Particle filter estimation of duration-type models

Miguel A. G. Belmonte

April 25, 2012

Abstract

In this paper we model financial durations by discrete and continuous-time point processes in state-space form (SSF). We illustrate our analysis on a duration dataset analysed in Engle (2000). For estimation of intensity and static parameters, we resort to particle filters. We compare estimates delivered by simulation-based filters, with methodology suggested in Engle (2000) and Bauwens and Veredas (2004).

We conclude that the smooth particle filter (SPF) of Pitt (2002) is an efficient method for off-line parameter estimation and on-line filtering of univariate SSF duration models. We have applied the particle MCMC of Andrieu *et al.* (2008) to univariate models and offer a comparison with estimates from the SPF.

Keywords: Durations, particle filters, state space, smooth log-likelihood

1 Introduction

In this paper we model duration data, defined as times between arrivals of events. In particular, we focus on financial durations and fit the dataset analysed in Engle (2000). Financial durations comprise thousands of observations arising in quick succession. Markov models formulated in state space form (SSF) allow on-line efficient computational methods to infer about latent intensity $\{\psi_t\}_{t=1}^T$. From a modelling perspective, a Markovian formulation of latent intensity captures the dependence among durations. Also, discrete and continuous-time formulations of intensity dynamics are plausible.

To avoid biased inference introduced by local linearisation of nonlinear and non-Gaussian SSF models, we resort to numerical simulation. Sequential sampling of the filtering density $\{p(\psi_{0:t} \mid \mathcal{F}_t)\}_{t=1}^T$, by particle filters (PFs) allows learning about the state by using all the information currently available, $\mathcal{F}_t \equiv \{y_t\}_{t=1}^T$.

On-line parameter estimation of unknown parameters θ, is more challenging than filtering because PFs cannot sample from the joint $p(\psi_{0:t}, \theta \mid \mathcal{F}_t)$, accurately. For parameter estimation we consider two off-line approaches which make use of the PF output. Firstly, the smooth particle filter (SPF) of Pitt (2002), delivers at cost $\mathcal{O}(N)$ smooth log-likelihood estimates that can be maximised. When the latent state is not univariate, the alternative is to adopt the recently suggested particle Markov Chain Monte Carlo (PMCMC) of Andrieu *et al.* (2008). The PMCMC uses unbiased particle likelihood estimates to build a Markov chain with stationary distribution $p(\theta \mid \mathcal{F}_T)$.

Current methodology for inference of duration models is based on Shephard and Pitt (1997), where draws from nonlinear log-likelihood are obtained by using a combination of importance sampling and the Kalman filter and smoother. Koopman and Lucas (2008) applies the method of Shephard and Pitt (1997) to model default risk. Koopman *et al.* (2008) introduces a new model for credit rating transitions and inference is based on Monte Carlo maximum likelihood by adopting Shephard and Pitt (1997). Instead McNeil and Wendin (2007) apply the Gibbs sampling to fit a generalised linear mixed model of credit default risk (another class of data modelled by duration models). Inference of the stochastic conditional duration (SCD) model of Bauwens and Veredas (2004) is implemented in Strickland *et al.* (2006) by integrating latent variables out of the posterior by a hybrid Gibbs and Metropolis-Hastings (MH) algorithm.

In this paper we concentrate on methods which directly use the PF output, exploiting its strength in estimating the unobserved intensity without the need for careful choice of proposal densities or data-specific tuning. In cases of univariate latent models for durations, PFs do not rely on linear approximations and represent a valid alternative to inference by Markov Chain Monte Carlo methods (MCMC).

The structure of the paper is as follows. In section 1.1 we introduce duration data that we use throughout the article and review univariate models of duration. We present results of existing methodology for comparison with our approach to modelling. Section 2 highlights the filtering recursions that form the basis of simulation-based filters. We also discuss two methods for off-line parameter estimation. Section 3 estimates univariate models with the SPF and develops a methodology for inference of Cox processes. Section 4 presents extensions of univariate models to the bi-variate case and summarises with the main conclusions.

1.1 Duration dataset

Throughout this article we use a duration dataset that consists of three months of IBM trade durations, ranging from 01/11/90 to 31/01/91. Durations are measured in seconds. We have followed the coded routines of Engle (2000) to filter seasonality by piecewise linear splines with knots set every half an hour, starting at 10.00 am. The total number of seasonally adjusted durations is $T = 52144$. Summary statistics of the dataset are $Max = 23.87$, $Min = 0.027$, $\bar{Y} = 0.999$, $\sigma_Y^2 = 1.754$ and $Q_{LB}(15) = 8040$. We denote the Ljung-Box statistic based on 15 lags by $Q_{LB}(15)$.

1.2 ACD model

The autoregressive conditional duration (ACD) introduced in Engle and Russell (1998), belongs to the family of general autoregressive conditional heteroskedastic (GARCH) models. The ACD allows dependence by setting a deterministic linear difference equation for the intensity ψ_t. The ACD is an observation-driven model rather than parameter-driven and so the likelihood is directly evaluated. The conditional distribution of durations allows ML estimation through the prediction decomposition. The Weibull ACD(2,2) (WACD(2,2)) that we fit is of the form,

$$\begin{aligned}
\psi_t &= \omega + \textstyle\sum_{\ell=1}^{2} \alpha_\ell y_{t-\ell} + \sum_{\ell=1}^{2} \phi_\ell \psi_{t-\ell} \\
y_t &= \psi_t \xi_t, \;\; \xi_t \sim \mathcal{W}e(\kappa, \Gamma(1 + \tfrac{1}{\kappa}))
\end{aligned}$$

where y_t represents the time in seconds between successive trades. The parametrisation of ξ_t ensures that $\mathbb{E}[\xi_t] = 1$, so κ is identifiable. The WACD(2,2) is parametrised with restrictions in their parameter values ($\alpha_\ell > 0$ and $\phi_\ell > 0, \ell = 1, 2$), so that intensity is constraint to be positive. The WACD(2,2) estimates low values of ψ_t in periods of short duration or high frequency of arrivals.

