# A Flexible Stochastic Conditional Duration Model

Samuel Gingras*[1] and William J. McCausland†[1]

[1]Département de sciences économiques, Université de Montréal

November 14, 2019

**Last version available here**

**Abstract**

We introduce a new stochastic duration model for transaction times in asset markets. We argue against the common practice of (imperfectly) aggregating related trades and instead use a Markov mixture model to account for both the short *cluster* durations between related trades and the more variable *regular* durations between unrelated trades. We introduce a normalized conditional distribution for regular durations that is flexible, and also expressible as a perturbation of an exponential distribution, which we argue is a theoretically sound first order model for unrelated financial transactions. Due in part to efficient draws of latent trade intensity, and despite the flexible distribution, numerical efficiency of posterior simulation is considerably better than that of previous studies where duration distributions are parametric. In an empirical application, we find that the conditional hazard function for regular durations varies much less than what is found in many studies. We attribute this to better (and probabilistic) classification of trades as related or not, and using flexible duration distributions instead of parametric distributions whose hazard functions have implausible behaviour near zero.

*Keywords*: Transaction data, Trade duration; Hazard function; Latent variable model; Non-Gaussian state space model; Markov Chain Monte Carlo

*JEL Codes*: C11; C41; C51; C58; G10

*samuel.gingras@umontreal.ca

†william.j.mccausland@umontreal.ca

The authors would like to thank David Benatia, the members of Université de Montréal Laboratory for Macroeconomic Policy, the participants at CIREQ PhD Students Conference 2018, European Seminar on Bayesian Econometrics (ESOBE) Conference 2018, Seminar on Bayesian Inference in Econometrics and Statistics (SBIES) Conference 2019, and Computing in Economics and Finance Conference 2019 for their valuable comments.

# 1 Introduction

Duration models for financial transactions describe the irregular arrival times of trades, or other events such as price changes. They are useful because trading intensity is a measure, among others, of market liquidity and, unlike count models, duration models use all the data available to infer trading intensity. They also shed light on market microstructure phenomena.

Some stylized facts about durations are well known. Trading intensity varies over time. Part of the variation is predictable given time of day and other observables. The remaining, stochastic, part is highly persistent. Trades often arrive in clusters, within a very short interval and without a marked difference in trading intensity before and after the cluster. We will refer to the very short durations between *related* trades as cluster durations, and all others as regular durations. The qualifier "related" allows for the possibility that durations between unrelated trades may be very short as well, by coincidence. While regular durations are usually outnumbered by cluster durations, they account for nearly all of the clock time the market is open and the literature attaches more importance to them. Although the definition of clusters in our model will be precise, the interpretation will be left vague, as there are different, and non-exclusive, reasons for clustering and we will not attempt here to distinguish between different types of clusters. We know that some are due to large market orders being matched with more than one limit order on the other side of the market, the transactions being recorded nearly simultaneously. Others are due to orders triggered by news, or orders triggering immediate reactions from trading algorithms.

In many data sets, transaction times are truncated to the nearest second. When it is important, we will use the term *recorded* duration to refer to a difference between truncated trade times.

Our paper makes three main contributions. The first contribution is to propose a new way to classify trades as being part of a cluster or not. We identify some problems with commonly used rules for classifying trades, including the practice of dropping all trades having a recorded duration equal to zero from the sample. Rather than apply a deterministic rule prior to inference, we develop a mixture model, featuring a latent indicator variable for every observed duration. As a result, we obtain, for each transaction, the posterior probability that it is part of a cluster. Our second contribution is to propose a new model for regular durations. We identify some undesirable features of existing parametric conditional duration densities. We then propose a new family of conditional distributions that is flexible, but also centered around a theoretically appealing first-order model, towards which we can shrink. Our third contribution is to draw the vector of state variables as a single Gibbs block. This has been done before for certain parametric models but not, as far as we know, for flexible models. The result is relatively high numerical efficiency.

Two other features of our model are appealing and uncommon, but not original. First,

we jointly estimate the regular diurnal pattern of intensity, together with other features of the model. Second, the latent state process is a discretized Ornstein-Uhlenbeck process, rather than a time-homogeneous autoregressive process. In this way, the variance of the state innovation depends on the elapsed time between two values.

We illustrate the phenomenon of clusters in Table 1, which shows the empirical distribution of recorded durations—which are integers by construction—from 0s to 5s. The counts are fairly smooth for values from 1s to 5s, but there is a lot of zero-inflation. In fact, the amount of zero-inflation is greater than it might appear, since the proportion of regular recorded durations equal to 0 is only about half of the unconditional duration "density" at zero.[1] Extrapolation of the counts from $1s$ to $5s$ back to $0s$ and dividing by 2 gives us a reasonable idea of the degree of zero-inflation, the number of zeros in excess of what we would expect from a smooth histogram. Some zeros are easier to identify as cluster durations or regular durations than others. For example, in a sequence of 50 observed durations of zero, the 25th is most likely a cluster duration; a duration of zero following and followed by multi-second duration is much less clear.

The most common approach to handle clustering is to remove transactions whose recorded duration is zero. This practice has been defended by claiming that transactions arriving within the same second are likely to be initiated by the same trader. Whether or not this is a good reason is debatable. If a market order matches nine limit orders, five times as many traders are getting their orders filled than if a market order matches a single limit order. Even if we accept, for the sake of argument, that one should count all transactions initiated by the same trader as a single transaction, the indiscriminate deletion of transactions where the recorded duration is zero is clearly going to eliminate many regular durations. Deleting observations whose value is both the mode and lower bound of a distribution is particularly distorting. The consequence is to overstate durations and understate intensity and liquidity, especially when that intensity is particularly high.

Instead of removing all zero durations, one can use the rule proposed by Grammig & Wellner (2002) and identify a cluster as a sequence of transactions within which durations are smaller than 1s and prices are monotonic (non-decreasing or non-increasing). Figure 2 shows the histograms of durations obtained by following this aggregation rule. Some of the zero transactions survive the aggregation, but not enough for them to be compatible with a smoothly varying hazard function near zero. This is unsurprising, as we would expect many of the regular durations recorded as zero to involve a pair of transactions that happen to exhibit monotonic price changes.

[1] If this is unclear, consider the probability that a true duration $y$ between two transactions will be recorded as integer $i$. Assume the fractional part of the first transaction time is uniform in $[0, 1)$. If $0 < y \leq 1$, the duration is recorded as 0 with probability $(1 - d)$ and 1 with probability $d$. If $i - 1 < d \leq i + 1$, it will be recorded as $i - 1$ with probability $\min(0, i - y)$, $i$ with probability $1 - |y - i|$ and $i + 1$ with probability $\min(0, y - i)$. If the true distribution is uniform on an interval containing $[0, 2]$, then a recorded duration of 1s is twice as likely as one of 0s. If it is merely slowly varying on $[0, 2]$, the factor of two is approximate.

Sequences of regular durations are usually modelled using multiplicative error models—named according to Engle (2002)—where the scale of the conditional distribution of duration depends on the history of an autoregressive process.[2] The scale may also depend on observables other than the history of the process, such as time of day. Such dynamic processes can capture the time-varying trading intensity and clustering of regular durations observed in high frequency financial data. We will use the term *normalized* conditional distribution to refer to the conditional distribution of duration divided by scale; typically, the normalized conditional distribution is fixed and has unit mean, but this is not always the case. Two basic models are the Autoregressive Conditional Duration model (ACD), introduced by Engle & Russell (1998), and the Stochastic Conditional Duration model (SCD) introduced by Bauwens & Veredas (2004). The ACD model, like (G)ARCH models of market volatility, is data driven, with the scale depending deterministically on past durations. The SCD model, like stochastic volatility models, is parameter driven, and the scale is a latent stochastic process. Although these two approaches differ in their specification of the scale process, they both describe the conditional distribution in terms of a time-varying scale and a normalized conditional density.[3]

In the case of ACD models, the unqualified term "conditional" implicitly means conditional on the history of observables. In the case of SCD models, it means conditional on the history of the state variable; in the usual case where the state is usually a Markov process, the conditional distribution depends only on the current value. Since the states in SCD models are unobserved, the conditional distribution of duration given only the history of observables is a mixture distribution, with the mixing distribution being the filtering distribution of the latent states. This gives SCD models some more flexibility; Bauwens & Veredas (2004) provide empirical evidence favouring the SCD model over a similar ACD model in the four data sets they analyzed.

The three most commonly used parametric conditional distributions in SCD and ACD models are the exponential, the gamma and the Weibull (Engle & Russell 1998, Bauwens & Giot 2000, Bauwens & Veredas 2004, Feng et al. 2004, Strickland et al. 2006, Men et al. 2015). Engle & Russell (1998), and many others since, conclude that the exponential, with its constant hazard function, is not flexible enough in the context of financial durations and prefer either the Weibull or the gamma distributions, as both distributions reduce to the exponential distribution for suitable values of the parameters. Note, however, that these conclusions come largely from studies in which recorded durations of zero are eliminated from the sample, an issue we explore in more detail below.

Both the Weibull and the gamma distribution has the property that when the hazard

[2]For recent surveys on the analysis of high-frequency financial durations using multiplicative error models see Pacurar (2008), Hautsch (2012) and Bhogal & Variyam (2019).

[3]Bauwens & Giot (2000) propose a logarithmic version of the original ACD model, avoiding parameter restrictions in the scale process. The SCD model feature a similar logarithmic specification for the time varying scale.

function is not constant, it is either zero at a duration of zero and strictly increasing; or infinite at a duration of zero and strictly decreasing. We will now argue that both features are unsuitable and unrealistic, and that the most plausible conditional distributions are those whose hazard functions do not exhibit such extreme variation. Rather, they should be bounded away from zero and infinity, and the ratio of supremum to infimum of the hazard function should not be too large.[4]

In queueing theory, a simple model for arrival times (of, say, customers at an ATM) is the Poisson process. It is a reasonable model when there are a large number of potential customers, acting independently and homogeneously in time, and the probability of any *given* customer arriving in a given time interval is much smaller than the probability of *some* customer arriving in the same interval. It has the property that the durations between arrivals are exponentially distributed, with constant hazard equal to the reciprocal of the mean. The constant hazard property is also known as the memoryless property: the probability of a customer arriving in the next minute does not depend on how long you have waited for one.

We have seen, however, that trade intensity in financial markets varies over time; there is intraday seasonality (called diurnal pattern), where durations tend to be shorter at the beginning and at the end of the trading day, and the widespread use of SCD and ACD models suggests that trading intensity varies also stochastically over time, although even here, much of the variation is predictable in the short term, due to high persistence. Hence, after conditioning on relevant predictors and latent states measuring trading intensity, we would expect the conditional distribution to be not too far from an exponential distribution—its hazard rate a function of this conditioning information—due to the large number of potential traders, most of whom are small and anonymous players. We suggest that a suitable normalized conditional distribution should have these two features: first, it should be able to approximate, with a small number of terms, a rich variety of distributions whose hazard function varies only moderately; second, have the flexibility to capture distributions whose hazard function fluctuates more widely, if the data support this strongly enough.

A mixture of exponential distributions avoids the above-mentioned problems with the gamma and the Weibull distribution; the hazard functions of these mixture distributions are bounded away from zero and infinity.[5] However, all mixtures of exponential distributions have decreasing hazard functions—see Barlow et al. (1963)—which is quite a restrictive feature. Moreover, all decreasing hazard functions cannot be captured by a mixture of

[4]The generalized gamma and the Burr distribution generalize the gamma and Weibull distribution. They were proposed as conditional distribution for ACD models by Lunde (1999) and Grammig & Maurer (2000). These distribution, unlike the gamma and the Weibull, allow for non-monotonic hazard functions. However, they retain the inconvenient property that their density and hazard function are bounded away from zero and infinity only for a restricted set of parameter values.

[5]DeLuca & Gallo (2004) used a mixture of two exponential distributions in ACD models and found that this specification provides a better fit than a Weibull distribution. DeLuca & Gallo (2009) again use a mixture of two exponential but allow mixture weights to depend on observable market activity.

exponential distributions. In simulations not reported here, we find that increasing the number of mixture components beyond two provides only modest additional flexibility. Only the components with the largest and the smallest hazard appears to be important, in the sense that the posterior distribution of the weights of all other mixture components are highly concentrated near zero. This suggests that if one allowed the density to be *any* linear combination of exponential distributions the posterior distribution would assign high probability to the region where at least one coefficient is negative.[6]

In this paper, we propose a SCD model for *regular* durations that features a flexible normalized density capable of matching a wide variety of distributions with moderately varying hazard functions. The normalized conditional distribution takes the form of a distortion of an exponential distribution, which we argued to be a reasonable first order model. The specification of the distortion takes the form of a continuous distribution on $[0, 1]$; the uniform distribution corresponds to no distortion. Thus, it is easy to center a prior distribution for the distortion around the uniform distribution; this in turn shrinks the duration distribution towards an exponential distribution. The distortion is a Bernstein density, a mixture of Beta distributions with parameters that depend only on the number of components. Such specification of the normalized density allows to approximate any continuous (normalized) density with positive support. Note that the resulting duration density can be viewed as a linear combination of exponential densities that is always positive, but with coefficients that can be negative, and thus, generalized the mixture of exponential distribution. Moreover, instead of aggregating related trades, we propose a Markov mixture model with two discrete states to account for both the short cluster duration between related trades and the more variable regular durations between unrelated trades. This gives a probabilistic classification of each short duration is classified as a cluster or a regular duration.

