81 | 61.92 | 5.78 | 56.16 | 13.81 | 61.92 | 1.09 |
| Weibull | 17.12 | 52.11 | 172.12 | 500.17 | 93.57 | 295.53 | 0.97 |
| Lognormal | 9.72 | 65.67 | 6.60 | 56.54 | 24.77 | 71.54 | 1.17 |
| Log-Logistic | 6.10 | 52.92 | 6.00 | 52.83 | 74.10 | 119.18 | 1.18 |
| Burr | 6.10 | 52.89 | 6.00 | 52.83 | 74.06 | 119.14 | 1.18 |
| F | 8.71 | 66.61 | 6.37 | 56.47 | 35.19 | 81.43 | 1.20 |
| | | | | DOW Range | | | |
| Gamma | 75.18 | 169.00 | 66.23 | 161.38 | 75.18 | 169.00 | 1.02 |
| Weibull | 15.71 | 52.28 | 3212.10 | 8408.01 | 2292.53 | 5749.02 | 0.96 |
| Lognormal | 70.97 | 171.81 | 71.82 | 164.65 | 98.77 | 190.29 | 1.05 |
| Log Logisic | 70.04 | 166.25 | 69.97 | 165.94 | 189.39 | 283.25 | 1.06 |
| Burr | 71.12 | 171.25 | 70.06 | 167.45 | 237.58 | 329.87 | 1.10 |
| F | 72.60 | 172.59 | 70.41 | 163.14 | 99.03 | 191.35 | 1.05 |

Table 3a Diagnostics for one component DCS models fitted to CAC and Dow-Jones.

| | Scores | | PITs | | $\varepsilon_t$ | | | log-score |
|---|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE | |
| | | | | CAC Range | | | | |
| Gamma | 25.97 | 61.36 | 16.30 | 61.44 | 25.97 | 61.36 | 1.14 | -34566.02 |
| Weibull | 26.83 | 52.51 | 50.92 | 109.03 | 40.48 | 80.79 | 1.05 | -33613.05 |
| Lognormal | 26.55 | 66.16 | 13.75 | 57.73 | 46.96 | 78.72 | 1.24 | -34773.82 |
| Log -Logistic | 20.37 | 63.01 | 20.42 | 64.50 | 111.26 | 142.03 | 1.25 | -34657.23 |
| Burr | 20.37 | 63.00 | 20.41 | 64.49 | 111.22 | 141.98 | 1.25 | -34659.23 |
| F | 25.53 | 65.29 | 14.60 | 58.00 | 62.48 | 93.32 | 1.26 | -34773.01 |
| | | | | DOW Range | | | | |
| Gamma | 14.56 | 45.27 | 11.88 | 45.28 | 14.56 | 45.27 | 1.04 | -68329.32 |
| Weibull | 0.21 | 1.18 | 737.84 | 1259.44 | 574.30 | 819.92 | 0.95 | -63456.30 |
| Lognormal | 18.05 | 48.87 | 9.66 | 45.73 | 19.73 | 51.18 | 1.08 | -68527.93 |
| Log-Logistic | 8.26 | 48.78 | 8.27 | 49.09 | 44.05 | 76.00 | 1.09 | -69256.46 |
| Burr | 8.35 | 49.87 | 8.28 | 49.52 | 57.45 | 88.82 | 1.14 | -69320.71 |
| F | 16.47 | 47.66 | 9.30 | 45.75 | 20.04 | 51.58 | 1.07 | -68594.65 |

Table 3b Post-sample diagnostics for one component DCS models fitted to CAC and Dow-Jones.

| | $\omega$ | $\phi^L$ | $\phi^S$ | $\kappa^L$ | $\kappa^S$ | $\kappa^{Lev}$ | a | b | Fit |
|---|---|---|---|---|---|---|---|---|---|
| Gamma | -4.559 | 0.996 | 0.904 | 0.062 | 0.090 | 0.045 | 6.877 | - | 17310 |
| | (0.251) | (0.005) | (0.091) | (0.119) | (0.130) | (0.125) | (0.144) | - | -34607 |
| Weibull | -4.297 | 0.989 | 0.701 | 0.022 | 0.038 | 0.080 | 2.420 | - | 16860 |
| | (0.022) | (0.004) | (0.049) | (0.005) | (0.027) | (0.024) | (0.047) | - | -33707 |
| Lognormal | -4.818 | 0.997 | 0.953 | 0.046 | 0.104 | 0.038 | 0.146 | - | 17409 |
| | (0.141) | (0.001) | (0.012) | (0.014) | (0.015) | (0.007) | (0.000) | - | -34804 |
| Log-Logistic | -4.828 | 0.998 | 0.962 | 0.025 | 0.066 | 0.039 | 4.545 | - | 17349 |
| | (0.054) | (0.002) | (0.009) | (0.010) | (0.009) | (0.031) | (0.034) | - | -34685 |
| Burr | -4.821 | 0.998 | 0.962 | 0.025 | 0.065 | 0.039 | 4.517 | 1.018 | 17350 |
| | (0.401) | (0.001) | (0.018) | (0.197) | (0.174) | (0.097) | (0.250) | (0.011) | -34684 |
| F | -4.859 | 0.998 | 0.958 | 0.006 | 0.015 | 0.038 | 36.462 | 23.184 | 17406 |
| | (0.050) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) | (0.129) | (0.096) | -34796 |

Table 4a ML estimates for two component DCS models fitted to CAC data. Numbers in first brackets give analytic standard errors, figure below is the corresponding numerical SE. 'Fit' lists $\ln L$ and $AIC$

| | $\omega$ | $\phi^L$ | $\phi^S$ | $\kappa^L$ | $\kappa^S$ | $\kappa^{Lev}$ | a | b | Fit |
|---|---|---|---|---|---|---|---|---|---|
| Gamma | -4.013 | 0.997 | 0.866 | 0.087 | 0.166 | 0.033 | 22.609 | - | 34259 |
| | (0.075) | (0.001) | (0.016) | (0.059) | (0.066) | (0.015) | (0.313) | - | -68504 |
| Weibull | -3.922 | 0.990 | 0.744 | 0.009 | 0.037 | 0.037 | 3.279 | - | 31857 |
| | (0.071) | (0.002) | (0.028) | (0.001) | (0.004) | (0.012) | (0.037) | - | -63700 |
| Lognormal | -4.027 | 0.997 | 0.884 | 0.091 | 0.167 | 0.032 | 0.044 | - | 34345 |
| | (0.079) | (0.001) | (0.012) | (0.016) | (0.015) | (0.004) | (0.012) | - | -68676 |
| Log-Logistic | -3.983 | 0.997 | 0.909 | 0.032 | 0.058 | 0.028 | 8.922 | - | 34697 |
| | (0.004) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.008) | - | -69381 |
| Burr | -3.998 | 0.997 | 0.913 | 0.035 | 0.062 | 0.028 | 9.952 | 0.759 | 34727 |
| | (0.085) | (0.002) | (0.066) | (0.110) | (0.069) | (0.029) | (0.718) | (0.192) | -69439 |
| F | -4.022 | 0.997 | 0.884 | 0.004 | 0.007 | 0.031 | 85.361 | 101.23 | 34374 |
| | (0.029) | (0.002) | (0.005) | (0.001) | (0.001) | (0.002) | (0.457) | (0.876) | -68732 |