Maximum likelihood estimates (MLEs) are obtained by maximising the log-likelihood function

$$\ell(\theta) = \sum_{t=1}^{T} \left\{ \log\left(\frac{\kappa}{y_t}\right) + \kappa \log\left(\frac{\Gamma(1+\frac{1}{\kappa})y_t}{\psi_t}\right) - \left(\frac{\Gamma(1+\frac{1}{\kappa})y_t}{\psi_t}\right)^{\kappa} \right\}$$

by the method of Broyden, Fletcher, Goldfarb and Shanno (BFGS). The results are displayed in Table 1. Standard errors are in brackets. MLEs show that persistence of data is very high.

$\widehat{\omega}$	$\widehat{\alpha_1}$	$\widehat{\alpha_2}$	$\widehat{\psi_1}$	$\widehat{\psi_2}$	$\widehat{\kappa}$
0.014108 (0.00141)	0.063628 (0.00246)	0.062170 (0.00283)	−0.058487 (0.01774)	0.92042 (0.0165)	0.91332 (0.00303)
$\ell(\widehat{\theta}) = -47741.9173$ Persistence = 0.9877			$Q_{LB}(15) = 49.0696$ $\chi^2_{0.95(15)} = 25$		

Table 1: *MLEs for WACD(2,2)*

Standardised residuals $\widehat{\xi_t} := \frac{y_t}{\widehat{\psi_t}}$, are correlated as evidenced by $Q_{LB}(15) > \chi^2_{0.95(15)}$.

Figure 1.2 shows off-line estimates of intensity. We observe that the WACD(2,2) tracks high intensity by decreasing $\widehat{\psi_t}$. The correlogram shows significant spikes that evidence that some dependence remains in the residuals. This is confirmed by the histogram of $\widehat{\xi_t}$ passed through their cumulative distribution function (cdf).

$$\widehat{e_t} = 1 - \exp\left\{-\Gamma\left(1+\frac{1}{\widehat{\kappa}}\right)\widehat{\xi_t}\right\}^{\widehat{\kappa}}.$$

1.3 SCD model

The stochastic conditional duration (SCD) proposed in Bauwens and Veredas (2004) is a discrete time formulation of a univariate latent variable process

$$\begin{array}{rcl} \psi_t & = & \omega(1-\phi) + \phi\psi_{t-1} + \varepsilon_t, \ \ \varepsilon_t \sim \mathcal{N}(0, \sigma^2_\varepsilon) \\ y_t & = & \exp\{\psi_t\}\xi_t, \ \ \xi_t \sim \mathcal{W}e(\kappa, 1). \end{array} \tag{1}$$

The SCD models dependence by an stationary autoregressive of first order (AR(1)). Unlike the ACD model, the SCD introduces an error term, ε_t in the equation for intensity which is independent of ξ_t. Observed durations y_t are marginally distributed according to a mixture of log-normal and Weibull densities. In Bauwens and Veredas (2004), inference about ψ_t and $\theta = (\omega, \phi, \sigma^2_\varepsilon, \kappa)$ is implemented by quasi-maximum likelihood (QML). In section 3, we will consider more efficient methods for inference. Lack of accuracy in the approximation of a log-Weibull by a Gaussian density will yield biased state and parameter estimation. The log-linearised version of (1) results in an observation equation

$$\log\{y_t\} = -\frac{0.57722}{\kappa} + \psi_t + \zeta_t, \ \ \zeta_t \sim \mathcal{N}(0, \frac{\pi^2}{6\kappa^2}). \tag{2}$$

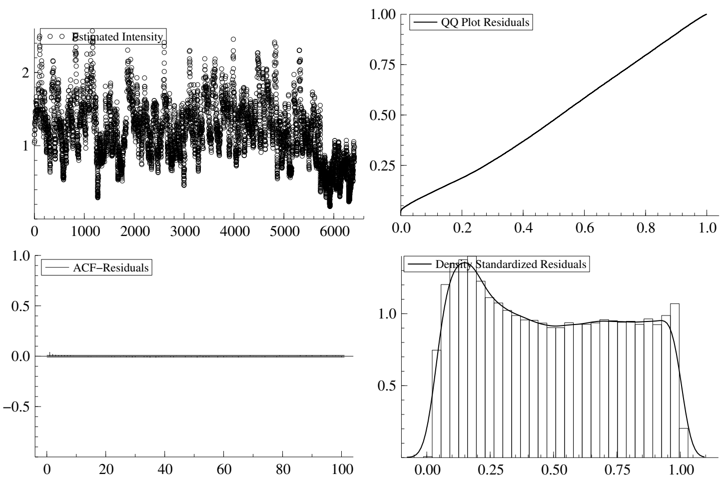

Figure 1: *Intensity and residuals of the WACD(2,2) model*. Top-left: Off-line estimates of intensity for first 6000 durations against arrival times. Top-right: QQ plot of standardised residuals against the $U(0,1)$ distribution. Bottom-left: Correlogram of standardised residuals. Bottom-right: Histogram of standardised residual passed through the cdf for the whole dataset.

To assess robustness of the log-linearised model, Bauwens and Veredas (2004) compute standardised residuals according to the original multiplicative formulation $\widehat{\xi}_t$ in (1) and to the log-linearised version $\widehat{\zeta}_t$ in (2). Results are displayed in Table 2. Clearly, both $Q^{15}_{LB(\widehat{\xi}_t)}$ and $Q^{15}_{LB(\widehat{\zeta}_t)}$ exceed the critical value $\chi^2_{0.95(15)} = 25$, so the hypothesis that the log-linearised model delivers random residuals is rejected.

