Non–exponential mixtures, non–monotonic financial hazard functions and the autoregressive conditional duration model

Emilio Gómez–Déniz, Jorge V. Pérez–Rodríguez
Department of Quantitative Methods. University of Las Palmas de Gran Canaria, Spain.

Abstract
This paper introduces a new specification of the standardised duration to specify and test unobserved heterogeneity in the autoregressive conditional duration model (ACD) when non-monotonic hazard distribution is allowed. Unlike other recently proposed unobserved heterogeneity models based on exponential mixtures, we propose the non-heterogeneity inverse Gaussian–ACD (IG–ACD), as an alternative to exponential, Weibull and other distributions, and the unobserved heterogeneity finite inverse Gaussian–ACD (FIGM–ACD) model, which assumes a finite mixture of an inverse Gaussian distribution with its reciprocal. IG–ACD and FIGM–ACD are easy to fit, and characterise the behaviour of the conditional durations reasonably well in terms of statistical criteria based on point and density forecasts.

Keywords: Autoregressive Conditional Duration model, Duration modelling, Inverse Gaussian Distribution

Acknowledgements

Emilio Gómez–Déniz thanks Ministerio de Ciencia e Innovación (project ECO2009–14152, MICINN). Jorge V. Pérez–Rodríguez thanks Ministerio de Economía y Competitividad (project ECO2011–23189). Also, Jorge V. Pérez–Rodríguez acknowledges the Department of Statistics and Centre for Research in Statistical Methodology (CRiSM), at the University of Warwick for their special support, since part of this paper was written while he was visiting the University of Warwick in 2013.

Address for correspondence: Jorge V. Pérez–Rodríguez, Department of Quantitative Methods in Economics, University of Las Palmas de Gran Canaria, 35017–Las Palmas de Gran Canaria, Spain. E–mail: `jvperez@dmc.ulpgc.es`

Non–exponential mixtures, non–monotonic financial hazard functions and the autoregressive conditional duration model

Abstract

This paper introduces a new specification of the standardized duration to specify and test unobserved heterogeneity in the autoregressive conditional duration model (ACD) when non-monotonic hazard distribution is allowed. Unlike other recently proposed unobserved heterogeneity models based on exponential mixtures, we propose the non-heterogeneity inverse Gaussian–ACD (IG–ACD), as an alternative to exponential, Weibull and other distributions, and the unobserved heterogeneity finite inverse Gaussian–ACD (FIGM–ACD) model, which assumes a finite mixture of an inverse Gaussian distribution with its reciprocal. IG–ACD and FIGM–ACD are easy to fit, and characterise the behaviour of the conditional durations reasonably well in terms of statistical criteria based on point and density forecasts.

Keywords: Autoregressive Conditional Duration model, Duration modelling, Inverse Gaussian Distribution

1 Introduction

The most widely-accepted econometric model for the intraday trading process is the autoregressive conditional duration (ACD) model (Engle and Russell, 1998) and its extensions; logarithmic ACD (Bauwens and Giot, 2000), stochastic conditional duration model (Bauwens and Veredas, 2004) or the stochastic volatility duration model (SVD) to cope with higher-order dynamics in the duration process (Ghysels et al., 2004), a nonlinear version based on self-exciting threshold autoregressive processes (TACD) (Zhang et al., 2001), or a family of ACD models (Box-Cox ACD) that encompass most specifications addressed in the literature (Fernandes and Grammig, 2006), and Markov-switching ACD models (Hujer et al., 2002; Calvet and Fisher, 2008; Chen et al., 2013), among others.

Since the original paper by Engle and Russell (1998) on financial duration models, there has been a gradual shift of emphasis from the estimation and interpretation of econometric results to model testing and the evaluation of the consequences of misspecification. In this respect, let us first note that various duration distributions have been proposed to estimate these models. Financial theory provides almost no basis on which to discriminate between competing models, and therefore sample information must be used. In general, it is held that the choice of a particular distribution should be guided by the aim to achieve a "correct" specification or by questions of convenience of estimation. The simplest distributional assumptions for conditional excess durations are the exponential and the Weibull distributions. However, these probability density functions (pdf) are far from capturing the most salient features of the errors, especially their variability, for example, the flexibility to accommodate certain stylised facts such as the adequacy of modelling the behaviour in the tail of the distribution (Luca and Gallo, 2009), the non-monotonic hazard function (intensity function conditional on past durations could be constant, increasing or decreasing with respect to duration; Grammig and Maurer, 2000), over-dispersion (standard deviation greater than the mean), slowly decreasing autocorrelations (Bauwens et al., 2008), or the conditions that the first and second duration moments should be finite in order to ensure that the mean and the variance both exist. Nowadays, various standardised financial duration distributions are available to estimate the parameters of ACD models: for example, Gamma, generalised Gamma, Burr, Log-normal, Pareto and Birnbaum-Saunders, among others (see Engle and Russell, 1998; Lunde, 1999; Grammig and Maurer, 2000; Bhatti, 2010; among others).

Second, tests based on ACD parameters have been developed to test the innovation distribution of ACD specifications, by gauging the distance between the parametric density and the hazard rate functions implied by the duration process, together with the nonparametric estimates (Fernandes and Grammig, 2005). Other testing frameworks for financial duration models include the density forecast evaluation technique (Bauwens et al., 2004) and the use of spectral density when testing for ACD effects, as well as for evaluating the adequacy of ACD models (Duchesne and Pacurar, 2008).

However, the most recent focus of attention with respect to the financial durations of transactions has been the unobserved heterogeneity that can be caused, for example, by differences in trading conditions, and which are not readily captured by covariates (observed heterogeneity). It is well known that duration analysis produces incorrect results if unobserved heterogeneity is ignored, because this can have serious consequences on the estimation of parameters, which can be sensitive to the presence of unobserved heterogeneity[1]. The link between statistical and financial aspects within a set of distributional assumptions is based on financial market microstructure theories. These theories divide traders into informed and non-informed, and the distribution of duration is assumed to be derived from a mixture of distributions. This basic idea is not new in microeconometric studies (Lancaster, 1990). The heterogeneity affects the hazard model in the sense that for traders who belong to distinct categories, durations might obey different probability laws. In fact, the assumption of interaction among agents, informed traders who possess private information and liquidity traders whose information set is publicly available (O'Hara, 1995; Ghysels, 2000), suggests that financial durations may obey different probability laws. On the other hand, there are many reasons to believe that arrival rates for informed and uninformed agents exhibit temporal dependence, each with its own distinct pattern.

Authors such as Luca and Zuccolotto (2003) and Luca and Gallo (2004, 2009) have proposed various mixtures of distributions to specify and estimate unobserved heterogeneity in the context of ACD models. The simplest formulation is the mixture of exponential distributions, which can be a finite

[1]Standard survival models assume homogeneity, such that all individuals are subject to the same risks embodied in the hazard or in the survivor functions. Models with covariates relax this assumption by introducing observed sources of heterogeneity. However, the presence of unobserved heterogeneity creates serious challenges for duration models because, as pointed out by Lancaster (1979) and Heckman and Singer (1984), by ignoring unobserved heterogeneity biased estimates of the hazard function may be obtained. In econometric duration research, the mixed proportional hazard model has been used extensively. This issue can be approached from various directions, such as non-parametric specifications (Heckman and Singer, 1984; Honoré, 1990; Bearse et al., 1994; among others) or by including non-parametric baseline hazards and unobserved heterogeneity parametrically (Cox, 1972; Kiefer, 1988; Meyer, 1990; among others). The results obtained by Abbring and Van den Berg (2007) rationalise the preference for the gamma distribution, and connect the results of mixed proportional hazard models with those of gamma heterogeneity as the option preferred to a wider class of models. More recently, with respect to hypothesis testing in the framework of unobserved heterogeneity, Cho and White (2010) proposed the likelihood ratio as a means of testing for unobserved heterogeneity in exponential and Weibull duration models. Unfortunately, estimators of the mixed proportional hazard model are usually biased if the functional form of the heterogeneity distribution is misspecified (Baker and Melino, 2000).

mixture, if the traders are assumed to be divided into a finite number of groups, or an infinite mixture, when every trader is considered to have his own behaviour. Therefore, the infinite mixture summarises a wide variety of agents or trading conditions, and thus different degrees of information, as well as different attitudes toward risk, budget constraints, and so on, can be taken into account, allowing for a complex unobserved heterogeneity. This context translates into a mixture provided the assumption holds of no interaction among agents: relaxing this hypothesis complicates matters and could be the object of further study. If test results indicate unobserved heterogeneity, then the crucial issue is to incorporate the mixing distribution of the heterogeneity term.

Because the parameter estimates are very sensitive to the choice of the mixing distribution (Heckman and Singer, 1984), other mixing distributions have sometimes been considered. Although there is no argument in favour of one choice over the other, we consider a distribution which can accommodate two facts: that the intensity function conditional on past durations (hazard function) is non-monotonic, and that the unobserved heterogeneity of traders can be modelled by means of finite mixtures of non-exponentials. Specifically, we consider a distribution that is widely used in studies of frailty: the inverse Gaussian distribution (Hougaard, 1984). The inverse Gaussian or Wald distribution has many applications in studies of life time, reaction times, reliability and number of event occurrences (Lancaster, 1972; Whitmore, 1975, 1979; Banerjee and Bhattacharyya, 1979; Chhikara and Folks, 1977, 1989; Jorgensen, 1982; Seshadri, 1983 and 1999; Abraham and Balakrishnan, 1998; and Balakrishnan and Nevzorov, 2003; among others), and it has been applied in fields such as economics, agriculture, demography, ecology, engineering, genetics, meteorology and the internet (Seshadri, 1999), in the study of many different topics, including financial asset returns, turbulent wind speeds, impulsive noise in radar, and radar and communication channels. It is a member of the natural exponential family of distributions and can be considered an alternative to exponential, log-Normal, log-logistic, Frechet and Weibull distributions, among others. The inverse Gaussian distribution, for example, is as suitable as the Gamma, both analytically and computationally, although it is not as widely used. Furthermore, the inverse Gaussian is a less complex alternative to the classical log-Normal model, and its hazard rate function has a $\cap$-shape like the log-Normal, generalised Weibull and log-logistic distributions, i.e. the inverse Gaussian is unimodal, which increases from 0 to its maximum value and then decreases asymptotically to a constant. It is also likely to prove useful in statistical applications as a flexible and tractable model for fitting duration data and other right-skewed unimodal data. Finally, it is a flexible closed form distribution that can be applied to model heavy-tailed processes.