We provide posterior simulation methods for Bayesian inference in our proposed SCD model for regular durations featuring probabilistic classification of cluster durations.[7] Using the HESSIAN method proposed by McCausland (2012), we jointly draw the latent state sequence and the parameters of the latent state process in a single block. This contributes to high numerical efficiency. In contrast with much of the previous literature, we incorporate the diurnal pattern directly into our model and jointly estimate it together with latent variables and parameters.[8]

[6] Other mixture distributions have been proposed for the normalized conditional density. Wirjanto et al. (2013) did a Bayesian analysis of the SCD model, with leverage, using three types of mixtures with two components: two exponential, two Weibull and two gamma distributions.

[7] Bayesian analysis of the SCD model has also been carried out by Strickland et al. (2006) for the basic specification, and by Men et al. (2015) for a model with leverage. However, none of those extensions innovates with respect to the normalized conditional density; they use exponential, Weibull or gamma distributions.

[8] Veredas et al. (2002) and Brownlees & Vannucci (2013) jointly estimated the diurnal patterns and the parameters for ACD models using respectively a semi-parametric approach and MCMC methods within a Bayesian framework.

We conduct an artificial data experiment to test for the correctness of our posterior simulator, including its implementation in code. We then illustrate our methods in an application on trade durations using transaction data on two equities traded on the Toronto Stock Exchange.

Our flexible density combined with probabilistic classification of cluster durations allows us to precisely estimate the normalized regular duration density. We find that that the resulting hazard functions first decrease and then remain stable at their long rue value. Our results highlight important differences in the hazard functions of regular durations when classified using a probabilistic classification compared to Grammig and Weller's *ex ante* rule. The conditional hazard functions for regular durations are bounded away from zero and infinity, and varies much less than what is found in previous studies. Together, these results suggest that commonly used normalized conditional distributions have undesirable features. We also find that the numerical efficiency of our posterior simulation, for a model with a flexible conditional distribution, compares favourably with the numerical efficiency obtained using auxiliary mixture models, which rely on particular parametric distributions.

The rest of the paper is organized as follows. We describe the transaction data analyzed in the empirical application in Section 2. We present our SCD model in Section 3 and describe estimation methods in Section 4. The artificial data experiment and the empirical application are presented in Section 5. We conclude in Section 6.

## 2 Data

The data analyzed in this paper comes from Tick Data historical database which contains data for securities listed on major stock exchanges around the world. In our empirical application, we will illustrate our methods using transaction data for two equities traded on the Toronto Stock Exchange (TSX); the Royal Bank of Canada (RY), and the Potash Corporation of Saskatchewan (POT). For each equity, we analyze transactions over 10 consecutive trading days, starting on March 16, 2014 and ending on March 29, 2014.

The trading hours of the TSX are 9:30 am to 4:00 pm, Monday to Friday. A pre-open session is held from 7:30 am where traders can submit market-on-open orders, i.e. an order that will be executed at the day's opening price. During this period, the opening price is continuously updated and displayed. At 9:30 am, all matching orders are sequentially executed at the opening price before the beginning of the continuous trading session and any remaining orders are placed on the limit order book. Limit-on-close orders, i.e. limit orders executed at the market close, can be submitted during the last twenty minutes of the continuous trading session. An extended trading session is also held between 4:15 pm and 5:00 pm, where only orders at the last sale price are accepted.

Intraday transaction data contain the time at which transactions occurred (called the time stamp), the price at which the trades were executed and the traded volume measured

| | Trades | Mean | Std. | Max | Kurt. | C.V. | 0 % | 1 % | 2 % | 3 % | 4 % | 5 % |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RY | 85490 | 2.74 | 7.20 | 176 | 43.25 | 2.63 | 69.0 | 5.8 | 3.4 | 2.7 | 2.5 | 2.0 |
| POT | 68643 | 3.41 | 10.53 | 289 | 56.59 | 3.09 | 74.3 | 4.1 | 2.5 | 1.9 | 1.5 | 1.3 |

Table 1: Descriptive statistics of the cleaned sample for March 2014. The first column gives the number of transactions recorded. The next five columns give the mean, standard deviation, maximum value, kurtosis and coefficient of variation of trade durations. Times are measured to the nearest second. The last six columns give the percentage of trade durations recorded as 0s, 1s, 2s, 3s, 4s, and 5s, respectively.

in number of shares. The trading session for which the transaction was recorded is also available as well as additional fields specifying if the transaction was delayed or subsequently corrected. Depending on the exchange and the era, transaction and quotation times are recorded to various levels of precision, from the nanosecond to the second. In our sample, times are measured to the nearest second.

As shown by Brownlees & Gallo (2006) careful data cleaning is an important aspect of high-frequency duration analysis and standard filters should be applied in order to remove observations that are obviously erroneous or resulting from atypical market conditions, such as opening and closing. To this end, we delete entries identified as incorrect, corrected, delayed or canceled and restrict our sample to transactions recorded during the continuous trading sessions. We remove transactions with aberrant prices in the following way. We first filter transaction prices for each day in the sample to compute a local mean and standard deviation for each transaction. We do this using a centered rolling window of 50 observations that excludes the observation under consideration. Then we delete transactions for which the price deviated from the local mean by more than five local standard deviations. Descriptive statistics of the cleaned data are reported in Table 1. For each duration series, we report the number of observations, followed by the sample mean, standard deviation, maximum, kurtosis and coefficient of variation. For comparison, an exponential distribution has a kurtosis of 9 and a coefficient of variation of 1. We see that the empirical distributions are heavy tailed and overdispersed relative to the exponential. The second part of the table reports the percentage of durations recorded as 0s, 1s, 2s, 3s, 4s and 5s. This shows a strong degree of zero-inflation. However, it also suggests that around 3% of all durations (and a much larger percentage of regular durations) are regular durations recorded as zero.

While the data cleaning procedure we adopt is standard—similar filters have been used by several researchers—our practice of not removing zero durations is unusual. Dating back to the seminal paper of Engle & Russell (1998), durations equal to zero are generally discarded based on the assumption that they come from the same trader a split transaction, i.e. a market order matched against and filled with several limit orders on the opposite side of the market. Zero durations can also occur if many traders submit limit orders to be

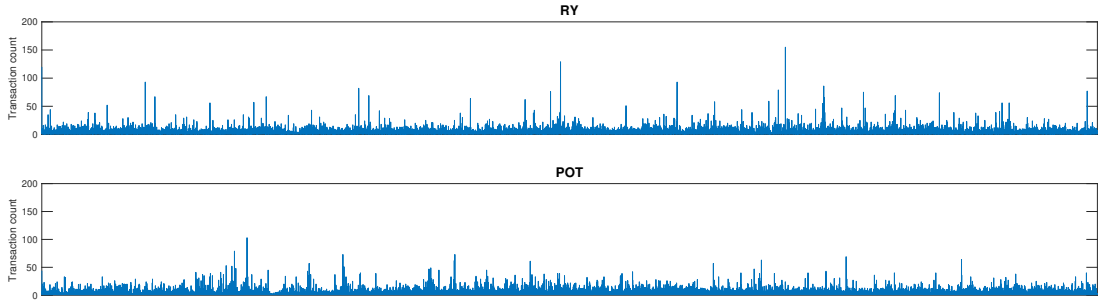
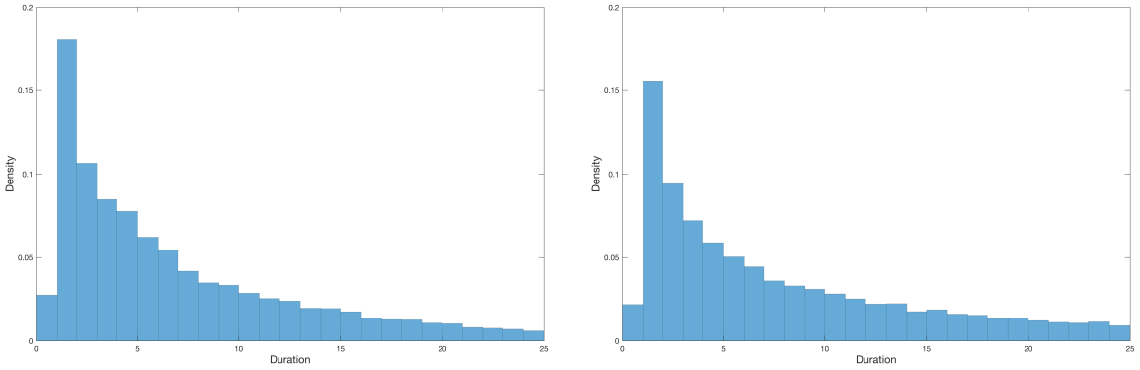



Figure 1: Number of transaction recorded for each time stamp during 10 consecutive days of trading for the Royal Bank of Canada (upper panel), and the Potash Corporation (lower panel).




(a) RY (b) POT

Figure 2: Histogram of durations following the filtering of zero durations using the Grammig and Wellner's rule. The histograms represent 25 seconds window starting from 0 with a bin width fixed at 1 second.

executed at a round price, as suggested by Veredas et al. (2002), or if important news synchronizes a flurry of trading, or if traders use algorithmic trading strategies that can be triggered by another trade. The pattern of durations associated with these different causes can look very similar.

Rather than propose a deterministic rule to distinguish between cluster and regular transactions, we do probabilistic inference for indicator variables in a mixture model where regular and cluster durations have different distributions. In the end, we obtain, for each transaction, the posterior probability that it is part of a cluster. While some transactions are easily classified—a transaction more than 2s away from any other is probably a regular transaction and the 25th transaction out of 50 recorded during the same second is probably a cluster transaction—others are more difficult. Two transactions within the same second, but more than 2s away from any other transaction, are more difficult to classify, but we will still be able to make a probabilistic statement.

# 3 A Stochastic Conditional Duration Model

In this section, we will present our model in several steps. We first describe a baseline model for regular durations, introducing a new family of normalized duration densities. We then extend the model to allow for both regular and cluster durations. Finally, we describe a family of prior distributions, which completes the model, giving us the joint distribution of data, latent variables and parameters. As with most models in the literature, time is continuous, which may not be fully suitable for data where trading intensity is high and times are rounded to the second. In Section 4.2 below, we discuss an adjustment to the model to make it more suitable for such data.

## 3.1 The Data Generating Process

We observe recorded transaction times for a given asset over a period of D consecutive trading days. All times of day are recorded in seconds after midnights. The opening and closing times are $t_{\text{open}}$ and $t_{\text{close}}$. For each day $d = 1, \ldots, D$, denote the ordered sequence of transaction times by $t_{d,0}, t_{d,1}, \ldots, t_{d,n_d}$ and construct the $n_d$ durations $y_{d,i} = t_{d,i} - t_{d,i-1}$, $i = 1, \ldots, n_d$. We condition on the first transaction time and the number of transactions in each day. For $d = 1, \ldots, D$, our model gives duration $y_{d,i}$ as

$$y_{d,i} = \exp(x_{d,i} + m_{d,i})\epsilon_{d,i}, \qquad i = 1, \ldots, n_d,$$

where $x_{d,i}$ is the value of a latent state process at time $t_{d,i-1}$, $m_{d,i}$ is the value of a function describing a diurnal pattern—equation (2) below—evaluated at time $t_{d,i-1}$, and $\epsilon_{d,i}$ is the contemporaneous value of a non-negative iid process with $\mathrm{E}[\epsilon_{d,i}] = 1$. The two random processes $x_{d,i}$ and $\epsilon_{d,i}$ are mutually independent.

The latent state process is a zero-mean Gaussian first order autoregressive process obtained by time-discretizing an Ornstein-Uhlenbeck process at the irregular time points where transactions occur. This gives

$$x_{d,1} \sim \mathcal{N}(0, \sigma^2), \quad x_{d,i+1} \,|\, x_{d,i}, y_{d,i} \sim \mathcal{N}(e^{-\phi y_{d,i}} x_{d,i}, \sigma^2(1 - e^{-2\phi y_{d,i}})) \quad i = 1, \ldots, n_d - 1, \quad (1)$$

where $\phi$ is a positive mean reversion parameter, and $\sigma$ is the marginal standard deviation of the latent state.