$\widehat{\omega}$	$\widehat{\phi}$	$\widehat{\sigma^2_\varepsilon}$	$\widehat{\kappa}$
-0.1846	0.9834	0.007880	1.07928
(0.02403)	(0.086562)	(0.08344)	(0.003319)
$\ell(\widehat{\theta}) = -84562.036$		$Q^{15}_{LB(\widehat{\xi}_t)} = 140.35$	$Q^{15}_{LB(\widehat{\zeta}_t)} = 61.09$

Table 2: *MLEs by QML for log-linearised SCD*

2 Filtering and parameter estimation

The following section describes the recursions that lead to state and off-line parameter estimation by PFs. Off-line parameter estimation uses particle log-likelihood estimates delivered at the end of the sample T.

2.1 Filtering recursions

To learn about $\{\psi_t\}_{t=1}^T$, the solution is to sample sequentially from the joint filtering density $\{p(\psi_{0:t} \mid \mathcal{F}_t, \theta)\}_{t=1}^T$. See Gordon *et al.* (1993), Pitt and Shephard (1999) and Doucet and

Johansen (2008) for further details. In this section, the vector of observations θ, is regarded as known.

The particle approximation to the filtering density is based on

$$p(\psi_{0:t} \mid \mathcal{F}_t, \theta) \propto f(\psi_t \mid \psi_{t-1}, \theta) g(y_t \mid \psi_t, \theta) p(\psi_{0:t-1} \mid \mathcal{F}_{t-1}, \theta),$$

where $p(. \mid .)$ represents the filtering density at $t-1$, $f(. \mid .)$ the transition density of the state equation and $g(. \mid .)$ the conditional likelihood.

A filter recursion updates N random samples from the filtering density at $t-1$

$$\left\{\psi_{t-1}^{(j)}, \frac{1}{N}\right\}_{j=1}^{N} \sim p(\psi_{0:t-1} \mid \mathcal{F}_{t-1}, \theta),$$

first by forecasting the state

$$\psi_t^{*(j)} \sim f(\psi_t \mid \psi_{t-1}^{(j)}, \theta)$$

and secondly by weighting according to the conditional likelihood once y_t is available

$$\omega_t^{(j)} \propto g(y_t \mid \psi_t^{*(j)}, \theta).$$

At t, the updated particle approximation to the filtering density

$$\left\{\psi_t^{*(j)}, \omega_t^{(j)}\right\}_{j=1}^{N} \sim p(\psi_{0:t} \mid \mathcal{F}_t, \theta)$$

is asymptotically consistent.

To avoid exponential grow of the variance of the weights, we occasionally re-sample existing particles according to their weights. Then, under mixing conditions on the latent process, the increase of the error over time is only linear.

2.2 Off-line parameter estimation

The constant of proportionality of the joint filtering distribution represents, by prediction decomposition, the likelihood function $L(\theta) := p(\mathcal{F}_T \mid \theta)$,

$$p(\mathcal{F}_T \mid \theta) = \prod_{t=1}^{T} p(y_t \mid \mathcal{F}_{t-1}, \theta). \tag{3}$$

Each independent term in the likelihood can be estimated because the densities involved can be evaluated

$$\begin{aligned} p(y_t \mid \mathcal{F}_{t-1}, \theta) &= \int g(y_t \mid \psi_t, \theta) f(\psi_t \mid \mathcal{F}_{t-1}, \theta) \; \mathrm{d}\psi_t \\ &= \mathbb{E}_{\psi_t \mid \mathcal{F}_{t-1}} \left[g(y_t \mid \psi_t, \theta) \right] \end{aligned}$$

Thus, a particle approximation to the log-likelihood, at the end of the sample is available

$$\widehat{\ell}(\theta) \approx \sum_{t=1}^{T} \log \left(\frac{1}{N} \sum_{j=1}^{N} \omega_t^{(j)} \right).$$

2.2.1 Smooth particle filter

If the mapping $\theta \mapsto \widehat{\ell}(\theta)$ were smooth, we could find the MLEs by maximising $\widehat{\ell}(\theta)$ with a BFGS. However, re-sampling introduces discontinuities in the approximation. A marginal change in θ, will dramatically affect the vector of re-sampled particles due to discreteness of the empirical distribution function $F_\theta(\psi_t)$.

For univariate states, the smooth particle filter (SPF) of Pitt (2002) bypasses discontinuities in $\theta \mapsto \widehat{\ell}(\theta)$ by approximating the discrete $F_\theta(\psi_t)$ with a continuous one, denoted by $\widehat{F}_\theta(\psi_t)$. A piecewise linear approximation ensures that $\widehat{F}_\theta(\psi_t) \to F_\theta(\psi_t)$ as $N \to \infty$. After smooth re-sampling $\widehat{F}_\theta(\psi_t)$ changes smoothly in θ while the computational cost is maintained at $O(N)$

2.2.2 Particle MCMC

In Bayesian inference, we are concerned with sampling from $p(\theta \mid \mathcal{F}_T) \propto L(\theta)p(\theta)$ to make inference about θ. The problem is that we do not directly know $L(\theta)$ but by (3) we generate an unbiased estimate $\widehat{L}(\theta)$.