Table 4b ML estimates for two component DCS models fitted to Dow-Jones data. Numbers in first brackets give analytic standard errors, figure below is the corresponding numerical SE. 'Fit' lists $\ln L$ and $AIC$

Diagnostics are shown in Table 3 for the one component model and in Table 5 for two components with leverage. These are based on (i) scores; (ii) PITs; and (iii) standardized residuals, that is $y_t \exp(-\lambda_{t|t-1})$. Box-Ljung statistics are shown for all three, though it should be noted that ten lags is not enough for the chi-square approximation to be effective as the loss in degrees of freedom is relatively large, particularly in the two component model. However including more lags risks diluting the effect of any residual serial correlation. Following the suggestion of Diebold *et al* (1998), we computed the autocorrelations of the squared PITs and the corresponding Ljung-Box statistics. Since the Ljung-Box statistics were not very different from those reported for the PITs themselves, they are not shown. However, the fact that they are similar probably indicates that there are no higher-order nonlinear effects.

For CAC, the values of the coefficients show a pattern that is not unusual for volatility data. The first component is highly persistent, and little would be lost by simply setting $\phi_1$ to unity. As regards the short-term component, the effect of leverage is that $\kappa$ is often close to zero for positive returns. The PITs, and hence the scores, indicate very lttle serial correlation, but this is not the case with the standardized residuals, where some positive serial correlation remains.

The log-normal distribution gives the best fit for both one and two component models. The estimates for Burr and log-logistic are not very different, but the maximized log-likelihoods are slighly smaller. The closeness of the log-likelihoods for Burr and log-logistic is reflected in the estimates of $\varsigma$. In the two component model the estimate of $\varsigma$ is 1.018, with a SE of 0.011. Neither the likelihood ratio test nor the Wald test would reject the null hypotheses of a log-logistic distribution ($\varsigma = 1$) at any conventional level of significance.

Unlike CAC, the lognormal model does not fit DJ well and is beaten in terms of goodness of fit by Burr, log-logistic and F. However, a LR or Wald test of $\varsigma = 1$ would clearly reject the log-logistic.

Overall the Burr seems a good choice. The graphs of the PITs, shown in Appendix **??** for the two component model, support this conclusion. Although the fit at the extremes is not perfect, the graph is much closer to the uniform than it is for any of the others.

The two-component model removes some, but not all, of the serial correlation in the PITs and scores, as well as in the residuals. Although, the autocorrelations are rather small when set against those for the raw data, there may be a case for adding AR and/or MA component to the long-run and/or short-run equations. In the short-run equation there is the additional possibility of augmenting the MA term by a lagged leverage variable.

| | Scores | | PITs | | $\varepsilon_t$ | | |
|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE |
| | CAC Range | | | | | | |
| Gamma | 11.95 | 52.43 | 10.09 | 50.04 | 11.95 | 52.43 | 1.09 |
| Weibull | 20.08 | 46.57 | 140.31 | 274.51 | 90.77 | 177.81 | 0.97 |
| Lognormal | 8.60 | 57.37 | 7.39 | 50.70 | 21.43 | 60.37 | 1.17 |
| Log Logisic | 6.81 | 47.95 | 6.50 | 47.77 | 59.08 | 97.71 | 1.18 |
| Burr | 6.80 | 47.95 | 6.50 | 47.80 | 59.26 | 97.67 | 1.17 |
| F | 7.99 | 59.04 | 6.73 | 50.21 | 30.52 | 68.87 | 1.19 |
| | DOW Range | | | | | | |
| Gamma | 30.44 | 103.49 | 41.30 | 111.42 | 30.44 | 103.49 | 1.02 |
| Weibull | 15.98 | 34.40 | 3147.75 | 7524.46 | 2309.23 | 5514.21 | 0.96 |
| Lognormal | 27.67 | 108.73 | 42.83 | 116.42 | 52.46 | 127.31 | 1.05 |
| Log Logisic | 42.92 | 121.65 | 42.92 | 121.49 | 152.23 | 233.03 | 1.05 |
| Burr | 45.24 | 127.07 | 42.26 | 122.71 | 209.37 | 293.47 | 1.10 |
| F | 29.55 | 109.39 | 41.41 | 114.43 | 50.88 | 126.03 | 1.05 |

Table 5a Post-sample diagnostics for two component DCS models fitted to CAC and Dow-Jones.

| | Scores | | PITs | | $\varepsilon_t$ | | | AIC |
|---|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE | |
| | | | | CAC Range | | | | |
| Gamma | 20.93 | 54.92 | 18.91 | 60.85 | 20.93 | 54.92 | 1.12 | -6459.77 |
| Weibull | 32.00 | 60.69 | 53.25 | 95.92 | 38.75 | 74.53 | 1.02 | -6316.42 |
| Lognormal | 22.11 | 60.78 | 21.00 | 64.36 | 45.40 | 77.22 | 1.22 | -6496.26 |
| Log Logisic | 6.85 | 59.78 | 35.03 | 76.61 | 107.15 | 135.70 | 1.23 | -6441.49 |
| Burr | 16.89 | 59.81 | 34.77 | 76.51 | 105.53 | 135.17 | 1.22 | -6439.80 |
| F | 21.42 | 60.32 | 23.32 | 65.58 | 60.63 | 91.49 | 1.24 | -6466.72 |
| | | | | DOW Range | | | | |
| Gamma | 15.04 | 47.41 | 20.34 | 56.19 | 15.04 | 47.41 | 1.03 | -5065.67 |
| Weibull | 0.24 | 0.98 | 1435.52 | 3671.75 | 565.28 | 1220.34 | 0.90 | -3976.68 |
| Lognormal | 14.41 | 45.89 | 19.58 | 56.58 | 23.58 | 55.59 | 1.06 | -5187.84 |
| Log Logisic | 9.60 | 47.23 | 17.52 | 54.54 | 57.69 | 87.16 | 1.08 | -5153.20 |
| Burr | 10.09 | 48.97 | 18.71 | 56.04 | 70.98 | 100.02 | 1.13 | -5161.93 |
| F | 13.28 | 45.23 | 18.47 | 55.25 | 23.70 | 55.69 | 1.06 | -5145.81 |

Table 5b Post-sample diagnostics for two component DCS models fitted to CAC and Dow-Jones.