The rest of this paper is structured as follows. Section 2 briefly describes the classical ACD model and presents the inverse Gaussian distribution as a suitable model in this scenario. Section 3 addresses the modelling of unobserved heterogeneity with finite mixture distributions. Section 4 presents a numerical application and, finally, some conclusions are drawn.

2 The inverse Gaussian ACD model

Engle and Russell's (1998) autoregressive conditional duration (ACD) model successfully models financial data that arrive at irregular intervals. This model is applied to the duration between two consecutive periods, $x_i = t_i - t_{i-1}$, where t_i is the time for period i, and can easily be formulated in terms of the expected conditional duration for the i–th trade: $\psi_i = E(x_i|x_{i-1}, \ldots, x_1; \theta_1)$, where $x_i = \psi_i\, \varepsilon_i$. Therefore, standardised or excess durations, $x_i/\psi_i \equiv \varepsilon_i \sim$ iid $D(\theta_2)$, where D is a general distribution defined within the interval $(0, \infty)$ with $E(\varepsilon_i) = 1$, and where θ_1 and θ_2 are vectors of unknown parameters. ψ_i is called the conditional duration and can be expressed as a linear function of past durations and lagged conditional durations. Hence, the ACD (p, q) model can be written as:

$$\psi_i = \omega + \sum_{j=1}^{q} \alpha_j\, x_{i-j} + \sum_{j=1}^{p} \beta_j\, \psi_{i-j}, \quad i = 1, 2, \ldots, N$$

where $\omega > 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$ for all j. Although not necessary, these sign restrictions are convenient to ensure the positivity of ψ_i in estimation.

Our attention in this paper is based on $D(\theta_2)$. Any distribution defined on a positive support can be specified for D. Previous studies have shown that several distributions can be used (exponential, Weibull, generalised gamma, Burr, log-normal, Pareto, Birnbaum-Saunders, among others (i.e., Engle and Russell, 1998; Lunde, 1999; Grammig and Maurer, 2000; Bhatti, 2010; among others)). However, these pdf's are far from capturing the most salient features of the errors, specifically their variability. [2] Nowadays, several choices of standardized financial duration distributions are

[2]For example, the flexibility to accommodate some stylized facts as the adequacy of

available to estimate parameters of ACD models, as for example, gamma, generalized gamma, Burr, log-normal, Pareto, Birnbaum-Saunders, among others (i.e., Engle and Russell, 1998; Lunde, 1999; Grammig and Maurer, 2000; Bhatti, 2010; among others). However, Engle and Russell (1998) show that consistent and asymptotically Normal estimates of parameters vector are obtained by maximising by QML, even if the distribution of the standardised duration, $D(\theta_2)$, is not exponential. Drost and Werker (2004) show that consistent estimates are obtained when the QML estimation method is based on the standard Gamma family (including the exponential). In this paper, we consider the inverse Gaussian distribution (Balakrishnan and Nevzorov, 2003; Chhikara and Folks, 1989; Jørgensen, 1982 and Seshadri 1983 and 1999).

2.1 The inverse Gaussian distribution

The inverse Gaussian distribution is a member of the natural exponential family of distributions. Let $\{W(t),\ t > 0\}$ be a one-dimension Wiener process with positive drift μ and variance σ^2, with $W(x) = x_0$. Then the time required for $W(t)$ to reach the value $u > x_0$ for the first time (first passage time, Schröndinger, 1915), is a random variable with inverse Gaussian distribution. A continuous random variable X follows the inverse Gaussian distribution with parameters μ and λ (henceforth $\mathcal{IG}(\mu,\lambda)$) if its probability density function is given by

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\,\pi\,x^3}}\,\exp\left[-\frac{\lambda(x-\mu)^2}{2\,x\,\mu^2}\right], \tag{1}$$

for $x > 0,\ \lambda > 0,\ \mu > 0$ and zero for $x < 0$.

Some useful properties of this distribution are given in Jørgensen (1982). Here we only reproduce the most important of them. The $\mathcal{IG}$ probability density is always positively skewed, with value $\gamma_2 = 3\sqrt{\mu/\lambda}$ and the excess kurtosis, which is $15\mu/\lambda$, is always positive. Figure 1 illustrates the shape of the $\mathcal{IG}$ distribution for selected parameter values. It is also important to note

modelling the behaviour in the tail of the distribution (Luca and Gallo, 2009), the non-monotonic hazard function (intensity function conditional on past durations could be constant, increasing or decreasing with respect to duration; Grammig and Maurer, 2000), over-dispersion (standard deviation greater than mean) and the slowly decrease of auto-correlations (Bauwens et al., 2008), or the conditions which the first and second duration moments should be finite to ensure that the mean and variance exist.

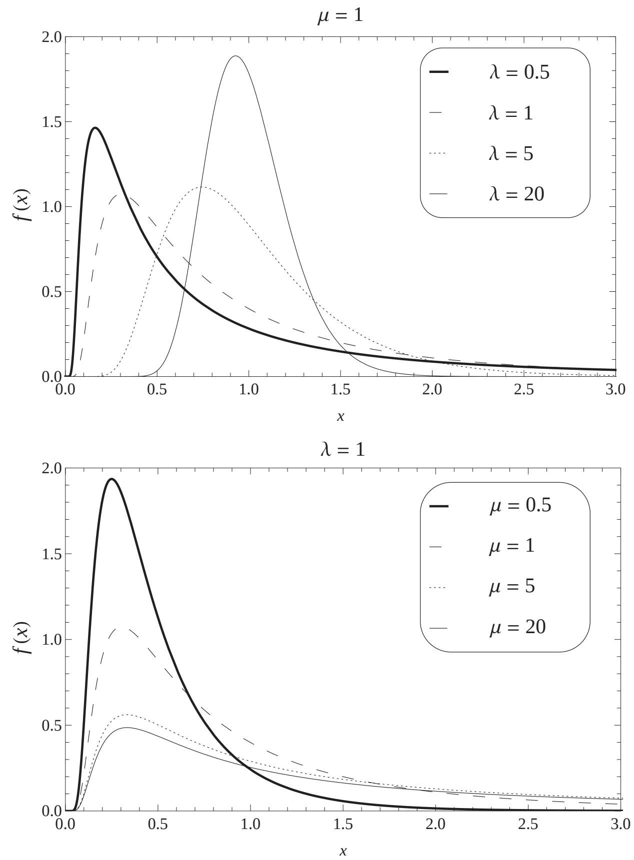

Figure 1: Pdf of the inverse Gaussian distribution for different parameter values

that the $\mathcal{IG}$ distribution is closed with respect to scale transformations since, if $X \sim \mathcal{IG}\,(\mu, \lambda)$ and $c > 0$, the transformed variable $c\,X$ has the distribution $\mathcal{IG}(c\,\mu, c\,\lambda)$. Moreover, if $x_i \sim \mathcal{IG}(\mu, \lambda)$, then $\sum_{i=1}^{n} x_i \sim \mathcal{IG}(n\,\mu, n\,\lambda)$ and furthermore, in this case, it is verified that the mean of durations $\bar{X} \sim \mathcal{IG}(\mu, n\,\lambda)$, where $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$. This property is useful because mean durations follow a $\mathcal{IG}$ distribution with the same mean as the individual distribution.

As λ tends to infinity, the inverse Gaussian distribution becomes more like a normal (Gaussian) distribution. The characteristic function is

$$\phi_X(t) = \exp\left\{\frac{\lambda}{\mu}\left[1-\sqrt{1-\frac{2\lambda^2 it}{\lambda}}\right]\right\}$$

and all positive and negative moments exist. In particular, the mean and variance of this distribution are $E(X) = \mu$ and $\mathrm{var}(X) = \mu^3/\lambda$, respectively.

The pdf of the $\mathcal{IG}$ distribution is unimodal and its mode is located in

$$x_{\text{mode}} = \mu\left(\sqrt{1+\frac{9\mu^2}{4\lambda^2}}-\frac{3\mu}{2\lambda}\right).$$

The cumulative distribution function (see Folks and Chhikara, 1978) is given by

$$F(x) = \Phi\left(t_1\sqrt{\frac{\lambda}{x}}\right) + \Phi\left(t_2\sqrt{\frac{\lambda}{x}}\right)\exp\left(\frac{2\lambda}{\mu}\right), \tag{2}$$

where $t_1 = -1 + x/\mu$, $t_2 = -1 - x/\mu$ and

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt$$

is the cumulative distribution function of the standard normal distribution.

2.2 The inverse Gaussian autoregressive conditional duration

From $Y = X/\mu$ we obtain $E(Y) = 1$ and after the changes of variable $y = \varepsilon_i = x_i/\psi_i$ we obtain

$$f(x_i|\psi_i) = \sqrt{\frac{\lambda\,\psi_i}{2\,\pi\,\mu\,x_i^3}}\,\exp\left[-\frac{\lambda(x_i-\psi_i)^2}{2\,\mu\,x_i\,\psi_i}\right], \tag{3}$$

which is an $\mathcal{IG}$ pdf where the λ–parameter is equal to $\lambda\psi_i/\mu$. Henceforth, this model is referred to as $\mathcal{IG}$–ACD (p, q).

Assuming that x_i is weakly stationary (i.e. the first two moments of x_i are time–invariant) we derive the variance of x_i in an $\mathcal{IG}$–ACD $(1, 1)$ model. It is easy to see that after the change of variable proposed $\varepsilon_i \sim \mathcal{IG}(1, \lambda/\mu)$

and then $\text{var}(\varepsilon_i) = \mu/\lambda$, which implies that the variance of the standardised durations is homoskedastic. Therefore $E(\varepsilon_i^2) = (\mu+\lambda)/\lambda$. After some computation, we obtain the unconditional mean and variance, given by[3],

$$\begin{aligned} E(x_i) &= \frac{\omega}{1-\alpha_1-\beta_1}, & (4) \\ var(x_i) &= \psi_i^2(\varepsilon_i) = \mu^2 \frac{(\mu+\lambda)(1-2\alpha_1\beta_1-\beta_1^2)+\lambda[\beta_1(2\alpha_1+\beta_1)-1]}{\lambda(1-\beta_1^2-2\alpha_1\beta_1)-\alpha_1^2(\mu+\lambda)}, & (5) \end{aligned}$$

respectively[4].