The main idea of the PMCMC of Andrieu *et al.* (2008) is to use $\widehat{L}(\theta)$ to build a Markov chain with stationary distribution $p(\theta \mid \mathcal{F}_T)$. Thus, $\widehat{L}(\theta)$ allows sampling from $p(\theta \mid \mathcal{F}_T)$ up to proportionality. This idea is exploited in Andrieu *et al.* (2008) to propose a marginal MH algorithm with acceptance probability

$$\alpha(\theta_{\text{new}} \mid \theta_{\text{old}}) = \min\left\{\frac{\widehat{L}(\theta_{\text{new}})p(\theta_{\text{new}})q(\theta_{\text{old}} \mid \theta_{\text{new}})}{\widehat{L}(\theta_{\text{old}})p(\theta_{\text{old}})q(\theta_{\text{new}} \mid \theta_{\text{old}})}, 1\right\}. \tag{4}$$

This sampler requires independent evaluations of $\widehat{L}(\theta_{\text{new}})$ per iteration, which is computationally expensive. The PMCMC accepts moves to areas of high likelihood $\widehat{L}(\theta_{\text{new}}) > \widehat{L}(\theta_{\text{old}})$ and seems and appealing alternative to models with multivariate states. In the latest case, the SPF cannot be applied.

However, $\alpha(\theta_{\text{new}} \mid \theta_{\text{old}})$ is not a continuous function of $\widehat{L}(\theta)$. Lack of smoothness of $\widehat{L}(\theta)$ implies that, contrary to Roberts and Rosenthal (2001), the acceptance rates of the MH will be low, despite decreasing the scaling of the proposals. A simple Monte Carlo experiment carried with simulated data from a noisy AR(1), illustrates this argument. Random walk

Observations	Scaling	$L(\theta)_{KF}$	$\widehat{L}(\theta)_{GSS}$	$\widehat{L}(\theta)_{SPF}$
$T = 200$	$\sigma = 0.8$	40.8	34.6	37.3
$T = 1000$	$\sigma = 0.43$	40.3	27.5	31.3

Table 3: *PMCMC acceptance ratios for exact and estimated log-likelihoods*

proposals for the dependence parameter ϕ of the AR(1) are scaled according to σ. Table 3 shows that acceptance rates are maintained only for the case of the exact likehood $L(\theta)_{KF}$, delivered by the Kalman filter. For the discontinuous $\widehat{L}(\theta)_{GSS}$ delivered by the Gordon *et al.* (1993) filter, acceptance rates are lower than those corresponding to estimates of $\widehat{L}(\theta)_{SPF}$ by the SPF.

Chopin (2004) proves that the variance of $\widehat{L}(\theta)$, increases linearly with the number of observations

$$\frac{\mathbb{V}\left[\widehat{L}(\theta)\right]}{L^2(\theta)} \leq D\frac{T}{N}. \tag{5}$$

Therefore, an increase of the number of particles N, proportional to T is necessary to control the variance of $\widehat{L}(\theta)$. In the following experiment implemented with simulated data, again generated from a noisy AR(1), we change the number of particles N and the number of observations. Table 4 shows that for fixed θ, we need to increase N proportionally to T in

N	$T = 1000$	$T = 5000$	$T = 10000$
200	5.645	9.487	13.043
1000	1.796	3.958	4.557
5000	0.975	1.576	1.977
10000	0.602	1.281	1.973

Table 4: *Standard deviations of particle log-likelihood estimates for* $\widehat{L}(\theta)_{GSS}$.

order to achieve moderate standard deviations for $\widehat{L}(\theta)_{GSS}$.

These results show that in the case of long time series, the PMCMC may be computationally very expensive.

3 Estimation of univariate duration models

3.1 Inference of the SCD by SPF and PMCMC

In this section we implement parameter estimation and filtering for the whole dataset $T = 52144$. The number of particles is set to $N = 2000$. MLEs, standard errors, value of $\ell(\widehat{\theta}_{MLE})$ and $Q^{15}_{LB(\widehat{\xi}_t)}$ are shown on Table 5. Computing time of MLEs barely took 1 hour on a desktop

$\widehat{\omega}$	$\widehat{\phi}$	$\widehat{\sigma^2_\varepsilon}$	$\widehat{\kappa}$
−0.154	0.979	0.013	0.970
(0.0233)	(0.0500)	(0.0317)	(0.0035)
$\ell(\widehat{\theta}) = -47377.00$		$Q^{15}_{LB(\widehat{\xi}_t)} = 32.57$	

Table 5: *MLEs for SCD by PFs*

PC. Significant differences with estimates for the WACD(2,2) of Table 1 are that the mean ω is negative. That is due to the different parametrisation of the WACD(2,2) that forces $\widehat{\psi}_t$ to be positive. The value of $\ell(\widehat{\theta})$ for the SCD is 364 log-likelihood units larger than for the WACD(2,2). Also residuals for the SCD have a lower $Q^{15}_{LB(\widehat{\xi}_t)}$ statistic than for the WACD(2,2). Regarding similarities, the persistence for the SCD also shows a high value.

Figure 3.1 shows on-line estimates of ψ_t, computed once the static parameters are estimated. Intensity is estimated on-line by the weighted mean

$$\widehat{\psi_t} = \frac{\sum_{j=1}^{N} \psi_t^{*(j)} \omega_t^{(j)}}{\sum_{j=1}^{N} \omega_t^{(j)}}.$$

In contrast with the WACD(2,2), the latent factor ψ_t in the SCD tracks periods of high intensity with more accuracy than $\widehat{\psi_t}$ for the WACD(2,2). Also estimates of ψ_t for the SCD decrease their magnitude to track periods of short durations.

The effective sample size (ESS) of Liu and Chen (1998) computed at arrival times, measures the performance of the filter, by giving an estimate of the variance of the weights. For filtering, we implement stratified re-sampling for each observed duration, so that the variance of the weights is zeros and the ESS becomes N. The histogram of residuals, for the whole dataset is obtained by

$$\begin{aligned} F(y_t \mid \mathcal{F}_{t-1}, \theta) &= \int F(y_t \mid \psi_t, \theta) p(\psi_t \mid \mathcal{F}_{t-1}, \theta) \; \mathrm{d}\psi_t \\ &\approx \frac{1}{N} \sum_{j=1}^{N} F(y_t \mid \psi_t^{(j)}) \end{aligned} \qquad (6)$$

introduced in Pitt and Shephard (1999) as a model diagnostic. The cumulative distribution $F(y_t \mid \psi_t)$ is the cdf of observations.