# 10 Duration

Duration models are widely used in financial econometrics to capture the changing intensity governing the time between events. Thus they may used, for example, to model the times between trades of an asset. In this context here is a relationship with volatility in that higher volatility tends to be associated with more trades.

Bauwens et al (2004) investigate a wide range of autoregressive conditional duration models for price, volume and trade duration data[6]. Diurnal effects are removed prior to estimation. In their conclusion they argue that price durations are perhaps the most interesting duration processes due to their close links to market microstructure and options pricing. They find that employing the basic MEM specifications with the exponential and Weibull

[6] A trade duration is given by the time interval between two consecutive trade events. A price duration is measured by the time interval between two bid-ask quotes during which a cumulative change in the mid-price of at least \$0.125 is observed. A volume duration denotes the time interval between two bid-ask quotes during which the cumulative traded volume amounts to at least 25,000 shares.

distributions is not advisable. An exponential link function gives much better results for the Weibull distribution. However, their preference is for the generalized gamma and Burr distributions, again with exponential link functions. Their 'log-ACD' specification has the conditional mean in (5) set to $\mu_{t|t-1} = \exp(\lambda^*_{t|t-1})$, where

$$\lambda^*_{t+1|t} = \delta + \beta\lambda^*_{t|t-1} + \alpha \ln y_t \qquad or \qquad \lambda^*_{t+1|t} = \delta + \beta\lambda^*_{t|t-1} + \alpha y_t \exp(-\lambda^*_{t|t-1}).$$

The first of the dynamic equations corresponds to the DCS model for a lognormal distribution, while the second is the DCS model for a gamma distribution. Neither resembles the DCS equation for any member of the generalized beta family, where the conditional score takes the form (7).

Bauwens et al (2004) reach similar conclusions regarding the best models when volume duration data is used. Table 6 shows the results of fitting various DCS models to their volume duration data[7] for Boeing. The (asymptotic) standard errors were computed analytically for the one component model. The first 1200 observations were used for estimation with the remaining 376 were reserved for post-sample evaluation. Diagnostics and post-sample diagnostics are given in tables 7 and 8 respectively. Some associated figures are shown in Appendix 12.2.

On the whole the gain from fitting a two component model is small and the diagnostics for the one component model seem perfectly fine. The Burr distribution gives the best fit, followed closely by Weibull. The Weibull shape parameter is greater than one, meaning that the distribution has a humped shape. The log-logistic distribution does not give a good fit and the hypothesis that the second shape parameter in the Burr, $\varsigma$, is unity is easily rejected using a LR test. The gamma and F-distributions[8] are only marginally worse than the Weibull, but the lognormal fit is very bad. The results are consistent with those reported by Bauwens et al (2004) for a range of companies.

[7] We are grateful to Luc Bauwens for providing us with his dataset

[8] The fitted gamma and $F$-distributions are quite close as a gamma distribution with $\gamma = \nu_1/2$ is obtained from an $F$-distribution in which $\nu_2 \to \infty$.

| | $\omega$ | $\phi$ | $\kappa$ | a | b | Fit |
|---|---|---|---|---|---|---|
| Gamma | -0.001 | 0.966 | 0.118 | 2.133 | - | -1012 |
| | (0.003) | (0.009) | (0.014) | (0.071) | - | 2033 |
| | (0.003) | (0.013) | (0.019) | (0.082) | - | 2028 |
| Weibull | 0.002 | 0.971 | 0.067 | 1.551 | - | -1014 |
| | (0.003) | (0.009) | (0.009) | (0.033) | - | 2036 |
| | (0.003) | (0.010) | (0.011) | (0.034) | - | 2031 |
| Lognormal | -0.011 | 0.961 | 0.102 | 0.589 | - | -1092 |
| | (0.004) | (0.012) | (0.014) | (0.022) | - | 2193 |
| | (0.005) | (0.015) | (0.018) | (0.889) | - | 2188 |
| Log-Logistic | -0.010 | 0.952 | 0.162 | 2.316 | - | -1064 |
| | (0.004) | (0.013) | (0.020) | (0.049) | - | 2136 |
| | (0.005) | (0.017) | (0.028) | (0.056) | - | 2131 |
| Burr | 0.036 | 0.971 | 0.082 | 1.679 | 8.006 | -1010 |
| | (0.001) | (0.001) | (0.011) | (0.012) | (0.084) | 2030 |
| | (0.013) | (0.010) | (0.012) | (0.043) | (0.751) | 2023 |
| F | -0.001 | 0.967 | 0.054 | 4.256 | 1000.03 | -1012 |
| | (0.003) | (0.009) | (0.007) | (5.419) | (1273.5) | 2035 |
| | (0.004) | (0.004) | (0.013) | (0.016) | (7.607) | 2029 |

Table 6a: ML estimates for DCS models fitted to Boeing volume duration. Numbers in first brackets give analytic standard errors, figure below is the corresponding numerical SE. 'Fit' lists $\ln L$ and $AIC$

| | $\omega$ | $\phi^L$ | $\phi^S$ | $\kappa^L$ | $\kappa^S$ | a | b | Fit |
|---|---|---|---|---|---|---|---|---|
| Gamma | -0.078 | 0.995 | 0.943 | 0.025 | 0.096 | 2.123 | - | -1012 |
| | (0.177) | (0.008) | (0.035) | (0.026) | (0.028) | (0.083) | - | 2036 |
| Weibull | 0.018 | 0.995 | 0.948 | 0.016 | 0.054 | 1.552 | - | -1013 |
| | (0.126) | (0.006) | (0.029) | (0.012) | (0.016) | (0.034) | - | 2039 |
| Lognormal | -0.312 | 0.994 | 0.934 | 0.020 | 0.087 | 0.522 | - | -1091 |
| | (0.147) | (0.010) | (0.036) | (0.021) | (0.026) | (0.001) | - | 2195 |
| Log-Logistic | -0.258 | 0.994 | 0.915 | 0.029 | 0.143 | 2.319 | - | -1063 |
| | (0.123) | (0.009) | (0.044) | (0.026) | (0.034) | (0.056) | - | 2138 |
| Burr | 1.201 | 0.995 | 0.949 | 0.019 | 0.067 | 1.680 | 8.001 | -1009 |
| | (0.412) | (0.008) | (0.033) | (0.015) | (0.029) | (0.058) | (0.751) | 2033 |
| F | -0.079 | 0.995 | 0.943 | 0.012 | 0.045 | 4.257 | 1178.5 | -1012 |
| | (0.004) | (0.001) | (0.019) | (0.002) | (0.005) | (0.108) | (3.033) | 2038 |

Table 6b: ML estimates for DCS models fitted to Boeing volume duration. Numbers in first brackets give analytic standard errors, figure below is the

corresponding numerical SE. 'Fit' lists $\ln L$ and $AIC$.