To model intraday seasonality, we follow the approach proposed by Eilers & Marx (1996) and specify the diurnal pattern $m_{d,i} \equiv m(t_{d,i-1})$ as a cubic B-spline function, a piecewise polynomial defined on a set of $M$ equally spaced knots $t_{\text{open}} = \kappa_1 < \cdots < \kappa_M = t_{\text{close}}$. The first and last knots have multiplicity 4 and the rest have multiplicity 1. This gives an

expansion that is the following linear combination of $L = M + 2$ B-spline basis functions:

$$m(t_{d,i-1}) = \sum_{l=1}^{L} \delta_l B_l(t_{d,i-1}), \tag{2}$$

where $B_l$ denotes the $l$-th basis cubic polynomial, and $\delta_l$ its coefficient. The basis functions depend on knot placements, and $B_l(t) = 0$ for $t \notin [\kappa_l, \kappa_{l+4}]$; note that the intervals where the basis functions are non-zero overlap. Setting the multiplicity of the knots $\kappa_1$ and $\kappa_M$ to 4 makes the values of the B-spline at $t_{\text{open}}$ and $t_{\text{close}}$ equal to the first and last coefficients, respectively. This is, $m(t_{\text{open}}) = \delta_1$ and $m(t_{\text{close}}) = \delta_L$. See de Boor (1978) and Dierckx (1993) for an in-depth treatment of B-splines.

We denote the density of $\epsilon_{d,i}$ by $p_\epsilon$ and call it the normalized duration density. The conditional density of duration $y_{d,i}$ given the contemporaneous value $x_{d,i}$ of the state, obtained through a change of variables, is

$$p(y_{d,i}|x_{d,i}, m_{d,i}) = e^{-(x_{d,i}+m_{d,i})} p_\epsilon(y_{d,i} e^{-(x_{d,i}+m_{d,i})}). \tag{3}$$

The quantity $e^{x_{d,i}+m_{d,i}}$ is the conditional mean of duration $y_{d,i}$ and gives the scale of the conditional distribution of $y_{d,i}$, the shape being determined by the normalized density $p_\epsilon$. We argued that some commonly used duration distributions are inappropriate for financial transactions data because of their restrictive or implausible hazard functions. To complete the model specification, we will now propose a new family of normalized densities capable of matching a wide variety of the most empirically relevant shapes of hazard functions.

### 3.2 A Normalized Density for Durations

We adopt a constructive approach similar to Ferreira & Steel (2006) and define the normalized density by applying an inverse probability integral transform to a parametric distribution. More precisely, we define the cumulative distribution function (c.d.f.) of the random process $\epsilon$ by

$$P_\epsilon(\epsilon) = G\big(F(\epsilon)\big),$$

where $F$ is a parametric continuous c.d.f. on $[0, \infty)$ with density $f = dF$ and $G$ is a flexible continuous c.d.f. on $[0, 1]$ with density $g = dG$. The distribution $P_\epsilon$ can be viewed as a distorted version of the parametric distribution given by $F$, with the distortion governed by $G$. When $G$ is uniform on $[0, 1]$, there is no distortion. As noted by Ferreira & Steel (2006), distributions defined in this way cover the entire class of continuous distributions since an appropriate choice of distorting distribution $G$ allows to choose values of $P_\epsilon$. This construction gives the following expression for the normalized density,

$$p_\epsilon(\epsilon) = f(\epsilon) g\big(F(\epsilon)\big). \tag{4}$$

We argued that the exponential distribution was a theoretically promising first order approximation of a conditional duration distribution. Thus we choose the exponential distribution for $F$, in the hope of being able to capture empirically relevant hazard functions using a distortion $G$ with a small number of terms. At the same time, we want to allow for very flexible $P_\epsilon$ if the data warrant it. For $G$, we choose a $J$'th order Bernstein polynomial which offers a lot of flexibility as Bernstein polynomials can approximate any continuous distribution on $[0,1]$ arbitrarily closely provided that the order of the polynomial is sufficiently large. The corresponding Bernstein density $g$ has the form of a $J$-component mixture of Beta densities where the parameters of each Beta component are fixed integers and the component weights are the coefficients of the distribution. Specifically,

$$g(z) = \sum_{j=1}^{J} \beta_j \,\mathrm{Beta}(z \,|\, j, J-j+1),$$

where $\beta_j \geq 0$ for all $j = 1, \ldots, J$, $\sum_{j=1}^{J} \beta_j = 1$, and

$$\mathrm{Beta}(z \,|\, j, J-j+1) = \frac{J!}{(j-1)!(J-j)!} z^{j-1}(1-z)^{J-j}, \qquad 0 \leq z \leq 1.$$

Bernstein polynomials from a partition of unity. Observe that for any choice of $J$, if $\beta_j = 1/J$ for all $j = 1, \ldots, J$, then $g$ is $U(0,1)$ and we get back the original, undistorted, exponential distribution. This characteristic also makes modelling the distortion distribution using Bernstein polynomials particularly attractive. We should point out that we do not treat $J$, the order of the Bernstein polynomial, as a parameter to estimate, but rather as being as part of the specification of the normalized density. For a discussion on Bayesian nonparametric density estimation using Bernstein polynomials with varying order $J$ see Petrone (1999*a*), Petrone (1999*b*) and Petrone & Wasserman (2002).

Putting the above expression for $g(z)$ in equation (4), together with the exponential distribution function $F(\epsilon) = 1 - e^{-\lambda\epsilon}$, gives the conditional duration density as the following polynomial in $e^{-\lambda\epsilon}$:

$$p_\epsilon(\epsilon) = \sum_{j=1}^{J} \beta_j \lambda \frac{J!}{(j-1)!(J-j)!} \left(1 - e^{-\lambda\epsilon}\right)^j \left(e^{-\lambda\epsilon}\right)^{J-j+1}. \tag{5}$$

We can write $p_\epsilon$ explicitly as this linear combination of exponential densities:

$$p_\epsilon(\epsilon) = \sum_{j=1}^{J} \alpha_j j \lambda e^{-j\lambda\epsilon}, \tag{6}$$

where $\alpha = (\alpha_1, \ldots, \alpha_J)$ is a linear function of the vector $\beta = (\beta_1, \ldots, \beta_J)$ of mixture weights satisfying $\sum_{j=1}^{J} \alpha_j = 1$. Non-negativity of the $\beta_j$ ensures the non-negativity of $p_\epsilon$, but the

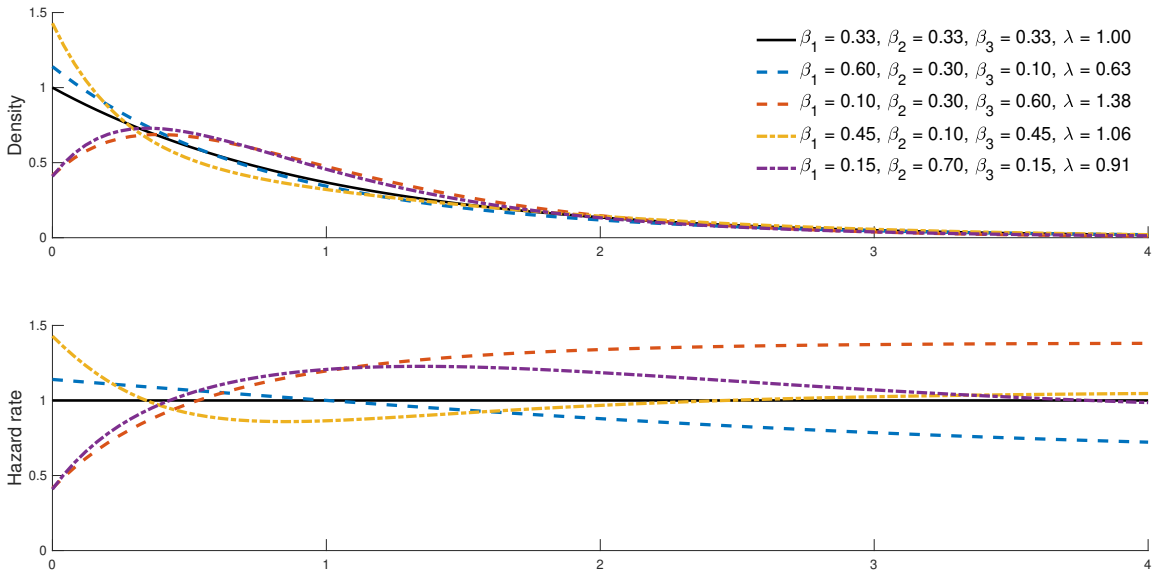



Figure 3: Examples of decreasing, increasing and non-monotonic hazard functions obtained for a four-term normalized density.

$\alpha_j$ are not necessarily all non-negative, so $p_\epsilon$ is not necessarily a mixture of $J$ exponential densities. The hazard function $h_\epsilon(\epsilon)$ for $p_\epsilon(\epsilon)$ is

$$h_\epsilon(\epsilon) = \sum_{j=1}^{J} j\lambda\alpha_j e^{-j\lambda\epsilon} \Big/ \sum_{j=1}^{J} \alpha_j e^{-j\lambda\epsilon} ,$$

which is bounded and positive on $[0, \infty)$. The hazard (or rate or inverse scale) parameter $\lambda$ of the original exponential distribution is the long run hazard of $p_\epsilon$, the limit of the hazard function $h_\epsilon(\epsilon)$ as $\epsilon$ goes to infinity. To normalize the duration distribution such that $\mathrm{E}[\epsilon] = 1$, we require

$$\lambda = \sum_{j=1}^{J} \frac{\alpha_j}{j}, \tag{7}$$

a condition that we use to substitute out $\lambda$ from the parameter vector governing $p_\epsilon$. This allows us to normalize the density without having to impose restrictions on the $\alpha_j$ or the $\beta_j$.

Figure 3 shows examples of decreasing, increasing and non-monotonic hazard rates that can be captured by our normalized density with $J = 3$ components. The solid lines are the density and (constant) hazard functions of the exponential distribution with mean equal to one, i.e. with no distortion. The other density and corresponding hazard functions are for various parameter values. In each case, the value of $\lambda$ is set to make the mean equal to one. Given the adding-up constraint, there are two degrees of freedom, the same as a mixture of two exponential distributions constrained to have unit mean. The dashed lines

show pairs of density and hazard functions where the hazard is monotonic. The distribution with increasing hazard is obtained by using an increasing sequence of parameters and the ones with a decreasing hazard by using the same sequence, but in a decreasing order. Pairs of density and hazard functions where the hazard function is not monotonic are shown using dash-dotted lines. Recall that no mixture of exponential distributions can have an increasing or non-monotonic hazard function.

With this specification, we are able to capture monotonic and non-monotonic hazard functions without extreme variation. This is important since if one could condition on enough relevant information, the conditional distribution of regular durations should be reasonably close to an exponential distribution, thus a constant hazard. However, our proposed distribution has the flexibility to capture more extreme variation using a larger number $J$ of terms, if warranted by the data. This make it appealing for duration modeling.

### 3.3 Modeling cluster Durations

The baseline model in Section 3.1 does not capture cluster durations and we would only recommend its use for transaction data that have been processed to retain only regular trade durations. In the following, we will refer to this version as the flexible SCD model (FSCD). We will now propose an extension which explicitly models cluster durations that we will called extended flexible SCD model (E-FSCD). This approach is more flexible than deterministic rules for identifying clusters of transactions, and instead delivers posterior probabilities of each duration as being a cluster duration. In most cases, the posterior probability will be very high or very low, in agreement with simple deterministic rules, but there will also be many cases that are difficult to classify, as we have seen. Based on queueing theory, the conditional distribution of regular durations should be close to an exponential distribution after having condition on the diurnal effect and the latent states measuring trading intensity, and therefore display a smooth hazard at zero. Hence, the distribution of regular duration can be used as addition source of information.