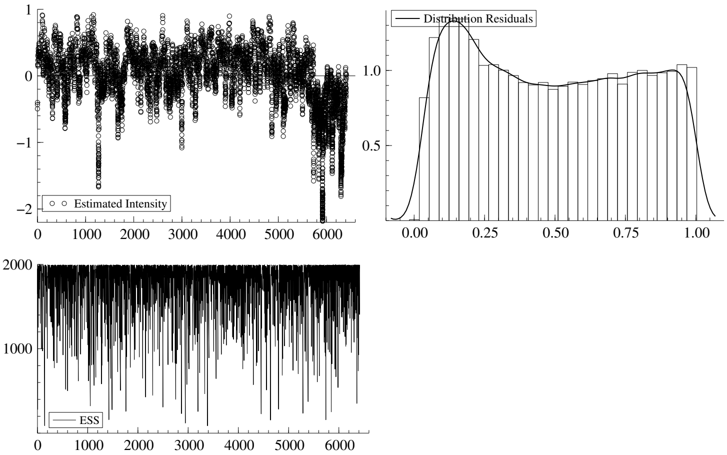

Figure 2: *Particle filtering applied to SCD model*. Top-left: Filtering means for first 6000 durations against arrival times, $N = 2000$. Top-right: Histogram of residuals for the whole dataset computed according to (6). Bottom: ESS for first 6000 durations approximated by $(\sum_{j=1}^{N} \omega_t^{(j)})^2 / \sum_{j=1}^{N} (\omega_t^{(j)})^2$.

Now, we show results of the PMCMC applied only to a subset of the first $T = 4000$ durations, for computational reasons outlined in section 2.2.2. Those durations have approximately accrued during two days of financial trading. We have set the initial values to the MLEs that we have found with the SPF (Table 7) and scale each proposal with corresponding standard deviations for the parameters. The number of iterations is Iter = 20000 and the number of particles is set to $N = 1000$. Acceptance rate is 0.362 and computing time was 10 hours. To dominate the tails of the posterior $p(\theta \mid \mathcal{F}_T)$ we choose as random term for the proposals a t distribution with 3 degrees of freedom. To avoid correlation in the chains for ϕ and σ_ε^2, we have parametrised the unknown variance as $\gamma := \frac{\sigma_\varepsilon^2}{1-\phi^2}$. Positive proposals for κ and σ_ε^2 are generated from a logarithmic random walk. Parameters ω and ψ are proposed from symmetric random walk and ψ is transformed to $(-1, 1)$ domain with a logit transformation. The prior that we have chosen for the parameters is $p(\theta) = \gamma^{-1/2} \exp^{-\sigma_\varepsilon^2}$.

	Quantiles		
Parameters	2.5%	50.0%	97.5%
$\widehat{\omega}$	0.0333	0.1069	0.1962
$\widehat{\phi}$	0.8819	0.9260	0.9537
$\widehat{\sigma_\varepsilon^2}$	0.0194	0.0270	0.03627
$\widehat{\kappa}$	0.9056	0.9321	0.9644
$\widehat{\ell(\theta)}$	−4746.84	−4742.83	−4740.43

Table 6: *Posterior summaries for SCD by PMCMC, $T = 4000$, $N = 1000$*

Posterior summaries of Table 6 are very close to the MLEs of Table 7 computed by PFs.

$\widehat{\omega}$	$\widehat{\phi}$	$\widehat{\sigma_\varepsilon^2}$	$\widehat{\kappa}$
0.1085	0.9224	0.02690	0.9311
(0.0342)	(0.0883)	(0.0796)	(0.0125)
$\ell(\widehat{\theta}) = -4742.1191$			

Table 7: *MLEs for SCD by PFs, $T = 4000$, $N = 1000$*

3.2 Modelling and inference of the Cox process

The third univariate model that we consider is a Cox process, quite popular for modelling clustered point patterns (e.g. Waagepetersen (2004)). The Cox process is a SSF model that formulates univariate intensity as an Ornstein-Uhlenbeck process that generates observed arrivals at discrete times. Continuous-time formulation of intensity is closer to the nature of this estimation problem and also allows appropriate scaling of the model according to inter-arrival times.

In this section, we change notation and denote arrival times by y_t, where y_t is the actual t^{th} arrival time. Durations are now defined by $y_t - y_{t-1}$, as time between two arrivals. We follow the convention of denoting random variables by upper case letters and their realisations by lower case. Assuming a Cox process,

$$\begin{aligned} d\Psi_s &= -\phi\,(\Psi_s - \omega)\,ds + \sigma_\varepsilon dB_s, \;\; s \geqslant 0 \\ p(y_t \mid \{\Psi_s; y_{t-1} \leq s \leq y_t\}) &= \nu(\psi_{y_t}) \exp\left\{ -\int_{y_{t-1}}^{y_t} \nu(\Psi_s)\,ds \right\} \end{aligned} \tag{7}$$

durations are generated by an inhomogeneous Poisson process with intensity function $\nu : \Psi_s \to \mathbb{R}_+$, defined as $\nu(.) \equiv \exp(\Psi_s)$. The Markov assumption on $\{\Psi_s; s \geqslant 0\}$ enables application of particle filters for on-line estimation of intensity $\nu(\psi_{y_t})$ at arrival times. Previous work in Fearnhead *et al.* (2008) implemented filtering for a Cox process by using Exact simulation of diffusions (as in Beskos *et al.* (2006)) and the auxiliary particle filter of Pitt and Shephard (1999). In this article, we follow a discretisation scheme to use the outcome of a particle filter algorithm for off-line estimation of the parameters.