| | Scores | | PITs | | $\varepsilon_t$ | | |
|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE |
| Gamma | 12.74 | 39.77 | 15.47 | 44.96 | 12.74 | 39.77 | 1.20 |
| Weibull | 8.20 | 32.78 | 17.32 | 47.97 | 10.59 | 37.52 | 1.08 |
| Lognormal | 16.13 | 46.00 | 17.05 | 47.28 | 14.94 | 42.92 | 1.56 |
| Log Logistic | 15.13 | 46.52 | 15.26 | 46.76 | 14.07 | 43.23 | 1.46 |
| Burr | 11.58 | 38.08 | 15.72 | 44.53 | 12.67 | 39.12 | 1.33 |
| F | 12.81 | 39.97 | 15.46 | 44.97 | 12.77 | 39.81 | 1.21 |

Table 7a: In-sample diagnostics for one component DCS models fitted to Boeing volume duration

| | Scores | | PITs | | $\varepsilon_t$ | | | AIC out |
|---|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE | |
| Gamma | 7.15 | 49.17 | 6.09 | 43.02 | 7.15 | 49.17 | 1.18 | -2017.56 |
| Weibull | 10.39 | 53.37 | 9.64 | 38.34 | 9.97 | 53.25 | 1.07 | -2020.69 |
| Lognormal | 3.42 | 41.04 | 4.01 | 43.14 | 6.47 | 46.83 | 1.52 | -1939.65 |
| Log Logistic | 3.41 | 39.91 | 3.81 | 38.84 | 5.35 | 45.06 | 1.40 | -2120.69 |
| Burr | 8.72 | 50.69 | 7.33 | 44.28 | 7.97 | 50.39 | 1.38 | -2010.69 |
| F | 7.09 | 49.07 | 6.04 | 42.94 | 7.11 | 49.11 | 1.18 | -2015.86 |

Table 7b: Post-sample diagnostics for one component DCS models fitted to Boeing volume duration

| | Scores | | PITs | | $\varepsilon_t$ | | |
|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE |
| Gamma | 11.38 | 39.63 | 14.40 | 44.33 | 11.38 | 39.63 | 0.67 |
| Weibull | 7.25 | 32.96 | 16.76 | 47.95 | 9.16 | 37.64 | 1.05 |
| Lognormal | 15.31 | 46.09 | 15.56 | 45.82 | 12.46 | 40.62 | 1.27 |
| Log Logistic | 13.92 | 45.30 | 14.02 | 45.49 | 11.78 | 40.82 | 0.58 |
| Burr | 10.58 | 38.01 | 14.78 | 44.07 | 11.35 | 39.09 | 0.89 |
| F | 11.44 | 39.74 | 14.39 | 44.32 | 11.40 | 39.66 | 1.00 |

Table 8a: In-sample diagnostics for two component DCS models fitted to Boeing volume duration

| | Scores | | PITs | | $\varepsilon_t$ | | | AIC out |
|---|---|---|---|---|---|---|---|---|
| | Q(10) | Q(50) | Q(10) | Q(50) | Q(10) | Q(50) | RMSE | |
| Gamma | 7.50 | 47.14 | 5.82 | 43.01 | 7.50 | 47.14 | 1.20 | -2012.53 |
| Weibull | 10.89 | 51.81 | 9.40 | 48.35 | 10.82 | 51.98 | 1.09 | -2015.49 |
| Lognormal | 3.09 | 39.52 | 4.03 | 43.74 | 7.20 | 45.11 | 1.55 | -2171.66 |
| Log Logistic | 3.69 | 38.86 | 3.79 | 39.66 | 5.37 | 42.45 | 1.43 | -2114.48 |
| Burr | 9.16 | 48.77 | 7.06 | 44.12 | 8.39 | 48.53 | 0.33 | -2005.79 |
| F | 7.46 | 47.07 | 5.78 | 42.97 | 7.47 | 47.09 | 1.20 | -2010.79 |

Table 8b: Post-sample diagnostics for two component DCS models fitted to Boeing volume duration

One particularly interesting feature of the results is that although the maximized likelihood function for the Weibull distribution is only marginally worse than that of the Burr distribution, its shape parameter of 1.57 means that, in contrast to the Burr distribution, it does not have a heavy tail. The QQ plots indicates that there are six or seven observations that are outliers for the Weibull, but not for the Burr. The corresponding graphs for the scores tell the same story, but the outlying Weibull observations do not show up in the histogram of the PITs.

Although all Burr distributions have a heavy tail, a value of less than one for the $\varsigma$ scale parameter means that the distribution of the logarithm of the variable is skewed to the left. Figure 2 shows the histogram of the residuals from the fitted Burr model, together with the histogram of their logarithms.

The diagnostics give little indication of residual serial correlation. In contrast to the Q-statistics for the Dow-Jones range data, the Q-statistics here are all rather similar for scores, residuals and PITs. The same is true in the post-sample period.

The duration literature tends to emphasize the estimation of location, but since the full conditional distribution may be very different for different types of data, this is unwise. Furthermore, the evidence showing a poor fit for the exponential distribution cautions against the use of QML.

# 11 Conclusions

Letting the dynamics for the scale in a time series model for a non-negative variables be driven by the score yields a class of models that can be applied to

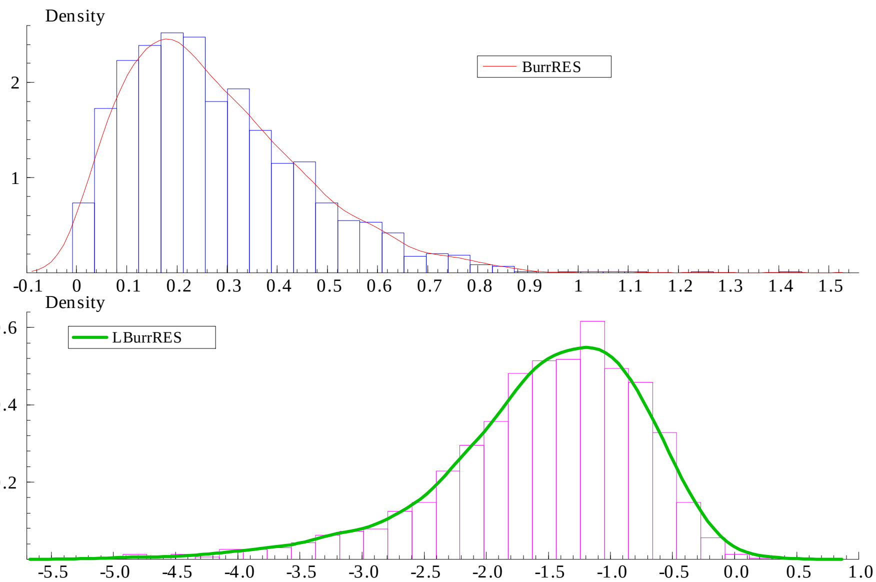



Figure 2: Histograms of Burr residuals and their logarithms for Boeing volume duration data.

a wide range of distributions. The generalized beta and generalized gamma distributions play a unifying role. The statistical properties of the models can be found because the scores are either beta or gamma distributed. For a first-order model, an analytic expression can be derived for the information matrix and we present Monte Carlo evidence showing that resulting asymptotic standard errors provide a good approximation in moderate size samples. Indeed they often appear to be more reliable than numerical standard errors.