Using (2) we obtain the cumulative distribution function of the pdf (3) given by

$$F(x_i|\psi_i) = \Phi\left(t_1^i\sqrt{\frac{\lambda\,\psi_i}{x_i\,\mu}}\right) + \Phi\left(t_2^i\sqrt{\frac{\lambda\,\psi_i}{x_i\,\mu}}\right)\exp\left(\frac{2\lambda\,\psi_i}{\mu^2}\right), \qquad (6)$$

where $t_1^i = -1 + x_i/\mu$, $t_2^i = -1 - x_i/\mu$. As formally shown by Engle and Russell (1998), instantaneous intraday price volatility is linked to the conditional hazard of price durations. Therefore, expression (6) can be used to obtain the conditional intensity process on past durations of the hazard function, given by

$$\begin{aligned} h(x_i|\psi_i) &= \left\{1-\Phi\left(t_1^i\sqrt{\frac{\lambda\,\psi_i}{x_i\,\mu}}\right) - \Phi\left(t_2^i\sqrt{\frac{\lambda\,\psi_i}{x_i\,\mu}}\right)\exp\left(\frac{2\lambda\,\psi_i}{\mu^2}\right)\right\}^{-1} \\ &\times \sqrt{\frac{\lambda\,\psi_i}{2\,\pi\,\mu\,x_i^3}}\exp\left[-\frac{\lambda(x_i-\psi_i)^2}{2\,\mu\,x_i\,\psi_i}\right] \end{aligned}$$

This function is non–monotonic with respect to duration, although it depends on parameter values of μ and λ (see Figure 2). In Figure 2 (top)

[3] *i*) If $1-\beta_1^2-2\alpha_1\beta_1-\alpha_1^2 = \alpha_1^2\mu/\lambda$, we have non–stationary variance.
ii) If $1-(\alpha_1+\beta_1)^2-\alpha_1^2\mu/\lambda > 0$, the variance of durations is positive and stationary (finite).

[4] For example, in an EACD$(1,1)$ model we have

$$\begin{aligned} E(x_i) &= \frac{\omega}{1-\alpha_1-\beta_1}, \\ var(x_i) &= \mu^2\left(\frac{1-\beta_1^2-\alpha_1\beta_1}{1-\beta_1^2-2\alpha_1\beta_1-2\alpha_1^2}\right). \end{aligned}$$

given a unit mean, when the shape parameter increases, the hazard function increases too, and more quickly for large durations. In Figure 2 (bottom), given the shape parameter λ, the higher the mean, the greater the intensity of trading and the faster the hazard function increases. Therefore, it could be increasing or decreasing depending on the values of the parameters. With the exception of $\lambda = 0.5$ and $\mu = 0.5$, the shape seems to be monotonically increasing.

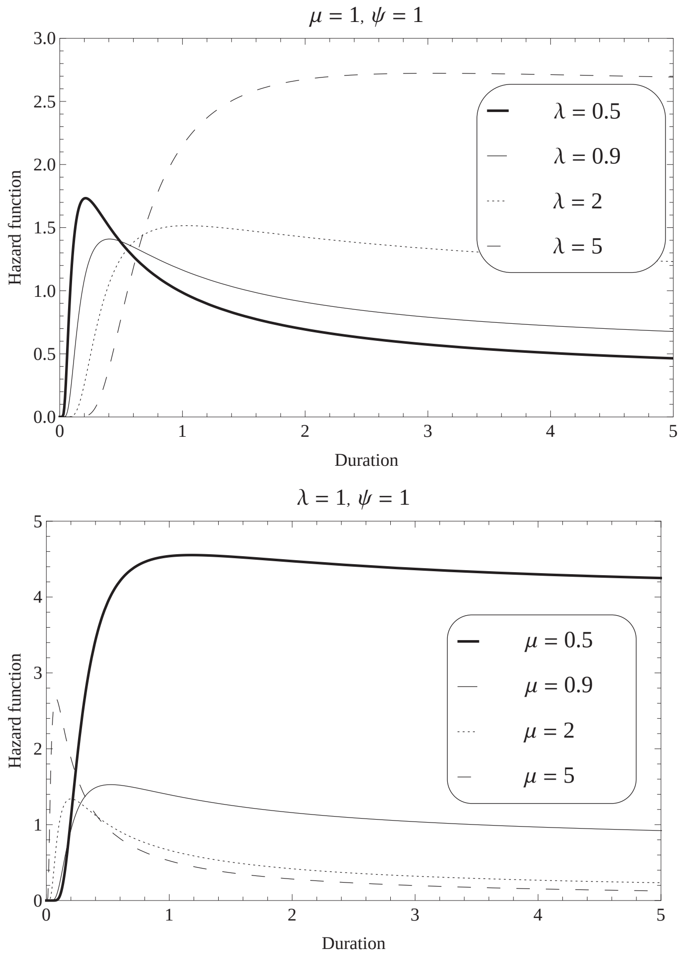

Figure 2: Hazard rate for the inverse Gaussian ACD model for different parameter values

From (3) we obtain the $\mathcal{IG}$–ACD(p,q) log–likelihood conditional function for the i-th observation, which produces a closed form, i.e. the parameters may be estimated by a numerical maximisation algorithm, such as:

$$\begin{aligned} \ell_i &= \log f(x_i|\psi_i) \\ &- frac12\,(\log\lambda + \log\psi_i - \log(2\,\pi) - \log\mu) - \frac{3}{2}\log x_i \\ &- \frac{\lambda(x_i-\psi_i)^2}{2\,\mu\,x_i\,\psi_i}, \end{aligned} \tag{7}$$

where $\ell = \sum_{i=1}^{N}\ell_i$, the first order conditions are as follows:

$$\begin{aligned} \frac{\partial\ell}{\partial\lambda} &= \frac{1}{2}\left[\frac{N}{\lambda} - \frac{1}{\mu}\sum_{i=1}^{N}\frac{(x_i-\psi_i)^2}{x_i\,\psi_i}\right], \\ \frac{\partial\ell}{\partial\mu} &= \frac{1}{2\mu}\left[-N + \frac{\lambda}{\mu}\sum_{i=1}^{N}\frac{(x_i-\psi_i)^2}{x_i\,\psi_i}\right], \\ \frac{\partial\ell}{\partial\omega} &= \frac{1}{2}\sum_{i=1}^{N}\frac{1}{\psi_i} + \frac{\lambda}{2\mu}\sum_{i=1}^{N}\frac{x_i^2-\psi_i^2}{x_i\,\psi_i^2}, \\ \frac{\partial\ell}{\partial\alpha_r} &= \frac{1}{2}\sum_{i=1}^{N}\left(\frac{1}{\psi_i} + \frac{\lambda}{2\mu}\frac{x_i^2-\psi_i^2}{x_i\,\psi_i^2}\right)\frac{\partial\psi_i}{\partial\alpha_r}, \quad r=1,2,\ldots,q, \\ \frac{\partial\ell}{\partial\beta_r} &= \frac{1}{2}\sum_{i=1}^{N}\left(\frac{1}{\psi_i} + \frac{\lambda}{2\mu}\frac{x_i^2-\psi_i^2}{x_i\,\psi_i^2}\right)\frac{\partial\psi_i}{\partial\beta_r}, \quad r=1,2,\ldots,p, \end{aligned}$$

where

$$\frac{\partial\psi_i}{\partial\alpha_r} = \sum_{j=1}^{q}\left(x_{r-j} + \beta_j\frac{\partial\psi_{r-j}}{\partial\alpha_r}\right), \tag{8}$$

$$\frac{\partial\psi_i}{\partial\beta_r} = \sum_{j=1}^{p}\beta_j\,\frac{\partial\psi_{r-j}}{\partial\beta_r}. \tag{9}$$

Remark 1 *A simpler model can be obtained by assuming that $\lambda=\mu^2$ (see Balakrishnan and Nevzorov, 2003, p. 238). In this simple case (7) is given by*

$$\log f(x_i|\psi_i) = \frac{1}{2}\,(\log\mu + \log\psi_i - \log(2\,\pi)) - \frac{3}{2}\log x_i - \frac{\mu(x_i-\psi_i)^2}{2\,x_i\,\psi_i}, \tag{10}$$

which is the log–likelihood of the first passage time to a point μ in a Brownian motion process with unit drift and unit variance, and (5) reduces to

$$var(x_i) = \mu^2 \frac{(1 - 2\alpha_1\beta_1 - \beta_1^2)(1+\mu) + \mu[\beta_1(2\alpha_1 + \beta_1) - 1]}{\mu(1 - \beta_1^2 - 2\alpha_1\beta_1) - \alpha_1^2(1+\mu)}.$$

3 Modelling the unobserved heterogeneity by finite mixtures of inverse Gaussian distributions

Economic theory does not provide much basis on which to choose a mixing distribution. Since financial market microstructure theories tend to divide the process into distinct categories (i.e., informed and uninformed) the distribution of duration is assumed to derive from a mixture of distributions.[5] Zhang et al. (2001) report that there are differences between informed trading (fast transacting regime) and uninformed trading (slow transacting regime). Accordingly, the differences between trading conditions is another argument in favour of the mixture hypothesis.

3.1 Finite mixtures of inverse Gaussian distributions

In this section, following the spirit of the work of Luca and Zuccoloto (2003), we assume that there exist two categories of Traders, and a finite mixture of the inverse Gaussian distribution is proposed. An initial distinction can be made between informed traders who possess private information with probability $1 - p$ and liquidity traders whose information set is publicly available with probability p. Given the assumption of no interaction among agents, the existence of different behaviours suggests that financial durations may obey different probability laws. Moreover, the fact of differences across trading conditions is another argument in favour of the mixture hypothesis.

Let us consider the model proposed by Jørgensen et al. (1991), where the random variable X is equal to

$$X = \begin{cases} X_1, & \text{with probability} \quad 1 - p, \\ X_2, & \text{with probability} \quad p, \end{cases}$$

[5]However, mixture models cannot model underdispersion (variance less than the mean), but this is not too restrictive as most data are overdispersed, see Cameron and Trivedi (1996).

where $X_1 \sim \mathcal{IG}(\mu, \lambda)$, $1/X_2 \sim \mathcal{IG}(1/\mu, \lambda\mu^2)$ and where $p = \mu/(\mu+\gamma) \in (0, 1)$. Here, X_2 is the complementary reciprocal of X_1.