In the extended flexible SCD model, the conditional duration distribution is a mixture of two distributions; one to capture cluster durations and one to capture regular durations. It will be helpful to use data augmentation here, as is common practice: we add an additional latent indicator process $s_{d,i} \in \{0, 1\}$, which is first-order Markov, stationary and homogeneous, with

$$\Pr[s_{d,i+1} = k \,|\, s_{d,i} = l] = \xi_{kl}. \tag{8}$$

When $s_{d,i} = 0$, the duration is a cluster duration and when $s_{d,i} = 1$ it is regular. By stationarity, the marginal distribution (see Hamilton (1994)) of $s_{d,1}$ is given by

$$\Pr[s_{d,1} = 0] = (1-\xi_{11})(2-\xi_{00}-\xi_{11})^{-1} \quad \text{and} \quad \Pr[s_{d,1} = 1] = (1-\xi_{00})(2-\xi_{00}-\xi_{11})^{-1}. \tag{9}$$

The conditional distribution of duration $y_{d,i}$, is then given by

$$p(y_{d,i} \mid s_{d,i}, x_{d,i}, m_{d,i}) = \begin{cases} p_0(y_{d,i}) & s_{d,i} = 0 \\ e^{-(x_{d,i}+m_{d,i})} p_\epsilon(y_{d,i} e^{-(x_{d,i}+m_{d,i})}) & s_{d,i} = 1 \end{cases} \tag{10}$$

where $p_0$ is a density function with non-negative support. Given the definition of cluster durations, $p_0$ should concentrate most of its probability near zero. To this end, we select

$$p_0(y_{d,i}) = \zeta e^{-\zeta y_{d,i}} \mathbf{1}\{y_{d,i} < 1\},$$

an exponential distribution with hazard $\zeta$, truncated according to the aggregation rule suggested by Grammig & Wellner (2002). Under this data generating process, the classification of cluster duration is probabilistic, with posterior probabilities depending not only on the duration itself, but also on market conditions more broadly. Only the regular duration density depends on the latent state $x_{d,i}$ and during periods of high trading intensity—when $x_{d,i}$ is small—the probability of observing short regular durations increases and durations less than one second will have higher posterior probability of being regular.

### 3.4 Prior Distributions

We now describe a prior distribution for the various parameter vectors: $(\phi, \sigma)$ of the latent $x_{d,i}$ process; $\delta$, of the diurnal pattern; $\beta$, of the normalized duration density, $(\xi_{00}, \xi_{11})$ of the latent $s_{d,i}$ process and $\zeta$ of the cluster duration density. The vectors are mutually independent.

We choose a multivariate log-Normal distribution for the vector $(\phi, \sigma)$, i.e. $\theta = (\log(\phi), \log(\sigma))$ such that $\theta \sim \mathcal{N}(\bar{\theta}, \bar{\Sigma})$. This prior distribution is motivated by computational considerations as this parametrization is better suited for posterior simulation.

Following Lang & Brezger (2004), we specify a prior for $\delta$ that enforces smoothness without favoring any particular shape of diurnal pattern. Since the derivatives of B-spline functions are a linear combination of the finite differences of adjacent B-spline coefficients, smoothness of the diurnal pattern can be controlled by choosing the variance of the difference between adjacent $\delta$ coefficients. We select a Gaussian distribution for the initial coefficient, which gives the level of the diurnal pattern at opening, and a first order Gaussian random walk for the remaining adjacent coefficients,

$$\delta_1 \sim \mathcal{N}(\bar{\delta}, \bar{h}^{-1}), \qquad \delta_l \mid \delta_{l-1}, \tau \sim \mathcal{N}(\delta_{l-1}, \tau^{-1}), \quad l = 2, \ldots, L,$$

where $\tau$ is a precision parameter having a chi-square prior distribution, $\bar{s}\tau \sim \chi^2(\bar{\nu})$. The parameter $\tau$ controls the degree of smoothness of the diurnal pattern and will be estimated simultaneously with the B-splines coefficients. This choice of prior mitigates overfitting,

even with a large number of knots, as small differences of adjacent B-spline coefficients imply local smoothness.

For the weights of the Beta mixture governing the normalized duration density, we use a Dirichlet distribution,

$$(\beta_1, \ldots, \beta_J) \sim \text{Dirichlet}(\bar{M}\bar{\beta}_1, \ldots, \bar{M}\bar{\beta}_J),$$

specified in term of a location parameter $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_J)$, with $\sum \bar{\beta}_j = 1$ and $\bar{\beta}_j > 0$, and a concentration parameter $\bar{M} > 0$. Choosing $\bar{\beta} = (1/J, \ldots, 1/J)$ centers the prior distribution around the special case of an exponential distribution. The choice of $\bar{M}$ controls the amount of shrinkage towards it.

We choose independent Beta distributions for the diagonal elements of the transition matrix of $s_{d,i}$, $\xi_{ii} \sim \text{Beta}(\bar{a}_i, \bar{b}_i)$, $i = 0, 1$, and a Gamma distribution for the hazard parameter $\zeta$ of the cluster duration distribution, $\zeta \sim \text{Gamma}(\bar{a}_2, \bar{b}_2)$. The prior for $\zeta$ is conditionally conjugate.

### 3.5 Joint Likelihood

To conclude the exposition of the model, we give an expression for the joint density of all parameters, latent variables and observations. This makes explicit many conditional independence relationships in the extended flexible SCD model. The joint density of the baseline model is obtained by removing factors related to $\xi$, $\zeta$, and the $s_{d,i}$ in the following expressions. First, define vectors for day $d$ quantities as follows: $y_d = (y_{d,1}, \ldots, y_{d,n_d})$, $x_d = (x_{d,1}, \ldots, x_{d,n_d})$ and $s_d = (s_{d,1}, \ldots, s_{d,n_d})$. Let $y$, $x$ and $s$ be the flat vectors of all durations, states and indicators, respectively. Then the joint density is

$$p(\delta, \tau, \phi, \sigma, \beta, \xi, \zeta, s, x, y) = p(\delta \,|\, \tau)p(\tau)p(\phi, \sigma)p(\beta)p(\xi)p(\zeta) \prod_{d=1}^{D} p(s_d, x_d, y_d \,|\, \delta, \phi, \sigma, \beta, \xi, \zeta),$$

where

$$\begin{aligned} p(s_d, x_d, y_d \,|\, \delta, \phi, \sigma, \beta, \xi, \zeta) = \; & p(s_{d,1} \,|\, \xi) \prod_{i=1}^{n_d - 1} p(s_{d,i+1} \,|\, s_{d,i}, \xi) \\ & \times \; p(x_{d,1} \,|\, \sigma) \prod_{i=1}^{n_d - 1} \Big[ p(y_{d,i} \,|\, s_{d,i}, x_{d,i}, \delta, \beta, \zeta) p(x_{d,i+1} \,|\, x_{d,i}, y_{d,i}, \phi, \sigma) \Big] \\ & \times \; p(y_{d,n_d} \,|\, s_{d,n_d}, x_{d,n_d}, \delta, \beta, \zeta). \end{aligned}$$

# 4 Bayesian Inference

In this section, we describe posterior simulation methods for Bayesian inference in our flexible stochastic conditional model. We describe methods for the extended version; methods for the baseline model require only straightforward modifications. We end this section by presenting some recommended adjustments to the model and simulation methods when transaction times are rounded to the nearest second.

## 4.1 Posterior simulation

We describe here MCMC methods to sample the joint posterior distribution of parameters, latent indicators and state variables. The full posterior density is denoted $p(s, x, \delta, \phi, \sigma, \beta, \xi, \zeta \,|\, y)$. The method consists of six Gibbs blocks, updating $(x, \phi, \sigma)$, $(\delta, \tau)$, $\beta$, $s$, $\xi$ and $\zeta$. For the baseline model, we only require the first three blocks. Each of the next sections describes one block.

### Sampling from $p(x \,|\, \delta, \phi, \sigma, y)$

When there is strong posterior dependence between state variables (here, $x$) and the parameters of their dynamics (here, $\phi$ and $\sigma$), it can improve numerical efficiency to update both together in a single block. However, even drawing state variables as a single block, without the parameters, in non-linear non-Gaussian state space models is difficult. Several methods have been proposed, and many of them have been applied to draw latent volatilities in stochastic volatility models. Here, an additional difficulty is the non-parametric nature of the measurement distribution, which appears to rule out methods based on auxiliary mixture models.

Instead of updating the parameter vector $(\phi, \sigma)$ directly, we use the log parameterization $\theta$, transforming $\theta$ back to $(\phi, \sigma)$ whenever we need to evaluate $p(x \,|\, \theta)$. The parameter $\phi$ often has considerable posterior mass near zero, its lower bound. The log function maps the parameter space $(0, \infty)$ to the real line, which facilitates posterior simulation. Since the posterior distribution of $\theta$ is more nearly Gaussian than the posterior of the untransformed parameters, sampling is more numerically efficient.

Our joint proposal of $(x, \theta)$ consist of a proposal of $\theta^*$ drawn from an independent proposal density $q(\theta \,|\, y)$ followed by a proposal of $x^*$ given $\theta^*$ drawn from a conditional proposal density $q(x \,|\, \theta, \delta, \beta, s, y)$. Thus, current values of $(x, \theta)$ are updated to $(x^*, \theta^*)$ with probability

$$\min\left\{1, \frac{p(y \,|\, s, x^*, \delta, \beta)p(x \,|\, \theta^*)p(\theta^*)}{p(y \,|\, s, x, \delta, \beta)p(x \,|\, \theta)p(\theta)} \times \frac{q(\theta \,|\, y)q(x \,|\, \theta, \delta, \beta, s, y)}{q(\theta^* \,|\, y)q(x^* \,|\, \theta^*, \delta, \beta, s, y)}\right\}.$$

The proposal density $q(\theta \,|\, y)$ is a multivariate Student's $t$ density with $\nu$ degree of

freedom,

$$q(\theta \,|\, y) \propto \left[1 + \frac{1}{\nu}(\theta - \hat{\theta})' \hat{\Sigma}^{-1} (\theta - \hat{\theta})\right]^{-(\nu+1)/2},$$

where $\hat{\theta}$ and $\hat{\Sigma}$ are approximations of the mean and covariance of the marginal posterior distribution of $\theta$, computed during a burn-in period. We use a Student's $t$ rather than a Gaussian to ensure that the proposal is fat-tailed with respect to the conditional posterior distribution. For posterior simulation, we use a value $\nu = 10$.

During the burn-in period, we use an adaptive random walk Metropolis algorithm, as described in Roberts & Rosenthal (2009), in which the random walk covariance adapts as the chain goes along. At burn-in iteration $b = 1, \ldots, B$, we draw a proposal $\theta^*$ as follows: if $b \leq 25$, we draw $\theta^* \sim \mathcal{N}(\theta, \Sigma_0)$ where $\Sigma_0 = 0.005 I_2$; otherwise, we draw $\theta^*$ from the scale mixture of two Gaussian proposal. The first mixture component, with mixing weight $(1 - \pi)$, is $\mathcal{N}(\theta, 2.8\Sigma_b)$; the second, with weight $\pi$, is $\mathcal{N}(\gamma, \Sigma_0)$. Here, $\Sigma_b$ is the covariance of $\theta$ based on the first $b$ elements of the chain, and $\pi$ is a tuning parameter we set to 0.05. At the end of the burn-in period, we use the second half of the burn-in values to compute $\hat{\theta}$ and $\hat{\Sigma}$.

Since the state vectors $x_d$ (one for each day) are conditionally independent across days, we can sequentially draw each $x_d$ without sacrificing numerical efficiency. We draw $x_d \,|\, \theta^*, \delta, \beta, s_d, y_d$ for $d = 1, \ldots, D$ using the HESSIAN method proposed by McCausland (2012). The HESSIAN method uses a close approximation $q(x \,|\, y)$ of the conditional posterior distribution $p(x \,|\, y)$ for univariate models in which $p(y \,|\, x) = \prod_{i=1}^{n} p(y_i \,|\, x_i)$ and $x \sim \mathcal{N}(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega}^{-1})$ with $\bar{\Omega}$ tridiagonal. Tridiagonality of $\bar{\Omega}$ corresponds to $x$ being Markov but not necessarily homogeneous. The method is highly generic, as the only model-specific code required consists of a routine to evaluate $\log p(y_{d,i} \,|\, \delta, \beta, s_{d,i}, x_{d,i})$, and its first five derivatives with respect to $x_{d,i}$, at a given point. We compute exact values of these derivatives without requiring analytic expressions for them; instead, we exploit automatic routines to combine derivatives of primitive function using Faá di Bruno's rule, which is much easier.

### Sampling from $p(\delta, \tau \,|\, \phi, \sigma, x, y)$

We update $(\delta, \tau)$ using a Gibbs sampler with two blocks; we first draw $\tau$ given the current value $\delta$ from $\bar{\bar{s}}\tau \,|\, \delta \sim \chi^2(\bar{\bar{\nu}})$, where $\bar{\bar{s}} = \bar{s} + \delta' R \delta$ and $\bar{\bar{\nu}} = \bar{\nu} + L$, and then use a Metropolis-Hasting algorithm to update $\delta$ given $\tau$ based on a Gaussian proposal distribution. Here, $R = \Delta'\Delta$ where $\Delta$ is the matrix of first order backward difference operator, giving the restriction on the adjacent coefficients imply by the random walk prior.