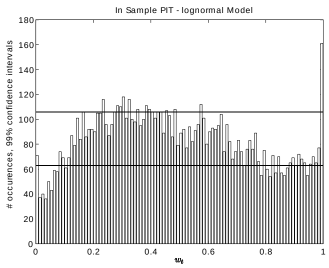
The practical value of our dynamic conditional score models was illustrated by fitting them to data on range and duration. A wide range of diagnostics were applied to check for goodness of fit of the distribution and a lack of serial correlation. The Burr distribution featured prominently for both range and duration. This has important implications for model performance since the response of dynamic conditional score models to large observations is bounded for generalized beta distributions.

Dynamic conditional score models can be used to model realized volatility, the case for their use being the same as for the range. Measures of realized volatility can be biased by market microstructure and so their logarithms may not be normally distributed. For example, Taylor (2005, pp 327-42) notes there appears to be significant skewness and kurtosis.

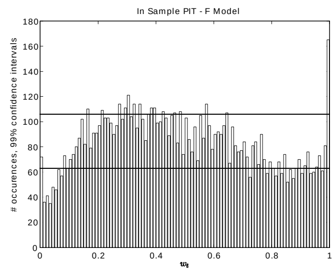
The structure of dynamic conditional score models is such that they can be extended to include time-varying trend and seasonal effects. For intra-day data, the seasonality translates into a diurnal effect; see Brownlees et al (2010, p 11). The usual approach in the literature is to remove such effects prior to any estimation. However, there is evidence to suggest that the diurnal effect is time-varying and future work will attempt to capture such effects within the model by using a limited number of trigonometric terms or by a time-varying periodic spline as in Harvey and Koopman (1993).