The resulting pdf for X is then given by

$$g(x; \mu, \lambda, \gamma) = \frac{\gamma + x}{\gamma + \mu}\sqrt{\frac{\lambda}{2\pi x^3}} \exp\left[-\frac{\lambda(x-\mu)^2}{2x\mu^2}\right], \tag{11}$$

which is always unimodal (see Jørgensen et al., 1991). Henceforth, this distribution is denoted as a finite inverse Gaussian mixture, or FIGM. Furthermore, because the distribution is infinitely divisible, all moments exist and the cumulants are positive.

Another interpretation of the above distribution can be stated as follows. Consider the random variable X_3 such that $X_3 \sim \frac{\mu^2}{\lambda}\chi^2(1)$, where $\chi^2(r)$ represents the χ^2–distribution with r degrees of freedom. Then (see Jørgensen et al., 1991), the random variable X following the pdf (11) can also be rewritten as $X = X_1 + W$, where X_1 and W are independent with

$$W = \begin{cases} 0, & \text{with probability} \quad 1-p, \\ X_3, & \text{with probability} \quad p, \end{cases}$$

This result shows, therefore, that the pdf (11) is the convolution of an inverse Gaussian distribution with a compound Bernoulli distribution. A possible interpretation of this sum in our duration process is that it has two distinct phases, i.e. the duration has two different intrinsic components. Variable V is the time it takes for a trade to be initiated in the market and X_1 is the subsequent time required for the trade to develop into a complete duration. If $V = 0$, the trade is initiated immediately. If $V = X_3$ there is a random delay before trade is initiated. Jørgensen et al. (1991) pointed out that the random X_1 from the initiation might be represented by the first passage time of a Wiener diffusion process and therefore would be inverse Gaussian distributed.

If $\gamma = 0$, there is no heterogeneity because $p = 1$. Therefore, a single formal Wald–test on γ could be employed to test heterogeneity in this mixed model. When $\gamma \to \infty$ the pdf in (11) reduces to the pdf of the inverse Gaussian distribution in (1) and when $\gamma = 0$ (i.e., when there is no heterogeneity) the pdf (11) gives the distribution of the complementary reciprocal of X_1, say X_2. Figure 3 contains a plot of the pdf (11) for selected parameter values.

Henceforth, we will use FIGM(μ, λ, γ) to denote the pdf of a random variable following the pdf given in (11). Some properties of this generalisation

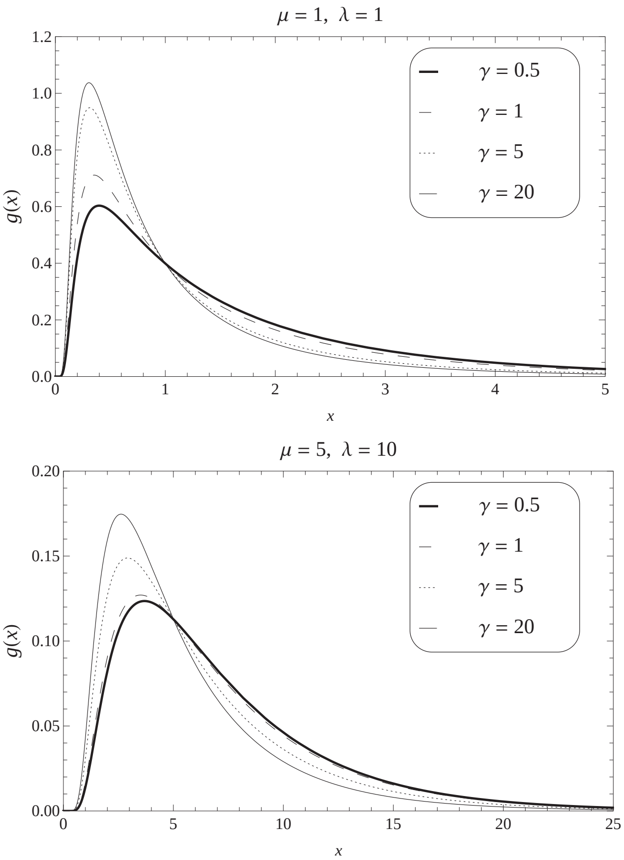

Figure 3: Pdf of the FIGM distribution for different parameter values

of the inverse Gaussian distribution are as follows. Its mean and variance are given by

$$\begin{aligned} E(X) &= \mu\left[1+\frac{\mu^2}{\lambda(1+\gamma)}\right], \qquad (12)\\ var(X) &= \mu^3\sigma^2\left[1+\frac{\sigma^2(2\mu+3\gamma)}{(\mu+\gamma)^2}\right], \end{aligned}$$

respectively.

The cdf becomes

$$G(x) = \Phi\left(t_1\sqrt{\frac{\lambda}{x}}\right) + \frac{\gamma-\mu}{\mu+\gamma}\,\Phi\left(t_2\sqrt{\frac{\lambda}{x}}\right)\,\exp\left(\frac{2\lambda}{\mu}\right).$$

3.2 The finite inverse Gaussian mixture autoregressive conditional duration model

From the Jorgensen et al. (1991) model, and after the appropriate change of variable, we obtain the pdf of the conditional duration, given by

$$g(x_i|\psi_i) = \frac{1}{\psi_i}\frac{\gamma\psi_i+\phi x_i}{\mu+\gamma}\sqrt{\frac{\lambda\psi_i}{2\pi\phi x_i^3}}\exp\left[-\frac{\lambda}{2}\frac{(\phi x_i-\mu\psi_i)^2}{\phi\psi_i x_i\mu^2}\right], \tag{13}$$

where ϕ is equal to the mean of the FIGM distribution in (12). It is easy to see that the pdf in (13) is again a finite mixture FIGM–ACD model in which the λ–parameter is equal to $\lambda\psi_i/\phi$, the μ–parameter is equal to $\mu\psi_i/\phi$ and the γ–parameter is equal to $\gamma\psi_i/\phi$. If $var(\varepsilon_i) > 1\,(< 1)$ there is conditional overdispersion (underdispersion).

The conditional hazard function for past durations is given by

$$\begin{aligned} h(x_i|\psi_i) &= \left\{1-\Phi\left(t_{11}^i\sqrt{\frac{\lambda\psi_i}{\phi x_i}}\right)+\frac{\gamma-\mu}{\mu+\gamma}\,\Phi\left(t_{22}^i\sqrt{\frac{\lambda\psi_i}{\phi x_i}}\right)\exp\left(\frac{2\lambda}{\mu}\right)\right\}^{-1} \\ &\times \frac{1}{\psi_i}\frac{\gamma\psi_i+\phi x_i}{\mu+\gamma}\sqrt{\frac{\lambda\psi_i}{2\pi\phi x_i^3}}\exp\left[-\frac{\lambda}{2}\frac{(\phi x_i-\mu\psi_i)^2}{\phi\psi_i x_i\mu^2}\right], \end{aligned}$$

where $t_{11}^i = -1 + \frac{x_i\phi}{\mu\psi_i}$ and $t_{22}^i = -1 - \frac{x_i\phi}{\mu\psi_i}$.

Figure 4 shows that the conditional hazard function for the FIGM–ACD distribution can be either increasing or decreasing, depending on the parameter values.

From (13) we obtain the finite mixture FIGM–ACD log–likelihood conditional function for the i-th observation which is also closed form, and so the parameters can be estimated straightforwardly.

$$\begin{aligned} \ell_i &= -\log\psi_i + \log(\gamma\,\psi_i+\phi\,x_i) - \log(\gamma+\mu) \\ &+ \frac{1}{2}\left[\log\lambda + \log\psi_i - \log(2\pi) - \log\phi\right] - \frac{3}{2}\log x_i \\ &- \frac{\lambda}{2}\frac{(\phi\,x_i-\mu\,\psi_i)^2}{\phi\,\psi_i\,x_i\mu^2}. \end{aligned} \tag{14}$$

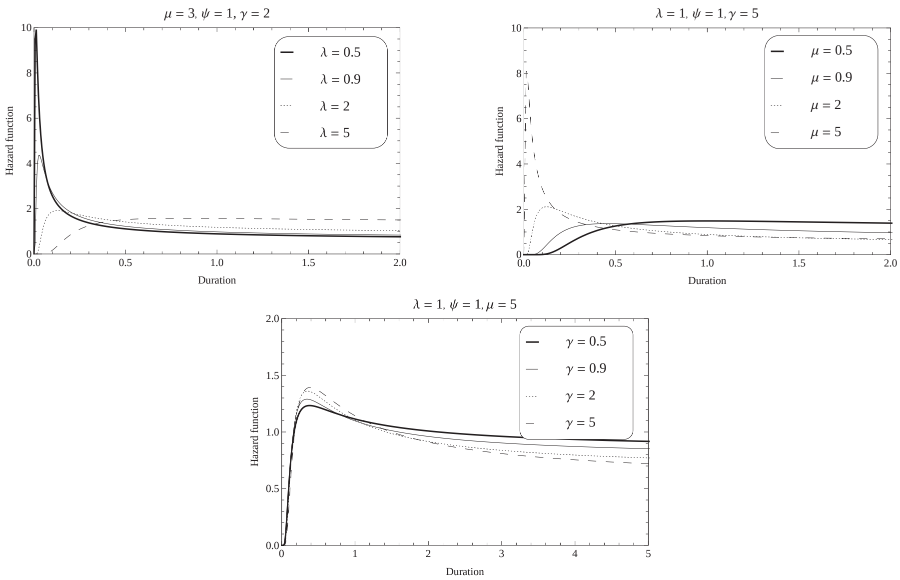

Figure 4: Hazard rate for the FIGM–ACD model for different parameter values

The first order conditions obtained from (14) for the i–th observation are:

$$
\begin{aligned}
\frac{\partial \ell_i}{\partial \lambda} &= \frac{x_i}{\gamma\psi_i - \phi\, x_i}\frac{\partial \phi}{\partial \lambda} + \frac{1}{2\lambda} - \frac{1}{\phi}\frac{\partial \phi}{\partial \lambda} \\
&- \frac{\lambda}{2}\left[\frac{1}{\lambda} + \frac{x_i\phi + \mu\psi_i}{\phi\, x_i - \mu\psi_i}\frac{\partial \phi}{\partial \lambda}\right]\frac{(\phi\, x_i - \mu\, \psi_i)^2}{\phi\, \psi_i\, x_i\mu^2}, \\
\frac{\partial \ell_i}{\partial \mu} &= -\frac{x_i}{\gamma\psi_i - \phi\, x_i}\frac{\partial \phi}{\partial \mu} + \frac{1}{\gamma + \mu} - \frac{1}{2\phi}\frac{\partial \phi}{\partial \mu} \\
&- \frac{\lambda}{2}\left[\frac{1}{\lambda} + \frac{x_i\frac{\partial \phi}{\partial \mu} + \psi_i}{\phi\, x_i - \mu\psi_i} - \frac{1}{\phi}\frac{\partial \phi}{\partial \mu} - \frac{2}{\mu}\right]\frac{(\phi\, x_i - \mu\, \psi_i)^2}{\phi\, \psi_i\, x_i\mu^2}, \\
\frac{\partial \ell_i}{\partial \gamma} &= \frac{\psi_i + x_i\frac{\partial \phi}{\partial \gamma}}{\gamma\psi_i - \phi\, x_i} - \frac{1}{\gamma + \mu} - \frac{1}{2\phi}\frac{\partial \phi}{\partial \gamma} + \frac{\lambda}{2}\frac{\partial \phi}{\partial \lambda}\frac{\mu^2\psi_i - \phi^2 x_i^2}{\phi\, \psi_i\, x_i\, \mu^2}, \\
\frac{\partial \ell_i}{\partial \omega} &= -\frac{\phi\, x_i}{\gamma\psi_i + \phi\, x_i} + \frac{1}{2\psi_i} + \frac{\lambda}{2}\frac{\mu + 1}{\psi_i}\frac{(\phi x_i - \mu\psi_i)^2}{\phi\psi_i x_i\mu^2}, \\
\frac{\partial \ell_i}{\partial \alpha_r} &= \left[-\frac{\phi\, x_i}{\gamma\psi_i + \phi\, x_i} + \frac{1}{2\psi_i} + \frac{\lambda}{2}\frac{\mu + 1}{\psi_i}\frac{(\phi x_i - \mu\psi_i)^2}{\phi\psi_i x_i\mu^2}\right]\frac{\partial \psi_i}{\partial \alpha_r},
\end{aligned}
$$

$$\frac{\partial \ell_i}{\partial \beta_r} = \left[-\frac{\phi\, x_i}{\gamma\psi_i + \phi\, x_i} + \frac{1}{2\psi_i} + \frac{\lambda}{2}\frac{\mu+1}{\psi_i}\frac{(\phi x_i - \mu\psi_i)^2}{\phi\psi_i x_i \mu^2}\right]\frac{\partial \psi_i}{\partial \beta_r},$$

where $\partial\psi_i/\partial\alpha_r$ and $\partial\psi_i/\partial\beta_r$ are given in (8) and (9), respectively, $i = 1, 2, \ldots, N$, $r = 1, 2, \ldots, q$ and

$$\begin{aligned}
\frac{\partial \phi}{\partial \lambda} &= -\frac{\mu^2}{\lambda^2(1+\gamma)}, \\
\frac{\partial \phi}{\partial \mu} &= \frac{3\mu^2 + \lambda(1+\gamma)}{\lambda(1+\gamma)}, \\
\frac{\partial \phi}{\partial \gamma} &= -\frac{\mu^3}{\lambda(1+\gamma)^2}.
\end{aligned}$$

4 An empirical example

As an illustration of the application of our specification to financial duration data, we estimated a simple ACD$(1,1)$ model using data obtained from the transaction durations of IBM stock on five consecutive trading days from November 1 to November 7, 1990, adjusted by removing the deterministic component (Tsay, 2002). The number of observations employed was 3534 positive adjusted durations.

Table 1 summarises the results of different distributions used in this paper to estimate the ACD$(1,1)$ model. We consider the exponential, Weibull, generalised gamma, Burr, inverse Gaussian with two and one-parameter distributions. Let Θ be a vector of unknown parameters to be estimated. Then, the density functions and the logarithm of the different duration models used in this paper are as follows:

$$\begin{aligned}
\textit{Exponential}: \ell(\Theta, x_i) &= \sum_{i=1}^{N}\left[-\log\psi_i - \frac{x_i}{\psi_i}\right], \\
\text{Weibull}: \ell(\Theta, x_i) &= \sum_{i=1}^{N}\left[\log\gamma - \log x_i + \gamma\log\left(\frac{x_i}{\phi_i}\right) - \left(\frac{x_i}{\phi_i}\right)^{\gamma}\right],
\end{aligned}$$

where $\phi_i = \psi_i\left[\Gamma\left(1 + 1/\gamma\right)\right]^{-1}$.

$$\text{Generalised gamma}: \ell(\Theta, x_i) = \sum_{i=1}^{N}\left[\log\theta - \log\Gamma(\gamma) + (\gamma\theta - 1)\log x_i\right.$$

$$- \quad \gamma\theta(\log\phi + \log\psi_i) - \left(\frac{x_i}{\phi\psi_i}\right)^{\theta}\Bigg],$$

where $\phi = \Gamma(\gamma)/\Gamma(\gamma + 1/\theta)$,

$$\begin{aligned} \text{Burr: } \ell(\Theta, x_i) &= \sum_{i=1}^{N} [\log\kappa - \kappa\log\epsilon_i + (\kappa - 1)\log x_i \\ &- \left(\frac{1}{\sigma^2} + 1\right)\log(1 + \sigma^2\epsilon_i^{-\kappa}x_i^{\kappa})\Bigg], \end{aligned}$$

where $\epsilon_i = \frac{\psi_i(\sigma^2)^{1+1/\kappa}\Gamma(1/\sigma^2+1)}{\Gamma\left(1+\frac{1}{\kappa}\right)\Gamma\left(\frac{1}{\sigma^2}-\frac{1}{\kappa}\right)}$.

The likelihood logarithms for the finite and infinite mixtures are:

$$\begin{aligned} \ell(\Theta, x_i) &= \sum_{i=1}^{N}\log\left[\sum_{k=1}^{K} p_k\frac{1}{\lambda_k}\exp\left(-\frac{x_i}{\psi_i\lambda_k}\right)\right] - \sum_{i=1}^{N}\log\psi_i, \\ \ell(\Theta, x_i) &= \sum_{i=1}^{N}[(\theta_1 + 1)\log\theta_1 - \log\psi_i + \log(\theta_1 + 1) \\ &- (\theta_1 + 2)\log\left(\theta_1 + \frac{x_i}{\psi_i}\right)\Bigg], \end{aligned}$$

for Luca and Zuccolotto's (2003) finite mixture and infinite mixture, respectively.

Table 1 shows the estimated parameters, with the asymptotic t–statistics corrected for the presence of unobserved heteroskedasticity, and also the value of the logarithm of maximum likelihood.

Because the tests for ACD models involve basic residual examinations, testing the functional form of the conditional mean duration or testing the distribution of the error term, we use some statistics based on standardized durations: mean, standard deviation, excess–dispersion,

$$\begin{aligned} E(x_i|\psi_i) &= \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{\psi_i}, \\ \sigma_{x_i|\psi_i} &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i}{\psi_i} - E(x_i|\psi_i)\right)^2}, \\ \text{MSE}_{x_i|\psi_i} &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \psi_i)^2, \end{aligned}$$

Table 1: Quasi maximum likelihood estimates, statistics and misspecification tests of the different ACD(1,1) models

	Exponential	Weibull	Generalised Gamma	Burr	Inverse Gaussian (2p.)	Inverse Gaussian (1p.)
γ		0.8788	4.0689			
		(78.48)	(5.66)			
θ			0.4038			
			(10.14)			
λ					0.4103	
					(9.06)	
μ					1.5367	0.2670
					(8.52)	(30.70)
κ				0.9812		
				(60.21)		
σ^2				0.1885		
				(5.81)		
ω	0.1803	0.1687	0.1411	0.1119	0.1442	0.1442
	(2.35)	(2.33)	(3.01)	(2.82)	(2.4)	(2.40)
α_1	0.065	0.0639	0.0627	0.0499	0.0449	0.0449
	(4.93)	(5.39)	(5.75)	(5.70)	(4.17)	(4.44)
β_1	0.8811	0.8852	0.8968	0.8966	0.9113	0.9113
	(27.9)	(30.61)	(47.59)	(42.69)	(36.21)	(36.65)
$\ell_{\max}$	–7688.09	–7633.65	–7583.66	–7615.51	–6973.39	–6973.39
AIC	4.3526	4.3224	4.2947	4.3127	3.9493	3.9493
SBIC	4.3579	4.3294	4.3034	4.3214	3.9580	3.9580
$E(x_i\|\psi_i)$	1.0014	1.0058	0.9894	1.2457	1.0033	1.0033
$\sigma_{x_i\|\psi_i}$	1.2276	1.2336	1.2157	1.2629	1.2321	1.2321
$\text{MSE}_{x_i\|\psi_i}$	13.701	13.658	13.860	13.373	13.599	16.350
$\text{MSE}_{x_i\|\psi_i}$	2.2906	2.2862	2.3025	2.2624	2.2887	2.7425
Excess–Dispersion	4.7728	4.8365	4.5032	5.2195	4.7780	4.7783
$Q(1)$	0.0680	0.0850	0.1566	0.146	1.236	1.236
	[0.79]	[0.77]	[0.69]	[0.70]	[0.27]	[0.27]
$Q(5)$	2.8626	2.8591	2.7562	2.9510	2.1950	2.1950
	[0.72]	[0.72]	[0.74]	[0.70]	[0.82]	[0.82]
$Q(10)$	5.1054	5.1193	5.1449	5.147	6.187	6.186
	[0.88]	[0.88]	[0.88]	[0.88]	[0.80]	[0.80]
$Q(20)$	11.2538	11.0604	10.6352	10.742	12.221	12.222
	[0.94]	[0.94]	[0.95]	[0.95]	[0.95]	[0.95]

Note: The t–values are shown in parentheses and the p–values in brackets. 1p and 2p: one and two-parameter models.

$$\text{MAE}_{x_i|\psi_i} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \psi_i|,$$

$$\text{Excess–Dispersion} = \sqrt{N}\frac{\sigma^2_{x_i|\psi_i} - 1}{\sigma_{(x_i|\psi_i - 1)^2}}.$$

And Akaike and Schwarz Bayesian information criteria (AIC and SBIC, respectively), durations ($\text{MSE}_{x_i|\psi_i}$, $\text{MAE}_{x_i|\psi_i}$), and also the Ljung–Box statistic (Q(k)) for k=2, 5,10 and 20 lags for autocorrelation in the standardized residuals shown in Table 1.