We construct the proposal distribution using the centered parameterization of the model which implies a relocation of the expression that captures the diurnal pattern in the state equation. Note that favoring a centered parametrization over a non-centered has not been shown to decrease numerical efficiency (see Strickland et al. 2008). However, it simplifies

the construction of a satisfactory proposal distribution as Bayesian linear regression theory can be used instead of having to rely on numerical optimization to match the mean of the proposal with the mode of the log likelihood. By defining $\tilde{x}_{d,i} \equiv x_{d,i} + m_{d,i}$ the state equation can be rewritten as

$$\tilde{x}_{d,i+1} - B_{d,i+1}\delta = e^{\phi y_{d,i}}(\tilde{x}_{d,i} - B_{d,i}\delta) + \sqrt{\sigma^2(1 - e^{-2\phi y_{d,i}})}e_{d,i} \qquad e_{d,i} \sim \mathcal{N}(0,1),$$

where $B_{d,i} = (B_{1,L}(t_{d,i-1}), \ldots, B_{L,L}(t_{d,i-1}))$, and the state dynamics as the following Gaussian linear regression:

$$v = W\delta + e, \qquad e \sim \mathcal{N}(0, I_N),$$

where each element of th $N \times 1$ vector $v = (v_{1,1}, \ldots, v_{1,n_1}, v_{2,1}, \ldots, v_{2,n_2}, \ldots, v_{D,n_D})$ and the $N \times L$ matrix $W = (W_{1,1} \,|\, \cdots \,|\, W_{1,n_1} \,|\, W_{2,1} \,|\, \cdots \,|\, W_{2,n_2} \,|\, \cdots \,|\, W_{D,n_D})'$ are given by

$$\begin{aligned} v_{d,1} &= \tilde{x}_{d,1}/\sigma, & v_{d,i+1} &= (\tilde{x}_{d,i+1} - e^{-\phi y_{d,i}}\tilde{x}_{d,i})\Big/\sqrt{\sigma^2(1 - e^{-2\phi y_{d,i}})}, \\ W_{d,1} &= B_{d,1}/\sigma, & W_{d,i+1} &= (B_{d,i+1} - e^{-\phi y_{d,i}}B_{d,i})\Big/\sqrt{\sigma^2(1 - e^{-2\phi y_{d,i}})}. \end{aligned}$$

Hence, to sample from the conditional posterior distribution under the centered parametrization, we draw a proposal $\delta^* \sim \mathcal{N}(\bar{\bar{\delta}}, \bar{\bar{H}}^{-1})$, where $\bar{\bar{H}} = \tau R + W'W$ and $\bar{\bar{\delta}} = \bar{\bar{H}}^{-1}W'v$, and update the current value with probability

$$\min\left\{1, \exp\left(-\frac{\bar{h}}{2}\Big[(\delta_1^* - \bar{\delta})^2 - (\delta_1 - \bar{\delta})^2\Big]\right)\right\}.$$

With Gaussian random walk prior, we obtain a posterior distribution that is almost conditionally conjugate and Gaussian. In this case, our choice of proposal leads to an high acceptance frequency as only the prior density of the leading terms as to be evaluated to obtain the Hastings ratio.

**Sampling from** $p(\beta \,|\, s, x, y)$

A popular approach for sampling mixture weights is to use Gibbs sampling with auxiliary variables. However, this method adds an additional layer of complexity as each observation need to be assigned to a specific mixture components at each iteration. Even though performing data augmentation is a valuable tool for posterior simulations, the computational requirement is not negligible, especially for larger sample sizes such as those in the empirical application. Also, since the normalized distribution has to be rescaled for each vector of mixture weights to ensure that its mean is equal to one, this strategy does not work well in our case as the conditional distribution of mixture is no longer conditionally conjugate for a Dirichlet prior distribution.

Instead of using data augmentation, we sample from the conditional posterior distribu-

tion using an independent Metropolis step. Our proposal relies on the logistic transformation of the vector $\beta$,

$$\gamma(\beta) = \left[\log\left(\frac{\beta_1}{\beta_J}\right), \ldots, \log\left(\frac{\beta_{J-1}}{\beta_J}\right)\right],$$

mapping the $J$-dimensional simplex to $\mathbb{R}^{J-1}$. This transformation is better suited for posterior simulation, as there are no equality or positivity constraints to deal with. Hence, we draw $\beta$ using proposal of $\gamma$. We use a Student's $t$ density with 10 degree of freedom as proposal density $q(\gamma)$ and transform back proposal $\gamma^*$ to the original parametrization $\beta^*$ to evaluate the Hasting ratio for the accept/reject step. Proposal are accepted with probability

$$\min\left\{1, \frac{p(y \mid \beta^*, s, x)p(\beta^*)q(\gamma)}{p(y \mid \beta, s, x)p(\beta)q(\gamma^*)} \times \frac{\prod_{j=1}^J \beta_j^*}{\prod_{j=1}^J \beta_j}\right\},$$

where the second factor consists of Jacobian factors for the logistic transformation. The mean $\tilde{\gamma}$ and covariance $\tilde{\Sigma}$ of the proposal density are approximation of the marginal mean and variance of the posterior distribution of $\gamma$. These approximations are computed at the end of a burn-in period during whitch the $\gamma$'s are updated using an adaptive random-walk Metropolis scheme. More precisely, we draw a proposal $\gamma^*$ at the $b$'th burn-in iteration as follows: if $b \leq 25$, we draw $\gamma^* \sim \mathcal{N}(\gamma, \Sigma_0)$ where $\Sigma_0 = (0.1)^2 I_{J-1}/(J-1)$; otherwise, we draw $\gamma^*$ from a proposal distribution that is the scale mixture of two Gaussians. The first mixture component, with mixing weight $(1-\pi)$, is $\mathcal{N}(\gamma, (2.4)^2\Sigma_b/(J-1))$; the second, with weight $\pi$, is $\mathcal{N}(\gamma, \Sigma_0)$. The tuning parameter $\pi$ is set to 0.05.

**Sampling from** $p(s \mid \beta, \xi, \zeta, x, y)$

Latent indicators are updated via a single-move sampler where each $s_{d,i}$ is drawn conditional on all other values of the latent indicator $s_{-(d,i)}$. For $d = 1, \ldots, D$, the probabilities of drawing indicator $j = 0, 1$ for are proportional to

$$\Pr(s_{d,i} = j \mid s_{-(d,i)}, -) \propto p(y_{d,i} \mid s_{d,i} = j, -)\xi_{s_{d,i-1},j}\xi_{j,s_{d,i+1}}$$

for $i = 2, \ldots, n_d - 1$, with the following modification for the first and last indicators:

$$\begin{aligned}
\Pr(s_{d,1} = j \mid s_{-(d,1)}, -) &\propto p(y_{d,1} \mid s_{d,1} = j, -)\Pr[s_{d,1} = j \mid \xi_{00}, \xi_{11}], \\
\Pr(s_{d,n_d} = j \mid s_{-(d,n_d)}, -) &\propto p(y_{d,n_d} \mid s_{d,n_d} = j, -)\xi_{s_{d,n_d-1},j}.
\end{aligned}$$

Recall that we only have to update the indicator for durations that are less than one second since the density of cluster durations is truncated at one. As mentioned in Fruhwirth-Schnatter (2006), single-move sampling possesses a computational advantage over multi-move sampling as running a filter is not required. Given the sample size of a typical transaction data set, this property is not negligible to keep the estimation time reasonable.

**Sampling from $p(\xi \mid s)$**

Conditional on the path of the latent indicators, sampling the transition probabilities can be made in a single step using a Metropolis-Hasting algorithm. A Beta prior for $\xi_{00}$ and $\xi_{11}$ is nearly conditionally conjugate. In this case, constructing an adequate proposal distribution is straightforward. The target can be written as

$$p(\xi|s) \propto \left[\prod_{d=1}^{D} \frac{(1-\xi_{11})^{1-s_{d,1}}(1-\xi_{00})^{s_{d,1}}}{2-\xi_{00}-\xi_{11}}\right] \xi_{00}^{N_{00}+\bar{a}_0-1}(1-\xi_{00})^{N_{01}+\bar{b}_0-1} \times \xi_{11}^{N_{11}+\bar{a}_1-1}(1-\xi_{11})^{N_{10}+\bar{b}_1-1},$$

where $N_{lk} = \sum_{d,i} \mathbf{1}\{s_{d,i+1} = k, s_{d,i} = l\}$ is the count of the number of transition from $l$ to $k$. Hence, to sample from the conditional posterior, we draw a proposal from two independent Beta distribution

$$\xi_{00}^* \mid s \sim \text{Beta}(N_{00} + \bar{a}_0, N_{01} + \bar{b}_0), \qquad \xi_{11}^* \mid s \sim \text{Beta}(N_{11} + \bar{a}_1, N_{10} + \bar{b}_1),$$

and update the current value $(\xi_{00}, \xi_{11})$ with probability

$$\min\left\{1, \prod_{d=1}^{D} \left(\frac{1-\xi_{11}^*}{1-\xi_{11}}\right)^{1-s_{d,1}} \left(\frac{1-\xi_{00}^*}{1-\xi_{00}}\right)^{s_{d,1}} \left(\frac{2-\xi_{00}-\xi_{11}}{2-\xi_{00}^*-\xi_{11}^*}\right)\right\}.$$

**Sampling from $p(\zeta \mid s, y)$**

A Gamma prior for $\zeta$ is conditionally conjugate and gives a Gamma conditional posterior distribution. To make this more transparent, the target for this block is

$$p(\zeta \mid s, y) \propto \zeta^{N_0+\bar{a}_\zeta} \exp(-\zeta(N_0\bar{y}_0 + \bar{b}_\zeta)),$$

where $N_0 = \sum_{d,i} \mathbf{1}\{s_{d,i} = 0\}$ is the number of duration classified as a cluster duration and $\bar{y}_0 = (1/N_0)\sum_{d,i} y_{d,i} 1\{s_{d,i} = 0\}$ is the mean of cluster duration. This gives the following conditional posterior distribution $\zeta \mid s, y \sim \text{Gamma}(N_0 + \bar{a}_\zeta, N_0\bar{y}_0 + \bar{b}_\zeta)$.

## 4.2 Adjustment for recording precision

As mentioned in Section 2, transaction times can be recorded to various precisions depending on the stock exchange and the time period considered for a given sample, making the data discrete. However, duration models are traditionally specified using continuous distribution and the discreteness of the data is generally ignored when estimating the model. For data recorded with high precision (e.g. with at least millisecond precision), this approach is certainly valid, but for lower precision, the rounding error can lead to estimation bias and make continuous models unreliable for hypothesis testing (see. Grimshaw et al. 2005, Zhang et al. 2010, Schneeweiss et al. 2010).

Motivated by those issues, we propose minor adjustments to the model to take into account the interval uncertainty of the observed durations caused by the recording accuracy. In the case of transactions data, the discreteness of the available data is a by-product of the recording technology and we think that assuming a continuous data generating process is reasonable. For this reason, we decided to make the following adjustment rather then specify our model using discrete distribution, even if our application use data recorded with low precision.

The modification is designed especially for data rounded to the second. First, we propose to substitute the conditional density of regular duration by the probability mass function obtained by integrating equation (3) over the uncertainty imply by the second precision,

$$p(y_{d,i} \mid x_{d,i}) = \frac{1}{2} \int_{y_{d,i}-1}^{y_{d,i}+1} e^{-x_{d,i}} p_\epsilon(y e^{-x_{d,i}}) dy. \tag{11}$$

Except for duration of zero, the conditional density is integrated over an interval of 2 seconds centered at the observed duration.

Second, note that the parameter of the cluster distribution in the extended flexible SCD is not identified for data with second precision as similar integration leads to a discrete distribution with equal probability mass on zero and one for any value of $\zeta$. Hence, to avoid misclassification of duration equal to 1s, we propose to simplify the cluster distribution to a degenerate distribution at zero instead of using the discretization of the cluster density. This approximation misclassify some duration of 1s, but overall has negligible effect on the distribution of regular duration. The estimation of the extended flexible SCD model requires one less Gibbs block. Blasques et al. (2019) recently proposed to analyze financial duration series using several discrete distributions as well as their zero inflated version instead of continuous distributions.

Apart for integrating the interval uncertainty into the likelihood, estimating the flexible SCD model using data recorded with second precision require a straightforward adjustment to one block of our posterior simulator. Recall that the latent state process is obtained by time-discretizing an Ornstein-Uhlenbeck where transactions occur, thus $y_{d,i} = 0$ imply that $x_{d,i+1} = x_{d,i}$ (see equation (1)). In this case, sampling from $p(x \mid \delta, \phi, \sigma, y)$ imply drawing the reduced latent state vector $x$ of length equal to the number of non zero duration. Hence, the computation of the derivatives required by the HESSIAN method needs straightforward adjustment. We will use those modification in our empirical application.