To apply PFs to the Cox process, we need the transition density and the conditional likelihood in closed form, as was shown in section 2.1. The transition density of an Ornstein-Uhlenbeck process is known and easy to sample from. On the other hand, the conditional likelihood

$$p(y_t \mid y_{t-1}, \psi_{y_{t-1}}, \psi_{y_t}) := \nu(\psi_{y_t})\mathbb{E}\left[\exp\left\{-\int_{y_{t-1}}^{y_t} \nu(\Psi_s)\ \mathrm{d}s\right\}\right] \tag{8}$$

involved in the weights is intractable because it depends upon the entire path of Ψ_s in the interval $[y_{t-1}, y_t)$.

A basic solution to approximate (8) is to discretise the interval $[y_{t-1}, y_t)$. This approach is equivalent to augment observed arrivals with additional observations in-between arrival times. We introduce a parameter $\Delta > 0$ which represents the maximum duration between two arrivals. If the duration is smaller than Δ we do not discretise. On the contrary, if $y_t - y_{t-1} > \Delta$, then the ratio $\frac{(y_t - y_{t-1})}{\Delta}$ defines the M number of discretised points and $\delta = \frac{y_t - y_{t-1}}{M+1}$ defines the width of each sub-interval. Since times between arrivals are informative of no arrival, we define data at times

$$\left\{y_{t-1} = \tau_0^{(t)} < \tau_1^{(t)} < \cdots < \tau_M^{(t)} = y_t\right\}_{t=1}^{T}$$

which is a superset of the observed arrivals. Thus we use the auxiliary variables $\tau_M^{(t)}$ to construct a data set $\{z_m\}_{m=1}^{M}$, where $z_m = 1$ if $\tau_m = y_t$ for some t, and 0 otherwise. In this setting, the conditional likelihood

$$P\left[Z_m = 0 \mid z_{m-1}, \psi_{\tau_{m-1}}, \psi_{\tau_m}\right] = \exp\left\{-\nu(\psi_{\tau_{m-1}})\delta\right\}.$$

depends only on state $\psi_{\tau_{m-1}}$ so that the SPF can be applied for parameter estimation.

Notice that Δ determines the accuracy of the approximation to (8) and crucially affects the particle approximation to the log-likelihood function. We have carried Monte Carlo experiments to understand the effect that discretisation Δ, has on parameter estimation. For 50 duration datasets of $T = 2000$, generated with fixed parameter values at $\omega = -1$, $\phi = 0.3$ and $\sigma_\varepsilon^2 = 0.15$ we have estimated MLEs, setting $N = 1000$. Table 8 illustrates bias, mean squared error (MSE) and variance of the MLEs for two levels of discretisation. Table 8 shows

Discretisation	$\Delta = 20$			$\Delta = 0.5$		
Parameters	$\widehat{\omega}$	$\widehat{\phi}$	$\widehat{\sigma_\varepsilon^2}$	$\widehat{\omega}$	$\widehat{\phi}$	$\widehat{\sigma_\varepsilon^2}$
Bias	0.113	0.026	−0.037	0.0126	−0.0107	−0.0096
MSE	0.0138	0.0208	0.0048	0.0015	0.0051	0.00217
Variance	0.0010	0.0201	0.0035	0.0014	0.0051	0.00208

Table 8: *Bias, MSE, and variance of parameter estimates for different* Δ

that for crude distretisation, represented by $\Delta = 20$, bias and MSE of MLEs is much larger than for finer discretisation, $\Delta = 0.5$. Therefore, it is recommended to work with integrated quantities to avoid biased parameter estimates.

After understanding discretisation bias, we estimate the parameters of a Cox process, setting $N = 2000$ and $\Delta = 0.5$ Due to identifibility problems we define $\gamma := \frac{\sigma_\varepsilon^2}{2\phi}$ to estimate ϕ,

$\widehat{\omega}$	$\widehat{\gamma} := \frac{\sigma_\varepsilon^2}{2\phi}$	$\widehat{\sigma_\varepsilon^2}$	$\widehat{\phi}$
−0.1424	0.2934	0.0316	0.0539
(0.0154)	(0.0295)	(0.0334)	N.A
$\ell(\widehat{\theta}) = -47437$		$e^{-\widehat{\phi}} = 0.9476$	
$Q^{15}_{LB(\widehat{\varepsilon_t})} = 37.62$			

Table 9: *MLEs for Cox process by SPF,* $N = 1000$

since there is more information for σ_ε^2. The log-likelihood $\ell(\widehat{\theta})$ is 60 log-likelihood units lower than that for the SCD but 304 larger than that for the WACD(2,2). Persistence is calculated by $e^{-\widehat{\phi}} = 0.9476$, which is moderately high. The Ljung-Box statistic is slightly worse than for the SCD model. These results, justify introducing a Cox process with a Negative-Binomial which models conditional over-dispersion better than a Poisson process.

With the MLEs from Table 9 we implement filtering. Residuals for the Cox process are