In general, these results indicate that all parameters in each estimated model are statistically significant at any conventional level of significance, but also that the standardized residual does not present serial dependence. Henceforth, misspecification tests based on the autocorrelation of standardized residuals indicate the non–presence of serial dependence. All the fitted models present a good behaviour in their standardized residuals, and so the models cannot be compared by means of this portmanteau test. On the other hand, in terms of the log–likelihood measures but also in terms of AIC and SBIC information criteria, IG fits better than other distributions. Moreover, with respect to statistical error measures, IG has a lower MSE than do other models. Finally, as can be seen, the durations are overdispersed in all models because the mean value is lower than the variance.

To analyze the effect of unobserved heterogeneity, we compared the finite and infinite mixtures of distributions based on exponentials derived by Luca and Zuccolotto (2003) and our model. Table 2) shows that the finite exponential mixture and FIGM have statistically significant parameters for unobserved heterogeneity. In the case of finite FIGM, γ is significant and in the case of Luca and Zuccolotto (2003), p_1 and p_2 are also significant at any significance level. Taking into account that liquidity traders have probabilities $\widehat{p}_2 = 0.86$ and $\widehat{p} = 0.81$, it can be seen that both models yield similar probabilities, but FIGM has a lower probability than the finite exponential mixture. Thus, $\ell_{\max}$ for FIGM is higher than the finite mixture model of Luca and Zuccolotto (2003), but it is also better in terms of AIC and SBIC information criteria. Therefore, we conclude that our model is statistically preferable.

The next subsection analyzes the forecasting capabilities of the different models used in this paper.

Table 2: Quasi–maximum likelihood estimates for ACD(1,1) models with heterogeneity. Finite and infinite mixtures.

	Finite exponential model	Infinite exponential model	Finite generalised inverse Gaussian
γ			0.0533
			(6.23)
λ			0.0859
			(44.99)
μ			0.2277
			(38.43)
p_1	0.1381		
	(5.2917)		
p_2	0.8645		
	(13.529)		
λ_1	0.6308		
	(1.9958)		
λ_2	2.6092		
	(2.1898)		
θ_1		3.7437	
		(8.81)	
ω	0.0554	0.1177	0.0596
	(1.87)	(2.71)	(2.74)
α_1	0.0260	0.0573	0.0227
	(2.10)	(5.54)	(4.31)
β_1	0.9123	0.9081	0.9181
	(45.32)	(43.77)	(41.20)
$\ell_{\max}$	–6947.05	–7613.70	–6654.77
AIC	3.9355	4.3111	3.7395
SBIC	3.9477	4.3181	3.7800
$E(x_i\|\psi_i)$	1.0527	0.9943	1.0018
$\sigma_{x_i\|\psi_i}$	2.4946	1.2095	1.4318
$\text{MSE}_{x_i\|\psi_i}$	19.1681	16.368	19.068
$\text{MAE}_{x_i\|\psi_i}$	2.5380	2.7582	2.5345
Excess–Dispersion	10.767	4.7324	10.7472
$Q(1)$	0.642	0.376	1.303
	[0.42]	[0.54]	[0.25]
$Q(5)$	2.878	3.043	2.669
	[0.72]	[0.69]	[0.75]
$Q(10)$	4.5	4.784	4.594
	[0.92]	[0.90]	[0.92]
$Q(20)$	10.547	10.658	10.806
	[0.96]	[0.95]	[0.95]

Note: The t–values are in parentheses and the p–values are in brackets.

4.1 Forecasting performance using statistical criteria

This subsection focuses on the out–of–sample forecasting ability of the different financial duration models used in Tables 1 and 2 in terms of statistical accuracy, using point forecasts and density forecasts. As is well known, in the case of ACD modelling, point forecasts do not depend on the shape of the residual distribution, which might render comparison of alternative ACD models difficult, but only provide evidence of how well a model captures the dynamics of the durations around the mean.

In general, the forecasting period is considered to be based on H observations. All the models are estimated from the first observation in the sample up to N-H-1 observations and then one–step–ahead forecasts are obtained for each model by considering the final H points.

The measures of accuracy for the one-step-ahead point forecasts used in this study are based on $h = 1, \ldots, H$ prediction periods for ψ_i, called $\widehat{\psi}_i$, and $H = 10, 25, 50, 100, 500$ and 1000 observations. This is because most empirical tests of market microstructure theories are concerned with the dynamics of (very) short or (very) long durations. For these predictions, the mean absolute error (MAE) and mean squared error (MSE) are determined.

Table 3 shows the forecast performance of both non–heterogeneity and heterogeneity models. In the first case, the exponential model seems preferable in terms of MSE for $H = 10, 25, 50, 500$. The inverse Gaussian is only preferable for MAE in $H = 10, 25$ and 100. However, with respect to unobserved heterogeneity models, the finite inverse Gaussian mixture model is always preferable to the finite exponential model, both in terms of MSE and of MAE. Although it is not usual in comparisons of duration–oriented financial forecasts, we also evaluated the equality of competitive forecasts. To do this, we used the Diebold and Mariano (DM, 1995) test, which examines whether the difference between the root MSE (RMSE) values for the two model forecasts is statistically significant. Table 4 shows the results for the modified version of the forecast evaluation DM test for non-heterogeneity models, for cases in which the null hypothesis of equal predictors is rejected. An asterisk indicates that forecasts 1 and 2 are equal at any significance level, that is, the null hypothesis cannot be rejected. However, p-values indicate that forecast 2 is better than forecast 1, that is, the null hypothesis can be rejected. We show the results corresponding to the MSE criteria applied to the difference: $d_h = e_{1h}^2 - e_{2h}^2$, where e_{1h} and e_{2h} are prediction errors formed by differences between the actual duration and its prediction, because the

Table 3: One–step–ahead forecasting results for models with and without heterogeneity. Short–and long–horizons of forecasts.

I. Short–horizon predictions

	$H = 10$		$H = 25$		$H = 50$	
	MSE	MAE	MSE	MAE	MSE	MAE
Panel A: Classic and Non–classic ACD models						
Exponential	29.2477	2.9587	27.1561	3.0612	20.8892	2.8169
Weibull	29.2989	2.9592	27.1998	3.0613	20.9352	2.8188
Generalised Gamma	29.4038	2.9699	27.3069	3.0659	21.0244	2.8243
Burr	29.7283	2.9569	27.4355	3.0416	21.3261	2.8274
Inverse Gaussian (1p)	29.6237	2.9427	27.4109	3.0343	21.1781	2.8228
Inverse Gaussian (2p)	29.6237	2.9427	27.4107	3.0343	21.1781	2.8228
Panel B: Unobserved heterogeneity in ACD models						
Finite exponential model (3p)	23.6497	3.1488	–	–	–	–
Finite Inverse Generalised inverse Gaussian model	23.6241	3.1465	21.6194	3.2436	16.0554	2.9156
Infinite exponential model	23.4041	3.1678	21.9798	3.3659	16.0302	2.9516

II. Long–horizon predictions

	$H = 100$		$H = 500$		$H = 1000$	
	MSE	MAE	MSE	MAE	MSE	MAE
Panel A: Classic and Non-classic ACD models						
Exponential	18.5476	2.6894	21.5633	2.8218	23.1841	2.8751
Weibull	18.5935	2.6916	21.6302	2.8259	17.9632	2.8578
Generalised Gamma	18.6821	2.6953	21.7397	2.8329	23.3027	2.8827
Burr	18.9817	2.7160	22.1071	2.8546	23.6847	2.9053
Inverse Gaussian (1p)	18.8237	2.6971	21.8908	2.8386	23.4829	2.8901
Inverse Gaussian (2p)	15.3272	2.6019	21.7453	2.8296	23.4829	2.8901
Panel B: Unobserved heterogeneity in ACD models						
Finite exponential model (3p)	14.1174	2.7065	20.1117	2.9152	17.7702	2.8297
Finite Inverse Generalised inverse Gaussian model	14.1149	2.7060	16.4896	2.8158	17.7747	2.8285
Infinite exponential model	14.0834	2.7323	16.3965	2.8573	17.6720	2.8695

Notes: MSE = Mean squared error, MAE = Mean absolute error. 1p, 2p and 3p represent 1, 2 and 3 parameter models

MAE criteria results indicate too much non-rejection for the null hypothesis.

Table 4: p-values for DM test with $d_h = e_{1h}^2 - e_{2h}^2$

		H					
e_1	e_2	10	25	50	100	500	1000
Weibull	Exponential	$*$	0.02	0.00	0.00	0.00	0.00
Generalised Gamma	Exponential	$*$	0.01	0.00	0.00	0.00	0.00
Burr	Exponential	$*$	$*$	0.00	0.00	0.00	0.00
Inverse Gaussian (1p)	Exponential	$*$	$*$	$*$	0.00	0.00	0.00
Inverse Gaussian (2p)	Exponential	$*$	$*$	$*$	$*$	0.00	0.00
Generalised Gamma	Weibull	$*$	0.01	0.00	0.00	0.00	0.00
Burr	Weibull	$*$	$*$	0.00	0.00	0.00	0.00
Inverse Gaussian (1p)	Weibull	$*$	$*$	$*$	$*$	0.00	0.00
Inverse Gaussian (2p)	Weibull	$*$	$*$	$*$	$*$	0.00	0.00
Weibull	Generalised Gamma	0.07	$*$	$*$	$*$	$*$	$*$
Burr	Generalised Gamma	$*$	$*$	0.00	0.00	0.00	0.00
Inverse Gaussian (1p)	Generalised Gamma	$*$	$*$	$*$	$*$	0.00	0.00
Inverse Gaussian (2p)	Generalised Gamma	$*$	$*$	$*$	$*$	$*$	0.00
Burr	Inverse Gaussian (1p)	0.07	$*$	0.00	0.00	$*$	0.00
Exponential	Inverse Gaussian (2p)	$*$	$*$	$*$	0.00	$*$	$*$
Weibull	Inverse Gaussian (2p)	$*$	$*$	$*$	0.00	$*$	$*$
Generalised Gamma	Inverse Gaussian (2p)	$*$	$*$	$*$	0.00	$*$	$*$
Burr	Inverse Gaussian (2p)	0.07	$*$	0.00	0.00	$*$	0.00

Note: An asterisk indicates that model forecasts 1 and 2 are statistically equal at any significance level.