# 5 Results

In this section, we first report results from an artificial data experiment meant to test for the correctness of our posterior simulators. We then illustrate the use of the flexible SCD model with an empirical application, using transaction data for two equities traded on the

Toronto Stock Exchange.

### 5.1 Getting it right

When posterior simulation methods are based on incorrect analysis or when there are coding errors in their implementation, we cannot rely on their results. Before applying our simulation methods to real data, we conduct a pre-data exercise to test their correctness. The tests described here are similar to those described in Geweke (2004) and the title of this section comes from the title of that paper.

We draw a sample from the *joint* distribution of parameters, latent state variables *and data* using the simulation methods described in Section 4 – with some slight modifications described below – augmented to include an additional Gibbs block updating the data $y$ from its conditional distribution given all parameters and latent variables. This additional computation is simply simulation of data from the model, taking care to keep track of the time of day $t_{d,i}$ of each transaction in order to evaluate the $m_{d,i}$. If the posterior simulation methods are conceptually sound and correctly implemented, then the marginal distribution of parameters in this sample is the same as their (known) prior distribution. This is a strong condition with many easily testable implications.

The dependence between the duration and the latent states prevents easy updating of the data conditional on all parameters and latent variables. This has the effect of increasing the amount of computation required to draw a new sample of artificial data. For the purpose of these simulations, we modify the latent state process described in (1) by

$$x_1 \sim \mathcal{N}(0, \sigma^2), \qquad x_{i+1} \,|\, x_i \sim \mathcal{N}(e^{-\phi} x_i, \sigma^2(1 - e^{-2\phi})).$$

Both specification shares the same parametrization in term of $(\phi, \sigma)$ and sampling from the posterior distribution can be done using the methods previously described with only a minor modification. However, this modified process allows us to directly simulate data from the model, which ease simulations, and to still verify the correctness of our implementation.

We set the number of components of the normalized distribution to $J = 3$ and use a B-spline function defined on two knots, $t_{\text{open}}$ and $t_{\text{close}}$, giving a diurnal pattern that is an expansion with $L = 4$ cubic polynomials. In this case, each B-spline basis function is positive on the whole interval $[t_{\text{open}}, t_{\text{close}}]$. Hence, the basis functions are global polynomials instead of local polynomials.

Recall that our data generating process is conditional on the first transaction time of each day and the number of transactions in each day; this is standard practice. However, it is not conditional on the event that the last transaction in each day occurs before $t_{\text{close}}$, a point not often emphasized. To avoid trades in simulations occuring after $t_{\text{close}}$, where the diurnal pattern is not defined, we choose the sample size $n$ (i.e. number of durations, not simulation sample size) and prior distributions such that the probability that the last

| $\vartheta$ | $\mathrm{E}[\vartheta] - \bar{\vartheta}$ | $\hat{\sigma}_{nse,\mathrm{E}[\vartheta]}$ | $t_{\text{stat}}$ | $\mathrm{E}[\vartheta^2] - \bar{\vartheta}^2$ | $\hat{\sigma}_{nse,\mathrm{E}[\vartheta^2]}$ | $t_{\text{stat}}$ |
|---|---|---|---|---|---|---|
| $\delta_1$ | 4.45e-04 | 5.13e-04 | 0.868 | 3.30e-03 | 3.59e-03 | 0.919 |
| $\tau$ | -2.79e-02 | 1.94e-02 | -1.441 | -5.56e+00 | 3.90e+00 | -1.425 |
| $\theta_1$ | 2.21e-04 | 3.27e-04 | 0.675 | -9.01e-04 | 1.31e-03 | -0.688 |
| $\theta_2$ | -2.71e-04 | 3.37e-04 | -0.805 | 1.91e-03 | 2.36e-03 | 0.808 |
| $\beta_1$ | 1.01e-04 | 1.13e-04 | 0.899 | 9.23e-05 | 1.13e-04 | 0.816 |
| $\beta_2$ | 7.88e-05 | 1.02e-04 | 0.776 | 4.16e-05 | 6.12e-05 | 0.679 |
| $\beta_3$ | -1.80e-04 | 8.52e-05 | -2.114 | -7.32e-05 | 3.47e-05 | -2.106 |
| $\xi_{00}$ | -3.72e-05 | 9.16e-05 | -0.407 | -2.13e-05 | 6.12e-05 | -0.348 |
| $\xi_{11}$ | 4.61e-05 | 1.09e-04 | 0.424 | 7.74e-05 | 1.81e-04 | 0.427 |
| $\zeta$ | 9.02e-04 | 2.36e-02 | 0.038 | 1.28e-01 | 4.77e+00 | 0.027 |

Table 2: Difference between sample and population first and second moments in the Getting it right experiment.

transaction of the day occurs after $t_{\text{close}}$ is extremely close to zero. We fix $D = 1$ and $t_0 = t_{\text{open}}$, choose a sample size of $n = 25$ observations, and set the length of the trading session to 2700 seconds. We select the following prior hyper-parameters: $\bar{\delta} = 3.5$, $\bar{h} = 200$, $\bar{s} = 5.0$ and $\bar{\nu} = 500$ for the prior distribution of the diurnal coefficients, $(\bar{a}_0, \bar{b}_0) = (200, 400)$ and $(\bar{a}_1, \bar{b}_1) = (500, 100)$ for that of the transition probabilities, $(\bar{a}_2, \bar{b}_2) = (200, 20)$ for that of the scale of the cluster distribution, $\bar{\theta} = (-3.5, -2.0)$ and $\bar{\Sigma} = 0.01 I_2$ for that of the latent intensity, and finally $\bar{\beta} = (0.50, 0.30, 0.20)$ and $\bar{M} = 250$ for that of the parameters of the flexible normalized density. This is a much tighter prior distribution and a much smaller number of observations than in our empirical application, to ensure high numerical precision and to be able to generate a very large sample in a reasonable amount of time.

Since we draw a new sample of artificial data at each iteration, the adaptive schemes implemented to approximate the mean and the covariance of $p(\theta \,|\, y)$ and $p(\gamma(\beta) \,|\, y)$ during the burn-in period do not work well. Instead, we use independence Metropolis-Hastings updates where the proposal distribution has the same mean and covariance as the parameter vector in question.

The initial draw of the parameter vector is from its prior distribution. We generate a sample of size $10^6$ for the analysis. Table 2 shows the results. The second columns gives the difference between the sample and population means of the various parameters; the third, the the numerical standard error (i.e. the simulation standard deviation quantifying error in finite simulation) of the sample mean. The t-statistics from the equality tests for the sample and population mean are reported in the fourth column. The fifth and sixth columns give the difference between the sample and population means of the squared values of the parameters and the numerical standard error for the sample second moment, followed by the t-statistics from the equality tests for the sample and population second moment.

Numerical standard errors are computed using the overlapping batch means method (see. Flegal & Jones 2010). Note that a Gaussian random walk prior implies a Gaussian marginal prior distribution centered at $\bar{\delta}$ for each of the B-splines coefficients with variance given as a function of $\tau$ and the position of the coefficient in the expansion. Hence, instead of reporting results for all coefficients that capture the diurnal pattern, we only report results for the first coefficient $\delta_1$ and the smoothing parameters $\tau$. Sample means and variances are close to true prior means and variances, relative to the numerical standard error. As can be seen from the table, the hypotheses of mean equality are not rejected at conventional significance levels for all but two parameters.

## 5.2 Empirical Application

We demonstrate our flexible SCD model and posterior simulation methods using transaction data from March 2014 for two equities traded on the Toronto Stock Exchange: the Royal Bank of Canada (RY) and the Potash Corporation (POT). See Section 2 for details about the data. We consider two different samples for each series: the subsample of regular durations obtained by applying Grammig and Wellner's rule, and the full sample. Subsamples are analyzed using FSCD models and full samples using E-FSCD models. This allows us to compare the density of regular durations obtained using an *ex ante* classification rule with that obtained using our proposed probabilistic classification.

### Model specification and prior distributions

We report results using a normalized conditional density with 3, 4, and 5 mixture components. Taking into account the mean normalization, the three component density has two degrees of freedom, the same as the generalized gamma and Burr distributions, and the same as a mixture of two exponential distributions.

We specify the diurnal pattern as a B-spline function defined on knots set on each half-hour, giving an expansion with $L = 16$ piecewise polynomials. Trade durations are often standardized using a cubic spline specification with knots set at each hour plus extra knots in the first and last half hour to better capture rapid decreases (resp. increases) of trading intensity near opening (resp. closing). Hence, our specification of diurnal patterns is similar to that used by others; recall, however, that we are estimating the parameters of the diurnal pattern jointly with other parameters, which is more unusual.

We use the same prior distribution for both series and model specifications. The prior mean of the log-transformed parameters of the latent intensity states $x_{d,i}$ is set to $\bar{\theta} = (-4.5, -1.5)$ and the values of the covariance matrix to $\bar{\Sigma}_{11} = 0.25$, $\bar{\Sigma}_{22} = 0.10$ and $\bar{\Sigma}_{12} = -0.01$. Table 3 shows the implied prior quantiles of the half-life[9] $t_{1/2}$ and the marginal standard deviation $\sigma$. We can see that this specification is quite diffuse. For the coefficients

[9]Two values of the latent state separated in time by the half-life have a correlation of 1/2.

| | $q_{0.01}$ | $q_{0.025}$ | $q_{0.1}$ | $q_{0.25}$ | $q_{0.5}$ | $q_{0.75}$ | $q_{0.9}$ | $q_{0.975}$ | $q_{0.99}$ |
|---|---|---|---|---|---|---|---|---|---|
| $t_{1/2}$ | 6.08 | 8.78 | 17.32 | 31.78 | 62.38 | 122.48 | 224.48 | 442.76 | 638.04 |
| $\sigma$ | 0.0218 | 0.0314 | 0.0619 | 0.1137 | 0.2231 | 0.4379 | 0.8037 | 1.5838 | 2.2852 |

Table 3: Prior quantiles of the half-life $t_{1/2}$, in seconds, and of the marginal standard deviation $\sigma$ of the latent state process.

of the B-spline function, we use $\bar{\delta} = 1.0$, $\bar{h} = 2$, $\bar{s} = 1.0$ and $\bar{\nu} = 100$. These values of $(\bar{s}, \bar{\nu})$ give a more diffuse prior distribution for the precision parameter $\tau$ than is typically used in the literature. We use a moderate number of knots to capture diurnal patterns, thus this choice implies slightly less shrinkage and ensures enough flexibility. For reasons outlined in the introduction, we choose $\bar{\beta} = (1/J, \ldots, 1/J)$, which centers the prior distribution around the exponential distribution, with its constant hazard. We set the concentration parameter to $\bar{M} = J + 1$. For the prior distribution of the transition probabilities of the latent indicator $s_{d,i}$ used in E-FSCD models, we select hyper-parameter values $(\bar{a}_0, \bar{b}_0) = (3, 2)$ and $(\bar{a}_1, \bar{b}_1) = (2, 3)$, which gives a quite diffuse distribution for the transition probability parameters. This complete the description of the prior distribution. Recall that for data with trade times recorded with second precision, the cluster distribution is replaced by a degenerate distribution with mass at zero since its parameter is no longer identified.

### Analysis of trade durations

Tables 6 and 7 shows results for the FSCD models, applied to the subsamples of regular durations. Tables 8 and 9 show results for the E-FSCD models applied to the full samples. For each parameter, we report the posterior mean and standard deviation, and the numerical standard error (NSE) and relative numerical efficiency (RNE) for the posterior mean. Defined in Geweke (1989), the relative numerical efficiency is a variance ratio that quantifies simulation error relative to a (hypothetical) iid sample from the posterior. Its product with the posterior sample size gives the size of an iid sample having the same numerical standard error. The estimated numerical standard errors of the posterior sample means are computed using the overlapping batch mean method (Flegal & Jones 2010). The posterior sample consists of 50,000 retained draws recorded after a burn-in period of 15,000 draws.