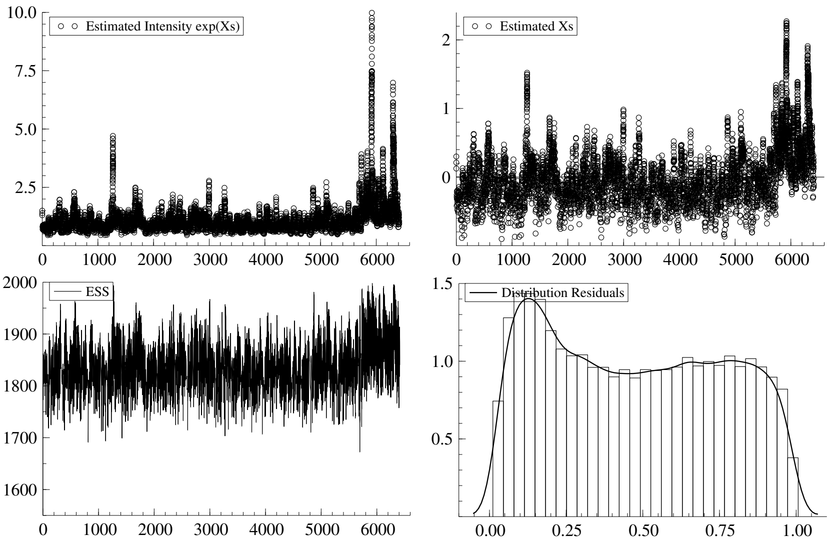

Figure 3: *Particle filtering applied to Cox process*
Top-left: Estimates of intensity $\nu(\psi_{y_t}) = \exp(\psi_{y_t})$ for first 6000 durations against arrival times, $N = 1000$. Top-right: Estimates of $\{\Psi_{y_t}\}$ against arrival times. Bottom-left: ESS for first 6000 durations. Bottom-right: Histogram of residuals for the whole dataset computed according to (9)

computed according to

$$F(y_t \mid \mathcal{F}_{t-1}, \theta) \approx \frac{1}{N}\sum_{j=1}^{N} 1 - \exp\left\{-\sum_{d=1}^{M} \nu(\psi_{\tau_d}^{(j)})\right\} \tag{9}$$

Figure 3.2, shows the results, where intensity and latent value of the Ornstein-Uhlenbeck equation are shown. For the Cox process, estimates of latent intensity and unobserved excursion of the Ornstein-Uhlenbeck increase. Periods of high latent intensity are clearly shown. Also, in periods of high intensity, more often re-sampling increases the ESS.

4 Conclusions

Motivated by the peak of the standardised residuals distributed around 0, we are interested in fitting multifactor models. All the models discussed in this paper allow superposition of intensities. The WACD(2,2) model fitted in section 1.2 can be superimposed, defining total intensity $\Upsilon_t := \psi_t + \lambda_t$, where ψ_t represents the short-term and λ_t the long-term intensity,

$$\begin{aligned}
\psi_t &= \lambda_t + \phi(\psi_{t-1} - \lambda_{t-1}) + \alpha(y_{t-1} - \lambda_{t-1}) \\
\lambda_t &= \omega(1-\rho) + \rho\lambda_{t-1} + \beta(y_{t-1} - \psi_{t-1}) \\
y_t &= \Upsilon_t \xi_t, \;\; \xi_t \sim \mathcal{W}e(\kappa, \Gamma(1 + \tfrac{1}{\kappa})).
\end{aligned}$$

The motivation of the two-component version of ACD models is discussed at length in Engle and Lee (1999). Inference on superimposed models is far from straightforward. In the equation for observed durations, there is information about the sum of intensities but not about each individual intensity. This makes estimation of parameters for each intensity difficult. For parameter estimation, we work instead with the reduced form of the model

$$\begin{aligned}
\psi_t &= \omega(1-\alpha-\phi)(1-\rho) + (\alpha+\beta)y_{t-1} + (-\alpha\beta - \beta\phi - \alpha\rho)y_{t-2} \\
&+ (\phi + \rho - \beta)\psi_{t-1} + (\alpha\beta + \beta\phi - \phi\rho)\psi_{t-2} \\
y_t &= \psi_t \xi_t, \;\; \xi_t \sim \mathcal{W}e(\kappa, \Gamma(1 + \frac{1}{\kappa})).
\end{aligned}$$

so that observed durations depend on one intensity. The dynamic equation for ψ_t is still a GARCH(2,2) but now parametrised in terms of its roots, so that these are constrained to be real. The MLEs are summarised on Table 10. The log-likelihood estimate $\ell(\widehat{\theta})$ from two-component WACD(2,2) is 80 log-likelihood units higher than the same value for WACD(2,2). The value of persistence given for the parameter with largest root $\widehat{\phi}$, is very close to persistence implied by the WACD(2,2). The value of $Q_{LB}(15)$ shows a dramatic improvement over equivalent statistic for the WACD(2,2). The two-component WACD(2,2) is clearly superior to the WACD(2,2).

$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\phi}$	$\widehat{\rho}$	$\widehat{\kappa}$
1.1395 (0.13265)	0.021102 (0.0025487)	0.049524 (0.0030485)	0.98712 (0.0026593)	0.95942 (0.0049055)	0.91536 (0.0030343)
$\ell(\widehat{\theta}) = -47661.0055956$			$Q_{LB}(15) = 29.4776$		
Persistence = 0.98712			$\chi^2_{15} = 25$		

Table 10: *MLEs for two-component WACD(2,2)*

The next model that allows superposition of intensities is the SCD model. Some preliminary results using QML are presented here. However, from simulations we have seen that the QML gives unreliable estimates.

The two-component SCD models intensity as the sum of two independent AR(1). Intensity is again, the sum of short-term ψ_t, and λ_t long-term intensity $\Upsilon_t := \psi_t + \lambda_t$. First, we apply the sub-optimal QML to the two-component SCD model, for the whole dataset $T = 52144$. With respect to the SCD, the two-component SCD shows an improvement of 90

$\widehat{\omega}$	$\widehat{\phi}$	$\widehat{\sigma_\varepsilon^2}$	$\widehat{\rho}$	$\widehat{\sigma_\eta^2}$	$\widehat{\kappa}$
−0.1818	0.9509	0.01425	0.9991	0.0002	1.0861
(0.06839)	(0.1121)	(0.1021)	(0.3545)	(0.3452)	(0.00345)
$\ell(\widehat{\theta}) = -84472.5085$		$Q^{15}_{LB(\widehat{\xi_t})} = 53.82$		$Q^{15}_{LB(\widehat{\zeta_t})} = 31.59$	

Table 11: *QML Two-component model SCD*

log-likelihood units. Also $Q^{15}_{LB(\widehat{\xi_t})}$ and $Q^{15}_{LB(\widehat{\zeta_t})}$ show less dependence. However, large standard errors evidence identifiability problems due to bi-modality of the linearised likelihood function.

