A Flexible Stochastic Conditional Duration Model

Samuel Gingras*
Université de Montréal

William J. McCausland†
Université de Montréal

Contact Information:
Département de Sciences Économiques
Université de Montréal
Montréal, Québec, Canada

*samuel.gingras@umontreal.ca
† william.j.mccausland@umontreal.ca

Université de Montréal

Introduction

A popular model for the analysis of sequences of durations between financial transactions is the Stochastic Conditional Duration model (SCD) introduced by Bauwens and Veredas (2004); a parameter driven model where the scale of the conditional distribution of duration is a latent state variable. In this paper, we argue that commonly used conditional distributions in SCD models possess unrealistic properties and propose a new flexible distribution more appropriate for financial duration modelling. Our proposed density is flexible in the sense that it can approximate any continuous density on a compact subset of $[0, \infty)$.

Contributions

1. Propose a flexible distribution capable of matching a wide variety of distribution with moderately varying hazard functions.
2. Include diurnal patterns in SCD models and jointly estimate the deterministic and stochastic parts of trading intensity.
3. Develop efficient posterior simulation methods for Bayesian inference in SCD models featuring our flexible distribution.

Motivation

A reasonable model for arrival time between transactions when there is a large number of homogenous trader acting independently is the Poisson process. This simple model has the property that the durations between arrivals are exponentially distributed, giving a constant hazard function.

Previous papers concluded that a distribution with constant hazard is not flexible enough in the context of financial durations and prefer either the Weibull or the gamma distribution. As shown in Figure 4, those two distribution have the property that when the hazard is not constant, it is either zero at a duration of zero and strictly increasing; or infinite at duration of zero and strictly decreasing.

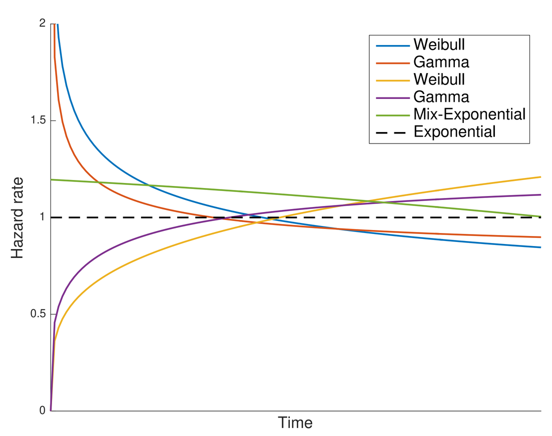

Figure 1: Hazard functions for popular conditional duration distributions.

Even though it is well known that trade intensity in financial markets varies over time, much of the variation is predictable; duration series are characterized by diurnal patterns and highly persistent stochastic components. This suggest that after conditioning on all relevant information, we would expect the conditional distribution to be not too far from an exponential distribution, and exhibit a slowly varying hazard.

Model

We observe all transaction times recorded to the nearest second in a market over a period of D consecutive trading days. The nominal opening and closing times are t_{open} and t_{close}, and any transaction times outside the interval $[t_{\text{open}}, t_{\text{close}}]$ are ignored.

For each day $d = 1, \ldots, D$, we construct durations from the ordered sequence of transaction times $\tilde{t}_{d0}, \tilde{t}_{d1}, \ldots, \tilde{t}_{dn_d}$ as $\tilde{y}_{di} \equiv \tilde{t}_{di} - \tilde{t}_{d,i-1}$.

We denote the flat vector regrouping all durations in temporal order by $y = (\tilde{y}_{11}, \ldots, \tilde{y}_{1n_1}, \ldots, \tilde{y}_{D1}, \ldots, \tilde{y}_{Dn_D})$ and let $t = (\tilde{t}_{10}, \ldots, \tilde{t}_{1,n_1-1}, \ldots, \tilde{t}_{D0}, \ldots, \tilde{t}_{D,n_D-1})$ such that the inter-trade interval i begins at time of day t_i and lasts y_i seconds.

Data Generating Process

The sequence of durations is given by

$$y_i = e^{x_i + m_i}\epsilon_i,$$

where x_i is a latent state process, m_i is a deterministic function of the time of day t_i and ϵ_i is a non-negative iid process with density $p_0(\epsilon_i)$ such that $E[\epsilon_i] = 1$. Both process ϵ_i and x_i are mutually independent.