The numerical efficiency of $\phi$ and $\sigma$, which are difficult to sample efficiently, compares favorably to the numerical efficiency reported for the block sampling method in Strickland et al. (2006) [10]. The highest numerical efficiency they obtained for the parameters characterizing the latent intensity process is more than four times lower than the lowest numerical

[10] Note that the latent intensity state in their analysis follows a Gaussian AR(1) process with fixed autocorrelation and innovation variance rather than a time discretized Ornstein-Uhlenbeck process at irregular intervals.

| | | RY | | | | POT | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std | $q_{0.025}$ | $q_{0.975}$ | Mean | Std | $q_{0.025}$ | $q_{0.975}$ |
| $t_{1/2}$ | FSCD(3) | 78.9 | 12.6 | 57.5 | 106.8 | 196.2 | 35.3 | 136.9 | 273.9 |
| | FSCD(4) | 63.5 | 9.0 | 48.3 | 83.0 | 167.7 | 31.7 | 114.9 | 237.4 |
| | FSCD(5) | 63.3 | 9.4 | 48.2 | 85.0 | 169.9 | 32.1 | 114.9 | 239.8 |
| | E-FSCD(3) | 73.7 | 12.0 | 53.4 | 99.7 | 133.2 | 27.4 | 86.5 | 194.2 |
| | E-FSCD(4) | 79.2 | 12.9 | 56.8 | 107.5 | 159.5 | 31.8 | 105.8 | 230.1 |
| | E-FSCD(5) | 90.1 | 14.4 | 65.6 | 122.0 | 169.6 | 32.9 | 114.1 | 243.6 |
| $\sigma$ | FSCD(3) | 0.397 | 0.012 | 0.373 | 0.422 | 0.385 | 0.016 | 0.354 | 0.416 |
| | FSCD(4) | 0.400 | 0.012 | 0.377 | 0.423 | 0.392 | 0.016 | 0.360 | 0.423 |
| | FSCD(5) | 0.398 | 0.011 | 0.376 | 0.421 | 0.392 | 0.016 | 0.360 | 0.423 |
| | E-FSCD(3) | 0.425 | 0.013 | 0.399 | 0.450 | 0.431 | 0.017 | 0.398 | 0.464 |
| | E-FSCD(4) | 0.423 | 0.013 | 0.396 | 0.449 | 0.423 | 0.017 | 0.390 | 0.456 |
| | E-FSCD(5) | 0.413 | 0.013 | 0.388 | 0.439 | 0.417 | 0.017 | 0.384 | 0.451 |

Table 4: Posterior quantiles and moments of the half-life measured in second $t_{1/2}$ and the marginal standard deviation $\sigma$ of the latent intensity state process $x_{d,i}$.

efficiency in all our posterior simulations. Moreover, they used parametric distributions amenable to auxiliary mixture model simulation methods, while we allow for more flexible distributions. The numerical efficiency for the coefficients of the diurnal pattern $\delta$ and the transition probabilities $\xi$ are quite high, all being between 0.25 and 0.81, due in part to our efficient sampling of the latent intensity state. The mixture weights have the lowest numerical efficiency, presumably because of high posterior correlation between these and the latent state process. Even here, though, efficiency is quite good for non-linear non-Gaussian state space models.

Table 4 reports posterior quantiles and moments of the half-life $t_{1/2}$, in seconds, and of the marginal standard deviation $\sigma$ of the latent intensity state, for each model specification and both series. Persistence of the latent intensity process is fairly high, and more so for POT. There is a fair degree of posterior uncertainty about $t_{1/2}$, and the posterior distribution is somewhat sensitive to how regular durations are classified and the number of terms in the normalized duration density. The marginal standard deviation $\sigma$ is estimated more precisely. Its distribution is less sensitive to model specification and it even varies little across series.

Each panel in Figure 5 shows 100 posteriors draws of the diurnal pattern $m(t)$. Panels on the left show diurnal patterns obtained using subsamples of regular durations classified using Grammig and Weller's rule, together with the FSCD(5) model; panels on the right, the same for full samples of all durations and the E-FSDD(5) model. Upper panels are for the RY series, lower panels for POT. Comparing across columns, we see that the posterior

| | RY | | | | POT | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Std | $q_{0.025}$ | $q_{0.975}$ | Mean | Std | $q_{0.025}$ | $q_{0.975}$ |
| G&W | 1973 | | | | 954 | | | |
| E-FSCD(3) | 2995 | 74 | 2851 | 3141 | 1492 | 49 | 1397 | 1589 |
| E-FSCD(4) | 3073 | 77 | 2923 | 3227 | 1538 | 49 | 1443 | 1637 |
| E-FSCD(5) | 3187 | 76 | 3039 | 3336 | 1610 | 50 | 1513 | 1709 |

Table 5: Posterior quantiles and moments for the number of zero-durations classified as regular. The first row reports the number of zero-durations classified using Grammig and Weller's rule.

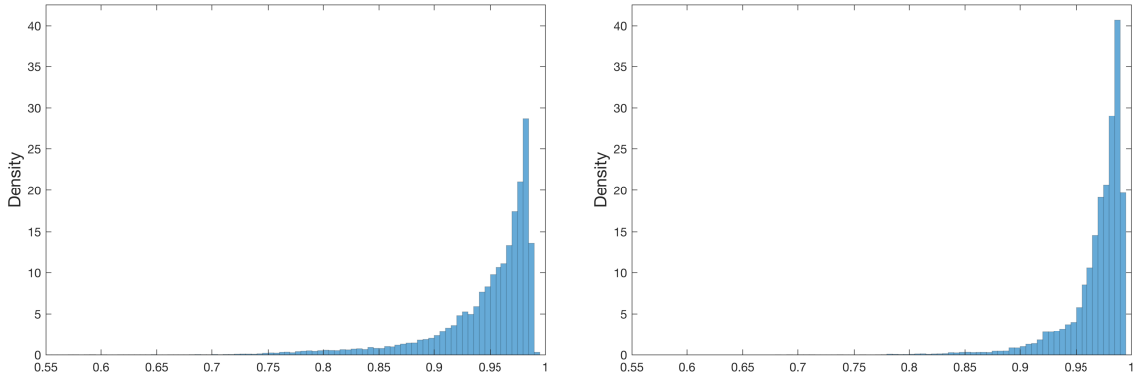



Figure 4: Posterior probability of each zero-duration to be classified as a cluster duration for an E-FSCD(5) model. The left histogram shows probabilities for the 58,954 trade durations equal to zero of the RY series and the right histogram for the 51,019 trade durations equal to zeros of the POT series. A point in both histogram corresponds to an observation.

distribution of the diurnal pattern is not very sensitive to the rule for classifying regular durations. Comparing across rows, we note that the diurnal pattern for RY (first row) is flatter than for POT (second row) which indicates less predictable variation in daily trading intensity. In all cases, we obtain the usual inverted U-shaped diurnal pattern found in most studies, with more trading intensity near opening and closing time in the middle of the trading day. There is some posterior variation in the diurnal patterns, but fairly small compared to the variation in intensity throughout the day. The posterior distributions of the diurnal pattern for other model specifications are very similar to these and are not reported.

At each posterior draw, we register, for each duration recorded as zero, whether it is classified as regular or not. Table 5 shows the posterior quantiles and moments of the number of such zero-durations classified as regular. The first row indicates the number of zero-durations classified as regular by Grammig and Weller's *ex ante* rule and the next three give posterior summaries for the E-FSCD(3), E-FSCD(4) and E-FSCD(5) models. Overall, our probabilistic approach classifies many more zero-durations as regular. The number of

zero-durations classified as regular increases with the number of terms in the normalized density; as flexibility increases, it is easier to fit a density with a sharp peak at zero.

For each individual zero-duration, we can compute the posterior probability that it is a cluster duration. Figure 4 gives, for each of the two full-sample series, a histogram of these posterior probabilities, under the E-FSCD(5) model. Analogous histograms for other E-FSCD model specifications are very similar and are not reported. To emphasize: each zero-duration is assigned to a histogram bin based on the posterior probability that it is a cluster duration; these are not histograms of draws from a posterior distribution. Individually, *all* zero-durations have a posterior probability of more than 0.5 of being cluster durations; for most, the probability is more than 0.95. At the same time, however, the probability that many zero-durations are regular is very high; there are a large number of zero-durations and their probabilities of being regular are far from negligible. These histograms demonstrate very well the danger of *ex ante* classification. As an isolated decision, each classification of a zero-duration as a cluster duration is the right choice (under symmetric and other reasonable loss functions). Collectively, however, they lead to a severe underestimation of the number of zero-durations that are regular. Observe that zero-durations for the RY series are more difficult to classify than those of the POT series, as the density for observations with probability lower than 0.90 is higher. For POT, zero-durations are less compatible with the lower average trading intensity shown in the diurnal pattern and are thus classified as cluster duration with higher probability.

Panels in Figure 6 show flexible density functions at posterior mean. Upper panels show density functions obtained with subsamples of regular duration classified using Grammig and Weller's rule, together with FSCD models; lower panels, the same for full samples of all durations and E-FSCD models. Left panels are for the RY series and right panels for the POT series. Comparing across panels, we see that the tail of the conditional distribution is not very affected by the number of components of the flexible density and that much of the variation is located near zero. For FSCD models (first row), increasing the number of components decreases the value of the density at zero. The variation is more pronounced effect for the RY series, giving a hump shaped density function. Recall that there is a larger difference between the number of zero-durations classified using Grammig and Weller's rule and the duration equal to one for the RY series that for POT (see Figure 2). Hence, increasing the number of components allows to better capture this feature of the data. However, the density obtained for E-FSCD models (second row) show that an hump shaped density is an artifact of the *ex ante* imperfect aggregation of the zero-durations. Compared to FSCD models, increasing the number of components increases the value of the density at zero, i.e. it allows to better capture the number of zero-durations classified as regular by our probabilistic approach. In this case, the density functions across series are more similar and the variation with the number of components is less pronounced. Panels in Figure 7 show the hazard functions corresponding to the density functions show in Figure 6. The organization

of the panels is the same in both figures. For FSCD models (first row), the hazard functions exhibit non-monotonic variations near zero while the hazard functions for E-FSCD models (second row) have a smoothly decreasing shape, an artifact of the misclassification of many regular durations as cluster durations.

Overall, a probabilistic classification of zero-durations combined with our flexible distribution gives a normalized regular duration density with a smoothly decreasing hazard near zero, which is a much more plausible behaviour for the hazard function of financial durations. Note that these last figures illustrate well the flexibility of our proposed normalized duration density; we are able to capture various hazard shapes with a single specification and a small number of degree of freedom.

# 6 Conclusion

Models in the literature are designed to capture regular durations, those between unrelated trades. They are not well suited to capture the observed clustering of related trades. Common practice is to try and aggregate clusters of related trades into single observations and then model only the "regular" durations between clusters. Even if trades could be classified as related or not without error, it is not clear that this would be desirable since it involves discarding information relevant to liquidity measurement and market microstructure. Furthermore, since it is not easy to tell related trades from unrelated trades that just happen to occur within the same second, errors of classification are inevitable. The most common rule to classify durations into cluster durations and regular durations is to aggregate all trades falling within the same second. This amounts to calling all recorded durations equal to zero cluster durations; and all others, regular durations. We have seen, however, that a large proportion of regular durations must be equal to zero—we don't always know which ones—and they are erroneously classified as cluster durations according to this rule. One consequence is to understate trade intensity and liquidity, especially at times of high intensity. Another is that the abrupt change in the number of recorded durations between 0s and 1s makes it difficult to fit a normalized conditional duration distribution. The rule suggested by Grammig & Wellner (2002) is clearly an improvement, but while it mitigates the problem, it does not entirely eliminate it.

The solution we proposed is to make our model a mixture model, with two discrete states representing cluster and regular durations. Identification of the two states comes from the very tight distribution of cluster durations and, more subtly, the shrinkage of the normalized conditional duration density towards an exponential distribution, which varies slowly near zero. However, we do not need to classify all durations perfectly. Instead, we obtain a probabilistic classification in the form of posterior probabilities for each binary state indicator.

Despite the probabilistic classification, we are able to estimate quite precisely the nor-

malized conditional duration density. Its hazard function does not exhibit the large changes near zero that occur using the Grammig & Wellner (2002) rule, which we claim is an artifact of misclassifying too many regular durations as cluster durations.

We introduced a flexible distribution for regular durations. Appealing to queueing theory, we argued that a good first order model for arrival times of unrelated trades is an exponential distribution, the distribution that naturally arises when there are many traders acting independently, each accounting for a small proportion of trades. Once allowances are made for variation in trading intensity by time of day or by market conditions, we expect conditional duration distributions not to differ too much from the exponential distribution. We introduce a normalized conditional distribution for regular durations that is flexible, and also expressible as a perturbation of an exponential distribution. This allows us to shrink towards the exponential distribution.

Due in part to efficient draws of latent trade intensity, and despite the flexible distribution, numerical efficiency of posterior simulation is considerably better than that of previous studies where duration distributions are parametric.

In the empirical application, we found that the conditional hazard function for regular durations varies much less than what is found in many studies. We attribute this to better (and probabilistic) classification of trades as related or not, and using flexible duration distributions instead of parametric distributions whose hazard functions have implausible behaviour near zero.

## References


Barlow, R. E., Marshall, A. W. & Proschan, F. (1963), 'Properties of probability distributions with monotone hazard rate', *The Annals of Mathematical Statistics* pp. 375–389.