# 12 Appendix

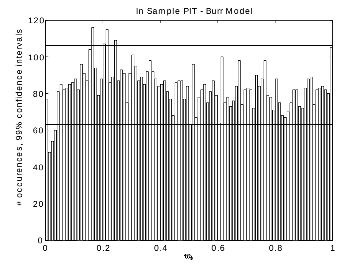
## 12.1 Diagnostic Figures Range Models



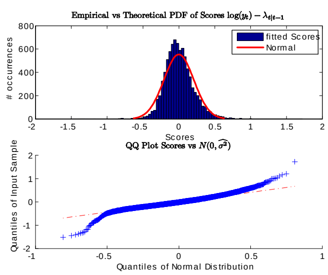

DOW PIT 2 comp DCS lognormal



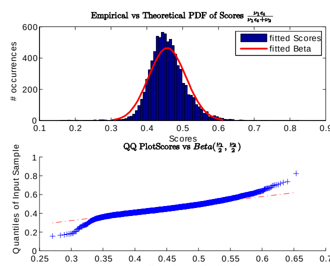

DOW PIT 2 comp DCS F



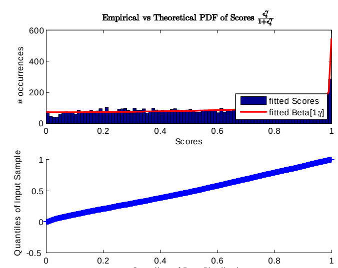

DOW PIT 2 comp DCS Burr



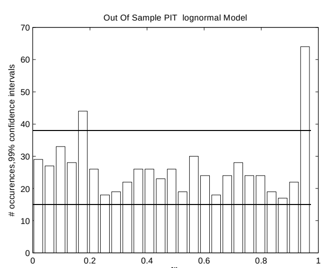

DOW Scores 2 comp lognormal



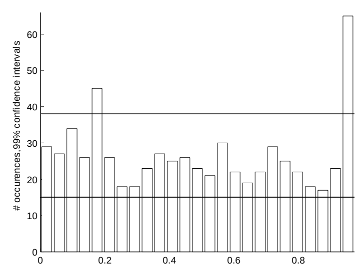

DOW Scores 2 comp F



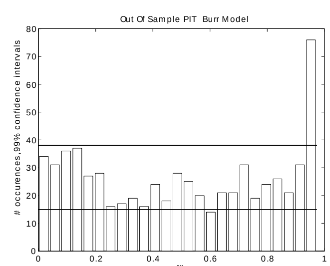

DOW Scores 2 comp Burr



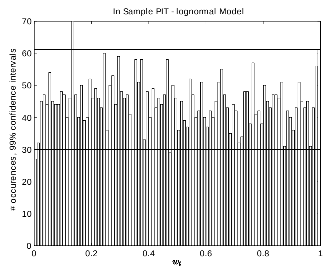

DOW 2 comp Out of sample PIT DCS lognormal



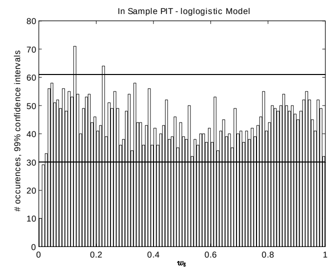

DOW 2 comp Out of sample PIT DCS F



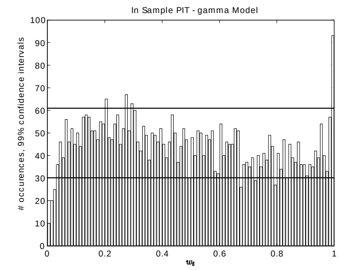

DOW 2 comp Out of sample PIT DCS Burr



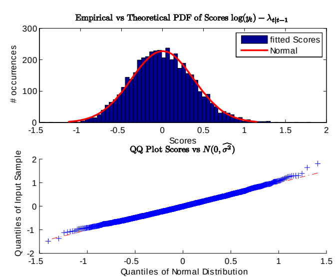

CAC PIT 2 comp DCS log-normal



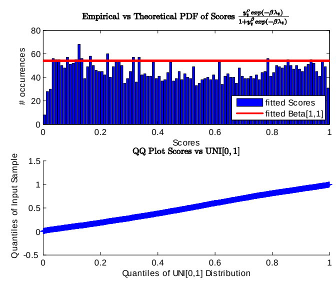

CAC PIT 2 comp DCS loglogistic



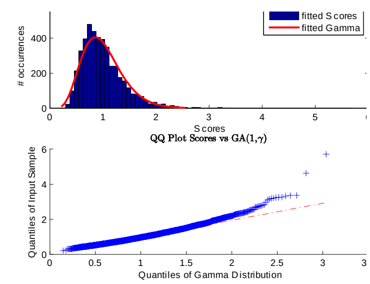

CAC PIT 2 comp DCS gamma



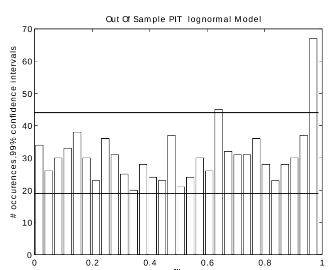

CAC Scores 2 comp DCS log-normal



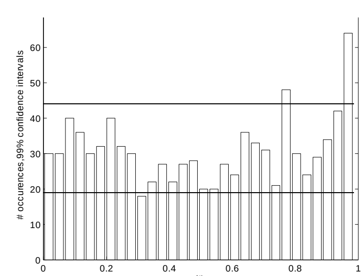

CAC Scores 2 comp DCS loglogistic



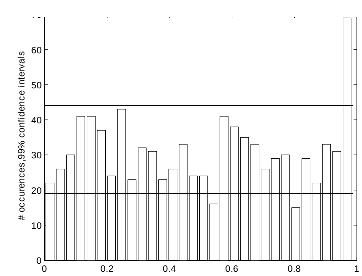

CAC Scores 2 comp DCS loglogistic




CAC Out of sample PIT DCS log-normal

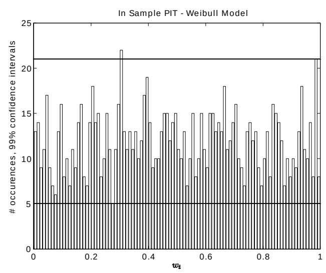



CAC Out of sample PITs 2 comp DCS loglogistic

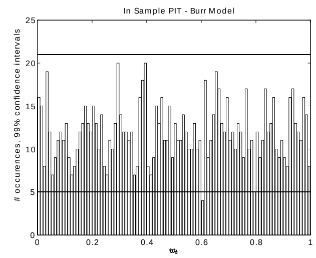



CAC Out of sample PITs 2 comp DCS loglogistic

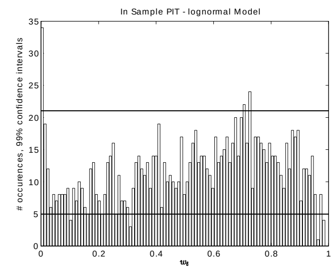
## 12.2 Diagnostic Figures Boeing Volume Duration Models



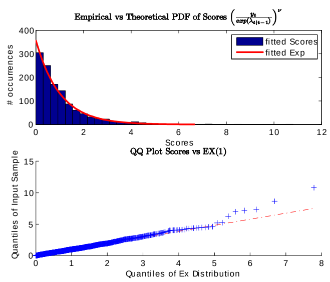

Boeing PIT 2 comp DCS Weibull



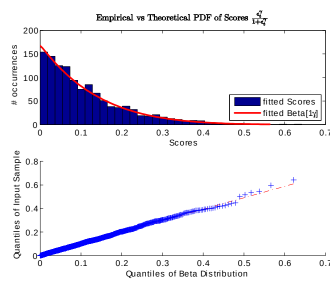

Boeing PIT 2 comp DCS Burr



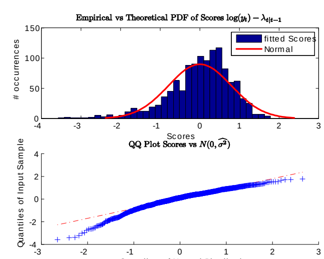

Boeing Out of sample PITs 2 comp DCS log-normal



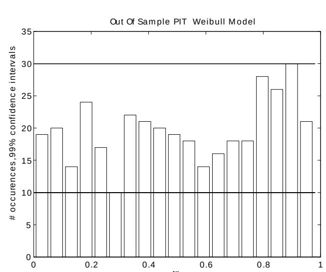

Boeing Score 2 comp DCS Weibull



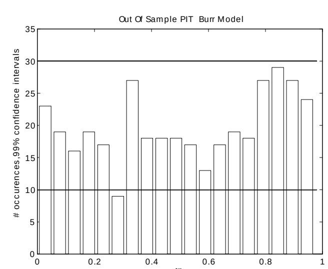

Boeing Score 2 comp DCS Burr



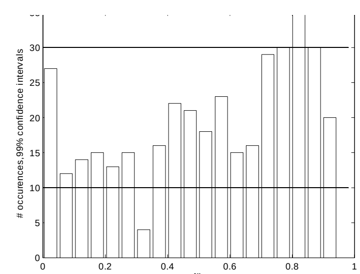

Boeing Scores 2 comp DCS log-normal




Boeing Out of sample PIT 2 comp DCS Weibull




Boeing Out of sample PIT DCS Burr




Boeing PIT 2 comp DCS log-normal

# Appendix: Asymptotic properties of the ML estimator

This appendix explains how to derive the information matrix of the ML estimator for the first-order model and outlines a proof for consistency and asymptotic normality for models from the GB2 family. The proofs of consistency and asymptotic normality for the GG family require a little more work; full details can be found in Harvey (2012).

As noted in the text, if the model is to be identified, $\kappa$ must not be zero or such that the constraint $b < 1$ is violated. A more formal statement is that the parameters should be interior points of the compact parameter space which will be taken to be $|\phi| < 1$, $|\omega| < \infty$ and $0 < \kappa < \kappa_u$, $\kappa_L < \kappa < 0$ where $\kappa_u$ and $\kappa_L$ are values determined by the condition $b < 1$.

The first step is to decompose the derivatives of the log density wrt $\boldsymbol{\psi}$ into derivatives wrt $\lambda_{t|t-1}$ and derivatives of $\lambda_{t|t-1}$ wrt $\boldsymbol{\psi}$, that is

$$\frac{\partial \ln f_t}{\partial \boldsymbol{\psi}} = \frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}, \quad i = 1, 2, 3.$$

Since the scores $\partial \ln f_t / \partial \lambda_{t|t-1}$ are $IID(0, \sigma_u^2)$ and so do not depend on $\lambda_{t|t-1}$,

$$\begin{aligned} E_{t-1}\left[\left(\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}\right)\left(\frac{\partial \ln f_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}}\right)'\right] &= \left[E\left(\frac{\partial \ln f_t}{\partial \mu}\right)^2\right] \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}'} \\ &= \sigma_u^2 \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}} \frac{\partial \lambda_{t|t-1}}{\partial \boldsymbol{\psi}'}. \end{aligned}$$

Thus the unconditional expectation requires evaluating the last term. In order to do this, we recall that the first derivative of the conditional score is as in (24), that is $-\nu^2(\xi + \varsigma)b_t(1 - b_t)$. Since, like $u_t$, this depends only on a beta variable, it is also IID. Hence the distribution of $u_t$ and its first derivative are independent of $\lambda_{t|t-1}$. All moments of $u_t$ and $\partial u_t / \partial \lambda$ exist for the t-distribution and the expressions for $a, b$ and $c$ are as in (12).