The quantity $e^{x_i+m_i}$ is the conditional mean of duration y_i and determines the scale of the conditional distribution of y_i,

$$p(y_i|x_i, t_i) = e^{-(x_i+m_i)}p_0(y_i e^{-(x_i+m_i)}).$$

We call the density $p_0(\epsilon)$ the *normalized conditional duration density*. The latent process is a stationary zero-mean Gaussian AR(1) process:

$$x_1 \sim N\left(0, \sigma^2(1-\phi^2)^{-1}\right), \quad x_i|x_{i-1} \sim N\left(\phi x_{i-1}, \sigma^2\right).$$

Diurnal patterns are captured as an Lth order expansion of Chebyshev polynomials of first kind,

$$m_i = \sum_{l=0}^{L} \delta_l T_l \left(\frac{2t_i - t_{\text{open}} - t_{\text{close}}}{t_{\text{close}} - t_{\text{open}}} \right).$$

A Normalized Conditional Duration Density

We use a linear combination of exponential densities as normalized conditional duration density,

$$p_0(\epsilon) = \sum_{j=1}^{J} \alpha_j j \lambda e^{-j\lambda\epsilon}.$$

For the density to integrate to one, we require that $\sum_{j=1}^{J} \alpha_j = 1$. If the α_j are all positive, p_0 can be viewed as a finite mixture of exponential densities, but we do not require this. We only require that the normalized density be non-negative on $[0, \infty)$.

The normalization $E[\epsilon] = 1$ requires that

$$\lambda = \sum_{j=1}^{J} \alpha_j / j,$$

a condition used to substitute out λ from the parameters governing $p_0(\epsilon)$.

The resulting hazard rates are finite and positive for all $\epsilon \in [0, \infty)$. Figure 2 shows examples of hazard functions that can be captured by our normalized conditional density with just three terms. In the left panels, α_2 is negative; in the right panels, α_3 is.

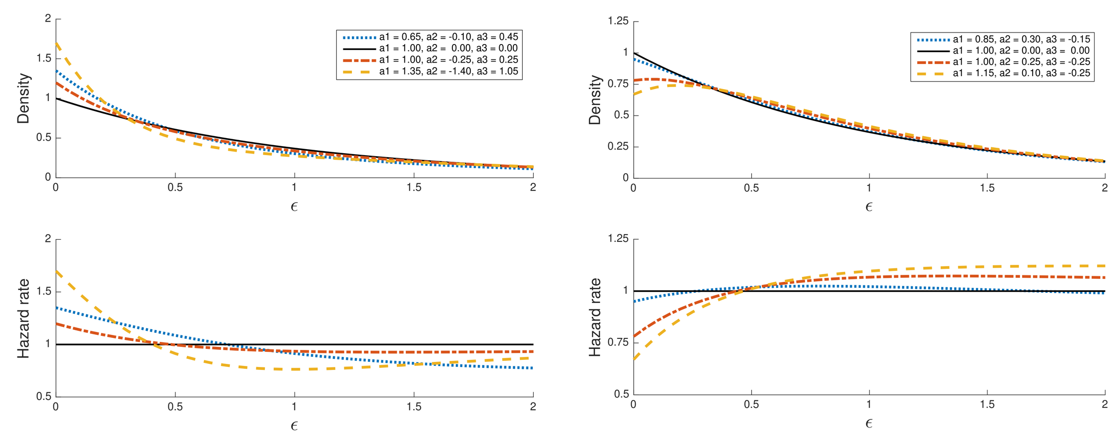

Figure 2: Examples of decreasing, increasing and non-monotonic hazard functions obtained for our proposed normalized conditional density with three terms.

Application

We demonstrate our flexible SCD model and our posterior simulation methods using a data set on IBM stocks from November 26, 1990 to December 21, 1990. We construct durations for transactions recorded between 10:00 am and 16:00 pm. This gives a time series of 14,794 trade durations. We fit our proposed SCD model using a 4 components normalized conditional density and a 7th order Chebyshev expansion for the diurnal pattern.

Estimation method: An Overview

Estimation is done via a MCMC approach by sampling from the joint posterior distribution of the parameters and the latent state variables. To simulate posterior samples, we implement a Metropolis-within-Gibbs algorithm with three blocks:

1. Sample from $x, \phi, \sigma | y, t, \alpha, \delta$.
2. Sample from $\delta | y, t, x, \phi, \sigma, \alpha$.
3. Sample from $\alpha | y, t, x, \phi, \sigma, \delta$.

Despite the flexible conditional distribution, we are able to draw the latent state sequence in a single block with high numerical efficiency.