As can be seen in Table 4, in terms of MSE criteria, inverse Gaussian forecasts are only statistically better than the other models for $H = 100$, but for other horizons the test is not conclusive, because the null hypotheses are not rejected at any significance level, with the exception of $H = 500$ and 1000, where Exponential, Weibull and generalised Gamma are always better than the inverse Gaussian models. On the other hand, with respect to the class of non–monotonic hazard distribution models, that is, the Burr model, the inverse Gaussian model (1p and 2p) is always preferable for $H = 10, 50, 100$ and 1000 observations.

The results for the unobserved heterogeneity models are shown in Table 5. Only the results for $H = 500$ and 1000 predictions are shown because for the remaining forecast periods the results indicate non–rejection of the null hypothesis. In all cases, according to the DM test the finite inverse Gaussian mixture is better than finite and infinite mixtures of exponentials, at different conventional significance levels.

We conclude, thus, that in terms of classical forecast evaluation criteria, the inverse Gaussian has a slight advantage with regard to other unobserved heterogeneity models, and the non-heterogeneity Burr distribution in all cases. As regards the other models, exponential forecasts seems to be better than the others for several prediction horizons. Finally, it is noteworthy, as also observed by Hujer et al. (2002), that the models which allow unobserved heterogeneity provided better forecast results in terms of MSE than did the ACD models without unobserved heterogeneity, in both short and long-run predictions. Thus, point forecasts do not tell us much about the suitability of the model in question, particularly regarding the appropriateness or otherwise of the residual distribution. Therefore, for a more comprehensive assessment of financial duration models, we employ the method of density

Table 5: p-values for DM test for predictions of unobserved heterogeneity models.

		H	
e_1	e_2	**500**	**1000**
Panel A: $d_h = e_{1h}^2 - e_{2h}^2$			
Finite exponential	Finite inverse Gaussian mixture	0.00	0.05
Finite exponential	Infinite exponential	0.00	0.04
Panel B: $d_h = \|e_{1h}\| - \|e_{2h}\|$			
Finite exponential	Finite inverse Gaussian mixture	0.05	0.10
Finite exponential	Infinite exponential	0.02	0.00
Infinite exponential	Finite inverse Gaussian mixture	0.00	0.00

forecasts. This technique can be employed to evaluate forecasts from both nested and non-nested models.

A density forecast is defined as the density for the next observation of the variables of interest. This method was derived by Diebold et al. (1998), based on Rosenblatt (1952). This approach allows us to evaluate forecasts without having to specify a loss function, and thus improves over most standard techniques for evaluating point forecasts, which typically assume quadratic loss. In our case, this density is implied by each model shown in Tables 1 and 2, and for the horizons defined by 100 and 1000 observations (for example, we use the first 2534 observations of the sample for estimation, and the last 1000 for out–of–sample density forecast evaluation).

To do this, we applied the probability integral transform (PIT): $z_i = \int_{-\infty}^{x_i} f(u|\psi_i)\, du$, where x_i is the duration realized and represents the one–step ahead conditional forecast density produced by the financial duration model. For example, in the exponential case $f_i(x_i|\psi_i) = \frac{1}{\psi_i}\exp(-x_i/\psi_i)$. Under the null hypothesis that the sequence of forecast densities $\{z_i\}_{i=1}^n$ is equal to the sequence defining the data generating process of x_i (which cannot be observed), that is, that the sequence of density forecasts is correctly specified, then $\{z_i\}_{i=1}^n \sim \text{iid}U(0,1)$. As Diebold et al. (1998) point out, this is a joint hypothesis which makes it difficult to distinguish the causes of a possible rejection. This suggests that it is necessary to evaluate the density forecasts $\{f_i(\cdot)\}_{i=1}^n$ by testing the hypothesis of iid U(0,1) for the sequence $\{z_i\}_{i=1}^n$. To do so, we can consider a number of tests of iid U(0,1), such as a simple inspection of the autocorrelation functions to analyse the iid hypothesis and histograms and QQ–plots of $\{z_i\}_{i=1}^n$ to assess uniformity. All of these measures could inform us about the goodness of the density forecasts.

Specifically, to test the iid assumption of $\{z_i\}_{i=1}^n$ –or even the presence of non-linear dependence in higher powers, for example, $\{z_i^2\}_{i=1}^n$–, we can employ the Ljung–Box statistic (Q), both for the level and the square of the sequence. These results are omitted by parsimony, but are available from the authors on request. In general, the test for serial independence of the z sequences yielded good results for all non-heterogeneity and heterogeneity ACD specifications, irrespective of the innovation distribution assumed. That is, the iid assumption for the z and the squared sequences cannot be rejected, at conventional significance levels, by employing different lags for LBQ tests.

We only report the results of histograms and QQ–plots for unobserved heterogeneity models, because the results for non-heterogeneity models indicate that these are not correctly specified, that is, the sequence $\{z_i\}_{i=1}^n$ is not iid U(0,1). For example, we can say that for $H = 1000$ predictions, the histograms suggest that density forecasts are under–estimated for Weibull (i.e., higher values for density or relative frequency at low values of probability integral transform, z_i), and over–estimated for exponential, generalised gamma, Burr and inverse Gaussian with one or two parameters. In other words, the probability integral transforms are not U(0,1).

Figure 5 shows the histogram for each unobserved heterogeneity financial duration model, which is scaled using the relative frequency and the bin width, by the Freedman-Diaconis method. In addition, we plot the theoretical U(0,1) density in a fill area behind the empirical histogram. In all figures, a comparison of the fill area for the U(0,1) distribution and the histogram shows that they are close to the U(0,1) assumption. Nevertheless, the finite exponential and finite inverse Gaussian mixtures have under–estimated forecasts. However, empirical distribution tests such as Kolmogorov-Smirnov, Anderson-Darling or Cramer-von Mises tests, reject null hypothesis of U(0,1) at 5

Finally, Figure 6 shows the QQ–plots (by Cleveland's method) between empirical and theoretical quantiles of U(0,1) for unobserved heterogeneity financial duration models. These indicate that the distributions lie on the straight line. However, the infinite exponential may be more accurate than the straight line for $H = 1000$ observations.

The main conclusion drawn with respect to density forecasts is that the unobserved heterogeneity models perform better than the non-heterogeneity ones because predictions are iid U(0,1); moreover, an infinite exponential mixture of distributions is better than a finite mixture, at least according to $H = 1000$ prediction results. However, these results cannot be considered conclusive.

5 Conclusions

This paper proposes a new specification of the disturbance in the autoregressive conditional duration model (ACD) to specify and test unobserved heterogeneity. It is assumed that standardized durations allow non-monotonic hazard distribution (i.e., constant, increasing or decreasing with respect to duration).

Unlike other recently proposed unobserved heterogeneity models based

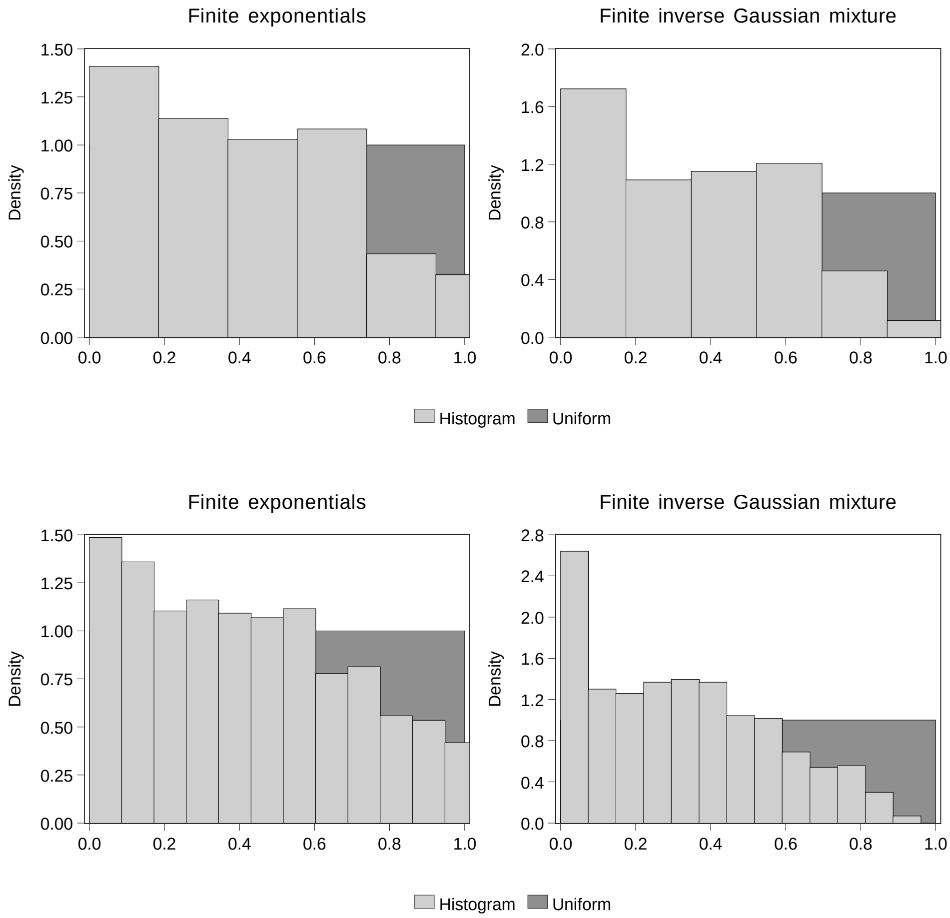

Figure 5: Histograms of the sequence of $\{z_i\}_{i=1}^{n}$ for unobserved heterogeneity financial duration models and the case of $H = 100$ (above) and $H = 1000$ (below) predicted observations.