The derivative of $\lambda_{t|t-1}$ wrt $\kappa$ is

$$\frac{\partial \lambda_{t|t-1}}{\partial \kappa} = \phi \frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + \kappa \frac{\partial u_{t-1}}{\partial \kappa} + u_{t-1}, \qquad t = 2, ..., T.$$

However,

$$\frac{\partial u_t}{\partial \kappa} = \frac{\partial u_t}{\partial \lambda_{t|t-1}} \frac{\partial \lambda_{t|t-1}}{\partial \kappa},$$

Therefore

$$\frac{\partial \lambda_{t|t-1}}{\partial \kappa} = x_{t-1}\frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + u_{t-1} \tag{30}$$

where

$$x_t = \phi + \kappa \frac{\partial u_t}{\partial \lambda_{t|t-1}}, \quad t = 1, ...., T. \tag{31}$$

Taking conditional expectations of $x_t$ gives

$$E_{t-1}(x_t) = \phi + \kappa E_{t-1}\left(\frac{\partial u_t}{\partial \lambda_{t|t-1}}\right) = \phi + \kappa E\left(\frac{\partial u_t}{\partial \mu}\right),$$

where the last equality follows because $\partial u_t / \partial \lambda_{t|t-1}$ is IID and so unconditional expectations can replace conditional ones. The unconditional expression defines the general expression for the quantity '$a$' in (12).

When the process for $\lambda_{t|t-1}$ starts in the infinite past and $|a| < 1$, taking conditional expectations of the derivatives at time $t-2$, followed by unconditional expectations gives

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right) = E\left(\frac{\partial \lambda_{t|t-1}}{\partial \phi}\right) = 0 \quad \text{and} \quad E\left(\frac{\partial \lambda_{t|t-1}}{\partial \omega}\right) = \frac{1-\phi}{1-a}.$$

The derivatives wrt $\phi$ and $\omega$ are found in a similar way.

To derive the information matrix, square both sides of (30) and take conditional expectations to give

$$\begin{aligned} E_{t-2}\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right)^2 &= E_{t-2}\left(x_{t-1}\frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + u_{t-1}\right)^2 \\ &= b\left(\frac{\partial \mu_{t-1|t-2}}{\partial \kappa}\right)^2 + 2c\frac{\partial \mu_{t-1|t-2}}{\partial \kappa} + \sigma_u^2. \end{aligned} \tag{32}$$

Taking unconditional expectations gives

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right)^2 = bE\left(\frac{\partial \mu_{t-1|t-2}}{\partial \kappa}\right)^2 + 2cE\left(\frac{\partial \mu_{t-1|t-2}}{\partial \kappa}\right) + \sigma_u^2$$

and so, provided that $b < 1$,

$$E\left(\frac{\partial \lambda_{t|t-1}}{\partial \kappa}\right)^2 = \frac{\sigma_u^2}{1-b}.$$

Expressions for other elements in the information matrix may be similarly derived; see Harvey (2012). Fulfillment of the condition $b < 1$ implies $|a| < 1$. That this is the case follows directly from the Cauchy-Schwartz inequality $E(x_t^2) \geq [E(x_t)]^2$.

Consistency and asymptotic normality can be proved by showing that the conditions for Lemma 1 in Jensen and Rahbek (2004, p 1206) hold. The main point to note is that the first three derivatives of $\lambda_{t|t-1}$ wrt $\kappa$, $\phi$ and $\omega$ are stochastic recurrence equations (SREs); see Brandt (1986) and Straumann and Mikosch (2006, p 2450-1). The condition $b < 1$ is sufficient[9] to ensure that they are strictly stationarity and ergodic at the true parameter value. Similarly $b < 1$ is sufficient to ensure that the squares of the first derivatives are strictly stationary and ergodic.

Let $\boldsymbol{\psi}_0$ denote the true value of $\boldsymbol{\psi}$. Since the score and its derivatives wrt $\mu$ in the static model possess the required moments, it is straightforward to show that (i) as $T \to \infty$, $(1/\sqrt{T})\partial \ln L(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi} \to N(0, \mathbf{I}(\boldsymbol{\psi}_0))$, where $\mathbf{I}(\boldsymbol{\psi}_0)$ is p.d. and (ii) as $T \to \infty$, $(-1/T)\partial^2 \ln L(\boldsymbol{\psi}_0)/\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}' \xrightarrow{P} \mathbf{I}(\boldsymbol{\psi}_0)$. The final condition in Jensen and Rahbek (2004) is concerned with boundedness of the third derivative of the log-likelihood function in the neighbourhood of $\boldsymbol{\psi}_0$. The derivatives of $u_t$, as well as $u_t$ itself, are affine functions of terms of the form $b_t^* = b_t^h(1 - b_t)^k$, where $h$ and $k$ are non-negative integers. Since

$$b_t = h(y_t; \boldsymbol{\psi})/(1 + h(y_t; \boldsymbol{\psi})), \qquad 0 \leq h(y_t; \boldsymbol{\psi}) \leq \infty,$$

where $h(y_t; \boldsymbol{\psi})$ depends on $y_t$ and $\boldsymbol{\psi}$, it is clear that, for any admissible $\boldsymbol{\psi}$, $0 \leq b_t \leq 1$ and so $0 \leq b_t^* \leq 1$. Furthermore the derivatives of $\lambda_{t|t-1}$ must be bounded at $\boldsymbol{\psi}_0$ since they are stable SREs which are ultimately dependent on $u_t$ and its derivatives. They must also be bounded in the neighbourhood of $\boldsymbol{\psi}_0$ since the condition $b < 1$ is more than enough to guarantee the stability condition $E(\ln |x_t|) < 0$.

Unknown shape parameters, including degrees of freedom, pose no problem as the third derivatives (including cross-derivatives) associated with them are almost invariably non-stochastic.

[9] The necessary condition for strict stationarity is $E(\ln |x_t|) < 0$. This condition is satisfied at the true parameter value when $|a| < 1$ since, from Jensen's inequality, $E(\ln |x_t|) \leq \ln E(|x_t|) < 0$ and as already noted $b < 1$ implies $|a| < 1$.

# References


[1] Alizadeh, S.; Brandt, M. W. & Diebold, (2002), F. X. Range-Based Estimation of Stochastic Volatility Models Journal of Finance, 57, 1047-1091

[2] Anderson, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X.: (2006). Volatility and correlation forecasting. In: Elliot, G., Granger, C., Timmerman, A. (Eds), *Handbook of Economic Forecasting*, 777-878. Amsterdam: North Holland.