Results

Approximate posterior moments and quantiles for parameters of the stochastic part of trading intensity are displayed in Table 1. We achieve very high numerical efficiency for all parameters, even for ϕ and σ which are more difficult to sample efficiently. With very high posterior probability, one term of the normalized conditional density is negative.

	$q_{0.01}$	$q_{0.5}$	$q_{0.99}$	Mean	Std	NSE	RNE
ϕ	0.9629	0.9730	0.9811	0.9728	0.0039	0.0000	0.2398
σ	0.0959	0.1138	0.1347	0.1140	0.0083	0.0001	0.2532
α_1	0.7156	0.8860	1.0575	0.8862	0.0718	0.0005	0.5628
α_2	0.0761	0.6656	1.2434	0.6643	0.2453	0.0016	0.5764
α_3	-3.0315	-2.2421	-1.4463	-2.2420	0.3346	0.0022	0.5703
α_4	1.3133	1.6903	2.0677	1.6914	0.1590	0.0011	0.5691

Table 1: Approximate posterior quantiles and moments computed for a sample 40,000 draws recorded after a burn-in period of 5000 draws. The numerical standard error (NSE) and the relative numerical efficiency (RNE) are computed for the posterior mean using the batch means method (Flegal and Jones, 2010).

We obtain inverted U-shaped diurnal patterns (left panel of Figure 3), as it is often the case in this literature, but a hazard function impossible to capture using usual conditional duration distributions. The estimated hazard rate first decreases and then gently increases with a ratio of the maximum to the minimum of less then three (right panel of Figure 3).

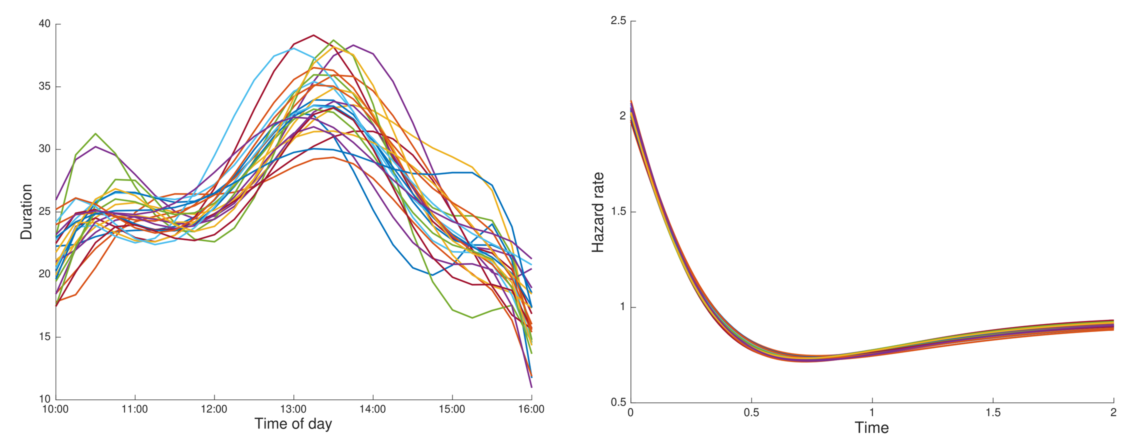

Figure 3: Diurnal pattern (left) and hazard function (right) for 25 posterior draw.

To check the specification of the model, we compare the distribution of residuals against the normalized conditional distribution implied by the parameters at the posterior mean. Figure 4 indicate that a conditional density with four terms provides a close fit to this sample of data.

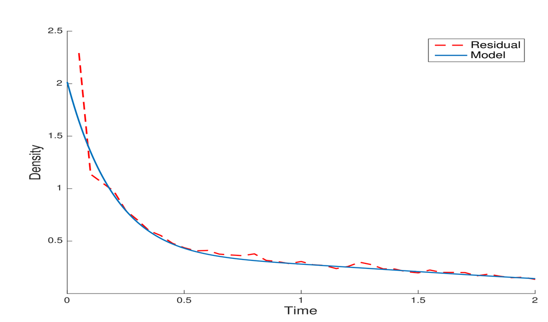

Figure 4: Empirical density vs Theoretical density.

Conclusions

We have proposed a new family of flexible conditional duration densities that can approximate durations distribution with moderate variation in their hazard rates. We argued theoretically that our distribution is realistic, and results from our application show that our proposed conditional distribution is empirically relevant.