Before applying the PMCMC to the two-component SCD, we fit the same model for $T = 4000$ observations and we find extremely bad estimates with large standard errors, which evidence that we cannot apply the PMCMC for identifiability issues. The same problem applies to the two-component Cox process.

We are currently experimenting with the PMCMC, however it is challenging to scale the method well for such long series.

We conclude summarising the findings of the applied methodology to durations and suggesting directions for future work.

This paper shows that latent variable models are preferable to observation driven models in scenarios of high frequency regime and long time series. Despite intractability of the likelihood function due to latent variables, particle filters offer a computationally convenient tool for on-line inference and parameter estimation of univariate duration models. Latent variables allow flexibility in the formulation of dynamics, whether in discrete or continuous-time, and lend themselves to formulate valid models to capture dependence.

The SCD model has proved superior to the WACD(2,2) in modelling the duration dataset of Engle (2000). By particle filters, the SCD is capable of tracking latent intensity, allows off-line parameter estimation and model diagnostics. Table 2 and Figure 3.1 evidence that the SCD is a valid model for dependence.

Attempts to model durations by a two-component WACD(2,2) and also by a two-component SCD by QML show that bi-variate models yield better residuals than univariate models. This is not new, and corroborates results of Roberts *et al.* (2004) and Griffin and Steel (2008).

Regarding the Cox process, we find desirable to model over-dispersion by a Negative-Binomial, rather than by a Poisson process. The smooth particle filter works well applied to a Cox process, which becomes another possible model for modelling duration-type data. The appealing feature of continuous-time formulation of intensity is that it allows appropriate scaling of the model according to inter-arrival times.

We are currently experimenting with the PMCMC, however it is challenging to tune the method so that it scales well for long series. Our aim is to extend the applicability of particle filters to bi-variate latent models by application of the PMCMC. At the moment, inference for those are attainable only by MCMC. Unfortunately, non-continuous likelihood estimates and evaluation of the likelihood at each iteration make the PMCMC algorithm very expensive computationally. Also, there are identifiability problems in the bivariate models that we are using, which we are trying to overcome.

References

Andrieu, C., Doucet, A., and Holenstein, R. (2008). Particle Markov chain Monte Carlo. Submitted.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent variable model for the analysis of financial durations. Journal of Econometrics, vol. 119(2), 381-412.

Beskos, A., Papaspiliopoulos, O., Roberts, G.O and Fearnhead, P. (2006) Exact and computationally efficient likelihood-based inference for discretely observed diffusions. Journal of the Royal Statistical Society, series B, vol. 68, 1-25.

Chopin, N. (2004). Central Limit Theorem for sequential Monte Carlo methods and its application to Bayesian inference. Annals of Statistics, vol. 32, 6, 2385-2411.

Doucet, A. and Johansen A.M. (2008). A Tutorial on Particle Filtering and Smoothing: Fifteen years Later. Handbook of Nonlinear Filtering (eds. D. Crisan and B. Rozovsky), Oxford University Press, to appear.

Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2008). Particle filters for partially observed diffusions. Journal of the Royal Statistical Society, series B, vol. 70, 755-777(23).

Engle, R. F. and Russell, J. R. (1998). Forecasting transaction rates: the autoregressive conditional duration model. Econometrica, vol. 66, 1127-1162.

Engle, R. F. (2000). The econometrics of ultra-high frequency data. Econometrica, vol. 68, 1-22.

Engle, R. and Lee, G. (1999). A long-run and short-run component model of stock return volatility. In: R. Engle and H. White, Editors, Cointegration, Causality and Forecasting, Oxford University Press.

Gordon, N. J., Salmond D. J. and Smith A. F. M., (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. IEE-Proceedings F, vol. 140, 107-113.

Griffin, J. and Steel, M. F. J. (2008). Bayesian inference with stochastic volatility models using continuous superpositions of non-Gaussian Ornstein-Uhlenbeck processes. CRISM Paper No. 08-23. www.warwick.ac.uk/go/crism

Koopman, S. J. and Lucas, A. (2008). A Non-Gaussian Panel Time Series Model for Estimating and Decomposing Default Risk. Journal of Business & Economic Statistics, vol. 26, 4. 510-525(16).

Koopman, S. J., Lucas, A. and Monteiro, A. (2008). The multi-state latent factor intensity model for credit rating transitions. Journal of Econometrics, vol. 142, 1, 399-424.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. Journal of the American Statistical Association, vol. 93, 1032-1044.

Mcneil, A. J. and Wendin, J. P., (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. Journal of Empirical Finance, vol. 14, 131-149.

Pitt, M.K. (2002). Smooth Particle Filters For Likelihood Evaluation And Maximisation. Working paper, available in `http://www2.warwick.ac.uk/fac/soc/economics/staff/faculty/pitt/`

Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association, vol. 99, 590-599.

Roberts, G. O., Papaspiliopoulos, O. and Dellaportas, P., (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. Journal of the Royal Statistical Society, series B, vol. 66, 369-393.

Roberts, G. O., Rosenthal, J. S. (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. Statistical Science, vol. 16, No. 4, 351-367.

Shephard, N. and Pitt, M. (1997). Likelihood Analysis of Non-Gaussian Measurement Time Series. Biometrika, 84, 653-667.

Strickland, C. M., Forbes, C. S. and Martin, G. M. (2006). Bayesian Analysis of the Stochastic Conditional Duration Model. Computational Statistics & Data Analysis, vol. 50, Issue 9, 2247-2267.

Møller, J. and Waagepetersen, R.P. (2004). Statistical inference and simulation for spatial point processes. Monographs on Statistics and Applied Probability, 100. Chapman & Hall/CRC, Boca Raton, FL.