[3] Andersen, T.G., Davis, R.A., Kreiss, J-P. and T. Mikosch (2009). *Handbook of Financial Time Series.* Springer, 2009. http://www.springerlink.com/content/978-3-540-71296-1

[4] Andersen, T.G. and Bollerslev (1998). Deutsche Mark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*, 53, 219-265.

[5] Anderson, T.G. et al. (2009) Handbook of Financial Time Series, 113-155. Berlin: Springer-Verlag.

[6] Asmussen, S. (2003).Applied Probability and Queues, Berlin, Springer Verlag.

[7] Atkinson, A. C. (1987) Plots, Transformations, and Regression. Oxford Statistical Science Series

[8] Bauwens, L., and D. Veredas (2004). The stochastic conditional duration model: A latent factor model for the analysis of financial durations, Journal of Econometrics, 119, 381-412.

[9] Bauwens, L.; Giot, P.; Grammig, J. & Veredas, D. (2004) A comparison of financial duration models via density forecasts International Journal of Forecasting, 20, 589-609

[10] Bauwens, L. & Giot, P (2000) The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks Annales d'Economie et de Statistique, 06

[11] Bauwens, L. and N. Hautsh (2009) Modelling high frequency data using point processes, in Andersen, T.G. et al. *Handbook of Financial Time Series*, 953-79. Springer, 2009. http://www.springerlink.com/content/978-3-540-71296-1

[12] Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics, 19*, 465-474.

[13] Brandt, A. (1986). The stochastic equation Yn+1 = AnYn + Bn with stationary coefficients. Advances in Applied Probability ,18, 211–220.

[14] Brandt, M.W. and Jones, C.S. (2006). Volatility Forecasting with Range-Based EGARCH Models. *Journal of Business and Economic Statistics* 24, No. 4.

[15] Brown, B.M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics,* 42, 59-66.

[16] Chou, R. Y. (2005) Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range (CARR) Model Journal of Money, Credit and Banking, 37, 561-82

[17] Chou, Heng Wang, Forecasting volatility on the UK stock market: A test of the conditional autoregressive range model Source International Research Journal of Finance and Economics, 10, pp.231-257

[18] Creal, D., Koopman, S.J., and A. Lucas: (2012). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* ( to appear). Earlier version appeared as: 'A general framework for observation driven time-varying parameter models', Tinbergen Institute Discussion Paper, TI 2008-108/4, Amsterdam.

[19] Creal, D., Koopman, S.J., Lucas, A., (2008). *A general framework for observation driven time-varying parameter models.* Tinbergen Institute Discussion Paper, TI 2008-108/4, Amsterdam.

[20] Creal, D., Koopman, S.J., Lucas, A., (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations, *Journal of Business and Economic Statistics,* 29, 552-63.

[21] Davidson, J. (2000). *Econometric Theory.* Blackwell: Oxford.

[22] Deo, R., Hurwich, C. and Y. Lu (2006). Forecasting realized volatility using a long memory stochastic volatility model: estimation, prediction and seasonal adjustment. *Journal of Econometrics,* 131, 29-58.

[23] Diebold F.X., Gunther, T.A., & Tay, A.S. (1998). Evaluating density forecasts. *International Economic Review, 39,* 863-83.

[24] Doornik, J. A., (2007). *Ox: An Object-Oriented Matrix Language.* 5th edition. Timberlake Consultants Press: London.

[25] Durbin, J., Koopman, S.J., (2001). *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford

[26] Engle, R.F., Lee, G.G.J., (1999). *A long-run and short-run component model of stock return volatility.* In: Engle, R.F., White, H. (Eds.), Cointegration, causality, and forecasting: a festschrift in honour of Clive W.J. Granger. Oxford University Press, Oxford.

[27] Engle, R.F. and J G Rangel (2008). The spline-GARCH model for lowfrequency volatility and its global macroeconomic causes. Review of Financial Studies, 21, 1188-1222.

[28] Engle, R.F. and J.R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66, 1127-1162.

[29] Engle, R.F. and G.M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131, 3-27.

[30] Gonzalez-Rivera, G., Senyuz, Z. and E. Yoldas (2010). Autocontours: dynamic specification testing. *Journal of Business and Economic Statistics* (to appear).

[31] Grammig, J. and Maurer, K-O. (2000). Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal* 3, pp. 16-38.

[32] Harvey A.C., (1989). *Forecasting, structural time series models and the Kalman filter* (Cambridge: Cambridge University Press).

[33] Harvey, A.C. (2012). Exponential conditional volatility models. Revised manuscript of *Working Paper 10-36, Statistics and Econometrics Series 20, 2010. Universidad Carlos III de Madrid.*

[34] Harvey, A.C. and Chakravarty, T. (2009). *Beta-t-EGARCH.* Working paper. Earlier version appeared in 2008 as a Cambridge Working paper in Economics, CWPE 0840.

[35] Harvey, A. C., and S. J. Koopman (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of American Statistical Association* 88: 1228-36.

[36] Harvey A.C., E. Ruiz and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies* 61: 247-64.

[37] Jensen, S.T. and A. Rahbek (2004). Asymptotic inference for nonstationary GARCH. *Econometric Theory,* 20, 1203-26.

[38] Kleiber, C. and S. Kotz (2003), Statistical Size Distributions in Economics and Actuarial Sciences, Wiley, New York.

[39] Koopman, S-J., Harvey A.C., Doornik J.A., Shephard, N., (2007). *STAMP 8.0 Structural Time Series Analysis Modeller and Predictor.* Timberlake Consultants Ltd., London.

[40] Mitchell, J., Wallis, K. (2011) Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness," Journal of Applied Econometrics, vol. 26(6), pages 1023-1040, 09.

[41] Nelson, D.B., (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6, 318-24.

[42] Nelson, D.B., (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59, 347-370.

[43] Nyblom, J. and Makelainen, T. (1983). Comparison of tests for the presence of random walk coefficients in a simple linear model, *Journal of the American Statistical Association*,78, 856-64.

[44] Russell, J. R. and R. F. Engle (2010). Analysis of high-frequency data. In Ait-Sahalia, Y. and L. P.Hansen, (eds) Handbook of Financial Econometrics,Volume 1 Tools and Techniques, 383-427. Amsterdam: North-Holland Publishing.

[45] Straumann, D. and T. Mikosch (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Annals of Statistics*, 34, 2449-2495.

[46] Tadikamalla, P. R. (1980) A look at the Burr and related distributions. International Statististical Review, 48, 337-344.

[47] Taylor, S. J, (2005). *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, Princeton.

[48] Victoria-Feser, M.-P. and E. Ronchetti (1994). "Robust methods for personal income distribution models". The Canadian Journal of Statistics 22, 247–258