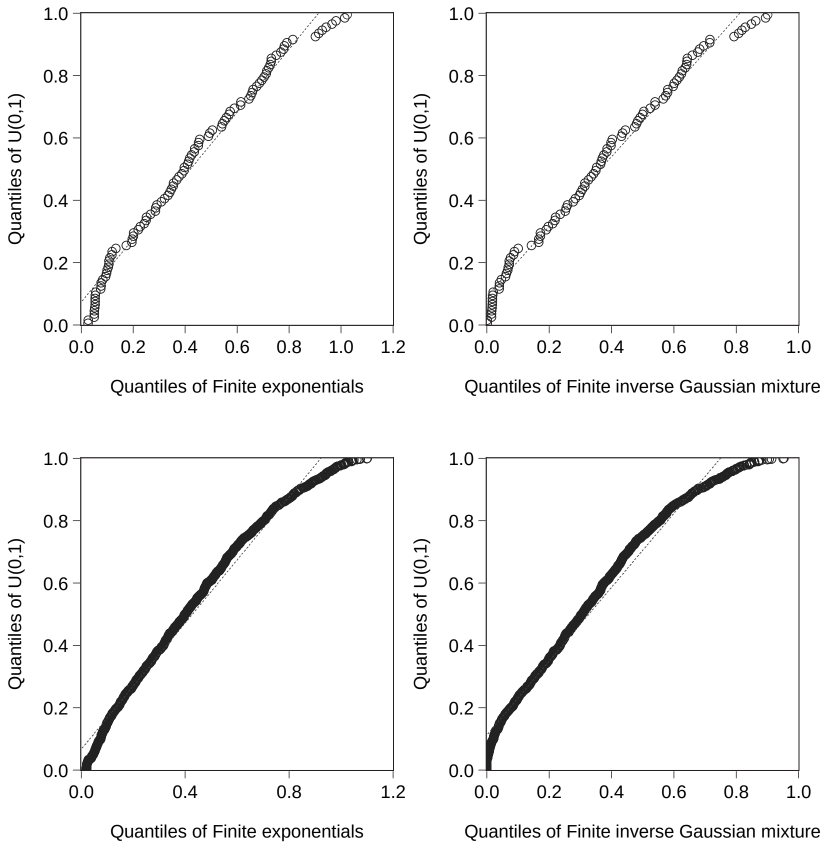

Figure 6: QQ-plots of the sequence of $\{z_i\}_{i=1}^n$ for all financial duration models and the case of $H = 100$ (above) and $H = 1000$ (below) predicted observations.

on exponential mixtures (Luca and Zuccolotto, 2003; Luca and Gallo, 2004, 2009), we propose the non-heterogeneity inverse Gaussian–ACD (IG–ACD) and an unobserved heterogeneity finite inverse Gaussian–ACD (FIGM–ACD) model. In the latter model, we assume the finite mixture of an inverse Gaussian distribution with its reciprocal, as proposed by Jørgensen et al, 1991, and we derive its statistical properties in the ACD(1,1) model.

The models we propose are easy to fit, and characterize the behaviour of the conditional durations reasonably well. Therefore, by using statistical measures and autocorrelation tests on standardized residuals based on in-sample estimates, all models are correctly specified, and the IG models perform better than those without heterogeneity, such as exponential, weibull, generalised gamma and Burr, and also, with regard to distributions, than the unobserved heterogeneity model based on exponential mixtures.

Finally, based on out–of–sample predictions and one-step-ahead forecast accuracy measures obtained for several prediction horizons, we only conclude - in terms of point forecasts - that the forecasts obtained by IG–ACD models are slightly better than those obtained from unobserved heterogeneity models based on exponentials. On the other hand, as regards density forecasts, the results of probability integral transforms for non-heterogeneity models (in all the horizons analysed), indicate that the models are not correctly specified. With respect to unobserved heterogeneity, although probability integral transforms are iid U(0,1), which indicates that the model is specified correctly, we cannot clearly determine any model as preferable. Therefore, these results are not conclusive.

References

Abbring, J.H. and Van den Berg, G. (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika*, 94, 1, 87–99.

Abraham, B and Balakrishnan, N. (1998). Inverse Gaussian autoregressive models. *Working Paper, RR–98–09*, University of Waterloo.

Anscombe, F. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, 37, 358–382.

Baker, M. and Melino, A. (2000). Duration dependence and nonparametric heterogeneity: a Monte Carlo study. *Journal of Econometrics*, 96, 357–393.

Balakrishnan, N. and Nevzorov, V.B. (2003). *A Primer on Statistical Distributions.* John Wiley and Sons, New York.

Banerjee, A.K. and Bhattacharyya, G.K. (1976). A purchase incidence model with inverse Gaussian interpurchases times. *Journal of American Statistical Association*, 71, 823–829.

Bauwens, L. and 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*, 60, 117–150.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. *Journal of Econometrics*, 119, 381–412.

Bauwens L., Giot P., Grammig J. and Veredas D.A. (2004). Comparison of financial duration models via density forecasts. *International Journal of Forecasting*, 20(4), 589–609.

Bearse, P., Canals, J. and Rilstone, P. (1994). Semiparametric maximum likelihood estimation of duration models with unobserved heterogeneity. *Discussion paper*, Department of Economics, York University.

Bhatti, C.R. (2009). On the interday homogeneity in the intraday rate of trading. *Mathematics and Computers in Simulation*, 79, 2250–2257.

Bhatti, C.R. (2010). The Birnbaum-Saunders autoregressive conditional duration model. *Mathematics and Computers in Simulation*, 80, 2062–2078.

Breslow, N. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, 33, 38–44.

Brock,W., Dechert,W., Scheinkman, J. and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15, 197–235.

Calvet, L. and Fisher, A. (2008). *Multifractal Volatility: Theory and Pricing.* Elsevier.

Cameron, A. and Trivedi, P. (1996). *Count data models for financial data.* Handbook of Statistics. Statistical Methods in Finance, North–Holland.

Chen, Y. and Kuan, C. (2002). Time irreversibility and EGARCH effects in us stock index returns. *Journal of Applied Econometrics*, 17, 565–578.

Chhikara, R.S. and Folks, J.L. (1977). The inverse Gaussian distribution as a life time model. *Technometrics*, 19, 461–468.

Chhikara, R. and Folks, T. (1989). *The inverse Gaussian Distribution. Theory, Methodology, and Applications.* Marcel Dekker, New York.

Chen, F., Diebold, F.X. and Schorfheide, F. (2013). A Markov-switching multifractal inter–trade duration model, with application to US equities. *Journal of Econometrics, forthcoming.* Available on line 17 April 2013.

Cho, J. and White, H. (2010). Testing for unobserved heterogeneity in exponential and Weibull duration models. *Journal of Econometrics*, 157, 2, 458–480.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society: Series B*, 34, 187–220.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39, 863–883.

Drost, F.C. and Werker, B.J.M. (2004). Semiparametric duration models. *Journal of Business and Economic Statistics*, 22, 40–50.

Duchesne, P. and Pacurar, M. (2008). Evaluating financial time series models for irregularly spaced data: A spectral density approach. *Computers & Operations Research*, 35, 130–155.

Engel, J. (1984). Models for response data showing extra-Poisson variation. *Statistical Neerlandica*, 38, 159–167.

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

Fernandes, M. and Grammig, J. (2005). Nonparametric specification tests for conditional duration models. *Journal of Econometrics*, 127, 35–68.

Fernandes, M. and Grammig, J. (2006). A family of autoregressive conditional duration models. *Journal of Econometrics*, 130, 1–23.

Ghysels, E., Gouriroux, C. and Jasiak, J. (2004). Stochastic volatility duration. *Journal of Econometrics*, 119, 413–433.

Giot, P. (2000). Time transformations, intraday data and volatility models. *Journal of Computational Finance*, 4, 31–62.

Ghysels, E. (2000). Some Econometric Recipes for High-Frequency Data Cooking. *Journal of Business & Economic Statistics*, 18, 154–163.

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

Hautsch, N. (2006). Testing the conditional mean function of autoregressive conditional duration models. *Finance Research Unit*. Department of Economics, University of Copenhagen.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52, 271–320.

Honoré, B. (1990). Simple estimation of a duration model with unobserved heterogeneity. *Econometrica*, 58, 453–473.

Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, 71, 75–83.

Hujer, R., Vuletic, S. and Kokot, S. (2002). The Markov switching ACD model. *Working paper series: Finance & Accounting*, 90, University of Frankfurt.

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution.* Lecture Notes in Statistics, 9. Springer–Verlag, New York.

Jørgensen, B., Seshadri, V. and Whitmore, G.A. (1991). On the mixture of the inverse Gaussian distribution with its complementary reciprocal. *Scandinavian Journal of Statistics*, 18, 77–89.

Kiefer, N. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, 26, 646–679.

Lancaster, T. (1972). A stochastic model for the duration of a strike. *Journal of the Royal Statistical Society, Series A (General)*, 135, 2, 257–271.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47, 939–956.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press.

Li,W.K. and Yu, P. (2003). On the residual autocorrelation of the autoregressive conditional duration model. *Economics Letters*, 79, 169–175.

Luca, G. De and Zuccolotto, P. (2003). Finite and infinite mixtures for financial durations. *Metron*, 61, 431–455.

Luca, G. De and Gallo, G. M. (2004). Mixture processes for intradaily financial durations. *Studies in Nonlinear Dynamics and Econometrics*, 8(2), 1–18.

Luca, G. and Gallo, G.M.(2009). Time–varying mixing weights in mixture autoregressive conditional duration models. *Econometric Reviews*, 28: 1, 102–120.

Lunde, A. (1999). A generalized gamma autoregressive conditional duration model. *Discussion Paper*, Aalborg Universiteit.

Meitz, M. and Teräsvitrta, T. (2006). Evaluating model of autoregressive conditional duration. *Journal of Business and Economics Statistics*, 24, 1, 104–124.

Meyer, B. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58, 757–782.

O'Hara, M. (1995). *Market Microstruture Theory.* Basil Blackwell Inc., Oxford.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathemarical Statistics*, 23, 470–472.

Seshadri, V. (1983). The inverse Gaussian distribution: Some properties and characterizations. *The Canadian Journal of Statistics*, 11, 131–136.

Seshadri, V. (1999). *The Inverse Gaussian Distribution: A Case Study in Exponential Families.* Oxford Science Publications.

Schrödinger, E. (1915). Zur theorie der fall-und steigversuche an teilchenn mit Brownscher bewegung. *Physikalische Zeitschrift*, 16, 289–295.

Stein, G., Zucchini, W. and Juritz, J. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82, 339, 938–944.

Tsay, R.S. (2002). *Analysis of Financial Time Series. Financial Econometrics.* John Wiley & Sons. USA.

Tweedie, M.C.K. (1957). Statistical properties of inverse Gaussian distributions I. *Annals of Mathematical Statistics*, 28, 2, 362–377.

Whitmore, G.A. (1975). The inverse Gaussian distribution as a model of hospital stay. *Health Services Research*, 10(3), 297–302.

Whitmore, G.A. (1979). An inverse Gaussian model for labour turnover. *Journal of Royal Statistical Society A*, 142(4), 468–478.

Willmot, G.E. (1987). The Poisson–inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 113–127.

Zhang, M.Y., Russell, J.R. and Tsay, R.S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, 104, 179–207.