Bauwens, L. & Giot, P. (2000), 'The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks', *Annales d'Économie et de la Statistique* **60**, 117–149.

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

Bhogal, S. K. & Variyam, R. T. (2019), 'Conditional duration models for high-frequency data: A review on recent developments', *Journal of Economic Surveys* **33**(1), 252–273.

Blasques, F., Holy, V. & Tomanova, P. (2019), Zero-inflated autoregressive conditional duration model for discrete trade durations with excessive zeros, Working Paper 004/111, Tinbergen Institute.

Brownlees, C. T. & Gallo, G. M. (2006), 'Financial econometric analysis of ultra-high frequency: Data handling concerns', *Computational Statisitcs and Data Analysis* **51**(4), 2232–2245.

Brownlees, C. T. & Vannucci, M. (2013), 'A bayesian approach for capturing daily heterogeneity in intra-daily durations time series', *Studies in Nonlinear Dynamics and Econometrics* **17**(1), 21–46.

de Boor, C. (1978), *A Pratical Guide to Splines*, Springer-Verlag New York.

DeLuca, G. & Gallo, G. M. (2004), 'Mixture processes for financial intradaily durations', *Studies in Nonlinear Dynamics and Econometrics* **8**(2).

DeLuca, G. & Gallo, G. M. (2009), 'Time-varying mixing weights in mixture autoregressive conditional duration models', *Econometric Reviews* **28**(1-3), 102–120.

Dierckx, P. (1993), *Curve and Surface Fitting with Splines*, Oxford University Press, Inc.

Eilers, P. H. C. & Marx, B. D. (1996), 'Flexible smoothing with b-splines and penalties', *Statistical Science* **11**(2), 89–121.

Engle, R. F. (2002), 'New frontier for arch models', *Journal of Applied Econometrics* **17**, 425–446.

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

Feng, D., Jiang, G. J. & Song, P. X.-K. (2004), 'Stochastic conditional duration models with "leverage effect" for financial transaction data', *Journal of Financial Econometrics* **2**(3), 390–421.

Ferreira, J. T. A. S. & Steel, M. F. J. (2006), 'A constructive representation of univariate skewed distribution', *Journal of the American Statistical Association* **101**(474), 823–829.

Flegal, J. M. & Jones, G. L. (2010), 'Batch means and spectral variance estimators in markov chain monte carlo', *The Annals of Statistics* **38**(2), 1034–1070.

Fruhwrith-Schnatter, S. (2006), *Finite Mixture and Markov Swithching Models*, Springer-Verlag New York.

Geweke, J. (1989), 'Bayesian inference in econometrics models using monte carlo integration', *Econometrica* **57**(6), 1317–1339.

Geweke, J. (2004), 'Getting it right: Joint distribution tests of posterior simulators', *Journal of the American Statistical Association* **99**, 799–804.

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

Grammig, J. & Wellner, M. (2002), 'Modeling the interdependence of volatility and intertransaction duration processes', *Journal of Econometrics* **106**, 369–400.

Grimshaw, S. D., McDonald, J., McQueen, G. R. & Thorley, S. (2005), 'Estimating hazard functions for discrete lifetimes', *Communications in Statistics - Simulation and Computation* **34**(2), 451–463.

Hamilton, J. D. (1994), State-space models, *in* R. F. Engle & D. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, chapter 50, pp. 3039–3080.

Hautsch, N. (2012), *Econometrics of Financial High-Frequency Data*, 1st edn, Springer-Verlag Berlin Heidelberg.

Lang, S. & Brezger, A. (2004), 'Bayesian p-splines', *Journal of Computational and Graphical Statistics* **13**(1), 183–212.

Lunde, A. (1999), A generalized gamma autoregressive conditional duration model, Technical report, Discussion Paper Aalborg University.

McCausland, W. J. (2012), 'The hessian method: Highly efficient simulation smoothing, in a nutshell', *Journal of Econometrics* **168**, 189–206.

Men, Z., Kolkeiwicz, A. W. & Wirjanto, T. S. (2015), 'Bayesian analysis of asymmetric stochastic conditional duration model', *Journal of Forcasting* **34**, 36–56.

Pacurar, M. (2008), 'Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature', *Journal of Economic Surveys* **22**(4), 711–751.

Petrone, S. (1999*a*), 'Bayesian density estimation using bernstein polynomials', *The Canadian Journal of Statistics* **27**(1), 105–126.

Petrone, S. (1999*b*), 'Random bernstein polynomials', *Scandinavian Journal of Statistics* **26**(3), 373–393.

Petrone, S. & Wasserman, L. (2002), 'Consistency of berntein polynomial posteriors', *Journal of the Royal Statistical Society Series B* **64**(1), 79–100.

Roberts, G. O. & Rosenthal, J. S. (2009), 'Examples of adaptive mcmc', *Journal of Computational and Graphical Statistics* **18**(2), 349–367.

Schneeweiss, H., Komlos, J. & Ahmad, A. (2010), 'Symmetric and asymmetric rounding: a review and some new results', *AStA Advances in Statistical Analysis* **94**(3).

Strickland, C. M., Forbes, C. S. & Martin, G. M. (2006), 'Bayesian analysis of the stochastic conditional duration model', *Computational Statisitcs and Data Analysis* **50**, 2247–2267.

Strickland, C. M., Martin, G. M. & Forbes, C. S. (2008), 'Parametrisation and efficient mcmc estimation of non-gaussian state space models', *Computational Statisitcs and Data Analysis* **52**(6), 2911–2930.

Veredas, D., Rodriguez-Poo, J. M. & Espasa, A. (2002), On the (intradaily) seasonality and dynamics of a financial point process: A semiparametric approach, CORE Discussion Papers 2002023, Université catholoque de Louvain, Center for Operations Research and Econometrics (CORE).

Wirjanto, T. S., Kolkeiwicz, A. W. & Men, Z. (2013), Stochastic conditional duration models with mixture processes, Technical report, Rimini Centre for Economic Analysis Working Papers series.

Zhang, B., Liu, T. & Bai, Z. (2010), 'Analysis of rounded data from depdendent sequences', *Annals ofr the Institute of Statistical Mathematics* **62**(6), 1143–1173.

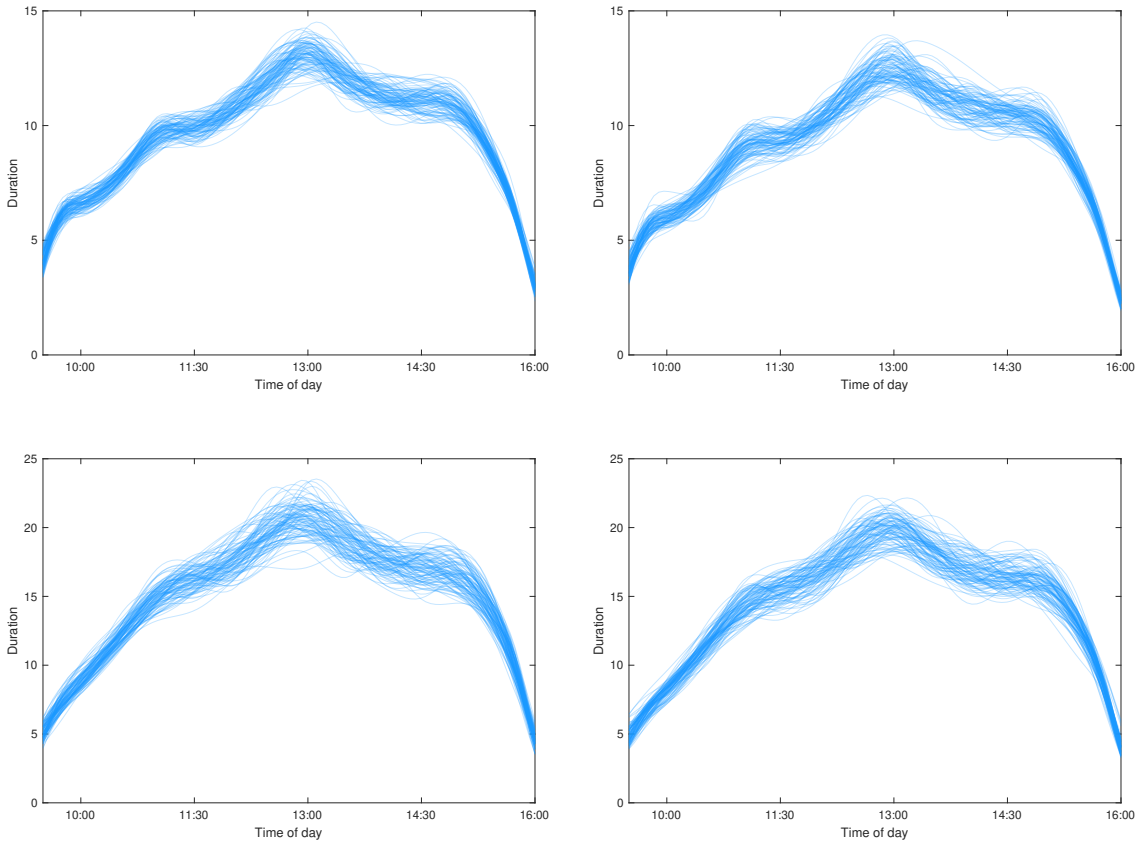



Figure 5: Diurnal pattern for 100 posteriors draws. Panels on the left show diurnal patterns obtained using subsamples of regular duration classified using Grammig and Weller's rule, together with the FSCD(5) model; panels on the right, the same for full samples of all durations and the E-FSCD(5) model. Upper panels are for the RY series, lower panels for POT.

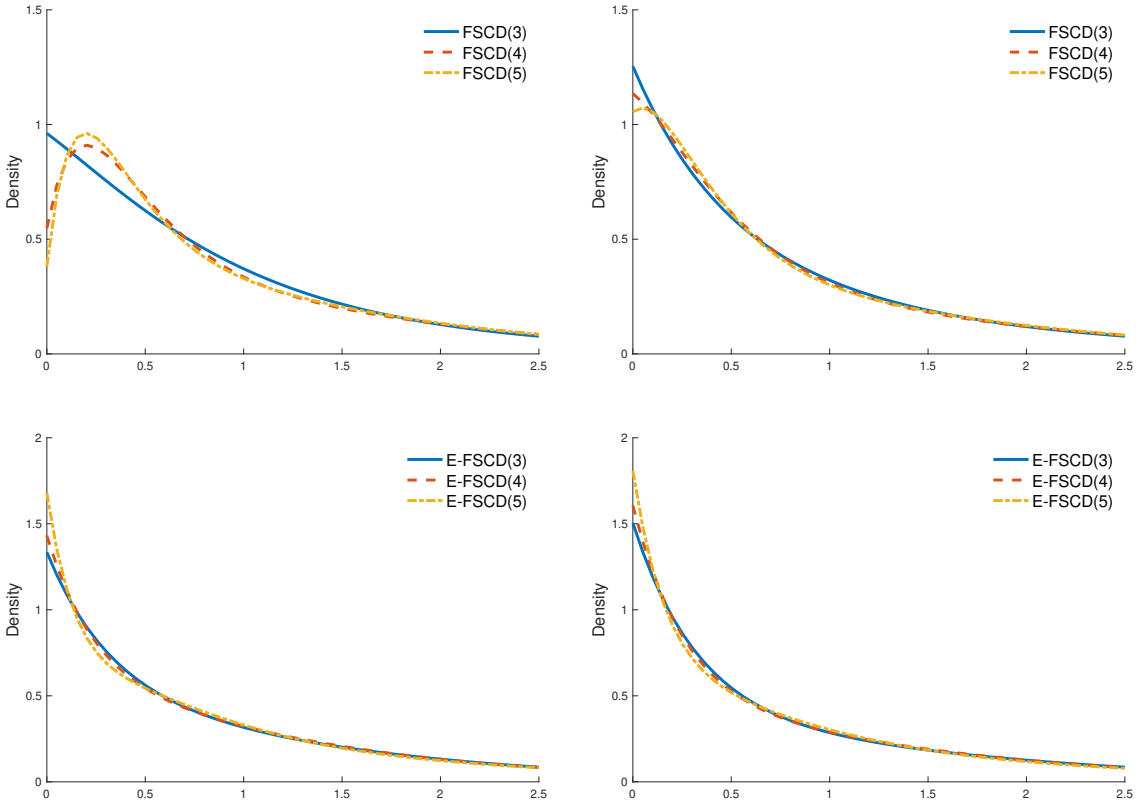



Figure 6: Normalized density functions at posterior mean. Upper panels show normalized density functions obtained with subsamples of regular duration classified using Grammig and Weller's rule, together with FSCD models; lower panels, the same for full samples of all durations and E-FSCD models. Panels on the right are for the RY series, panels on the left for POT.

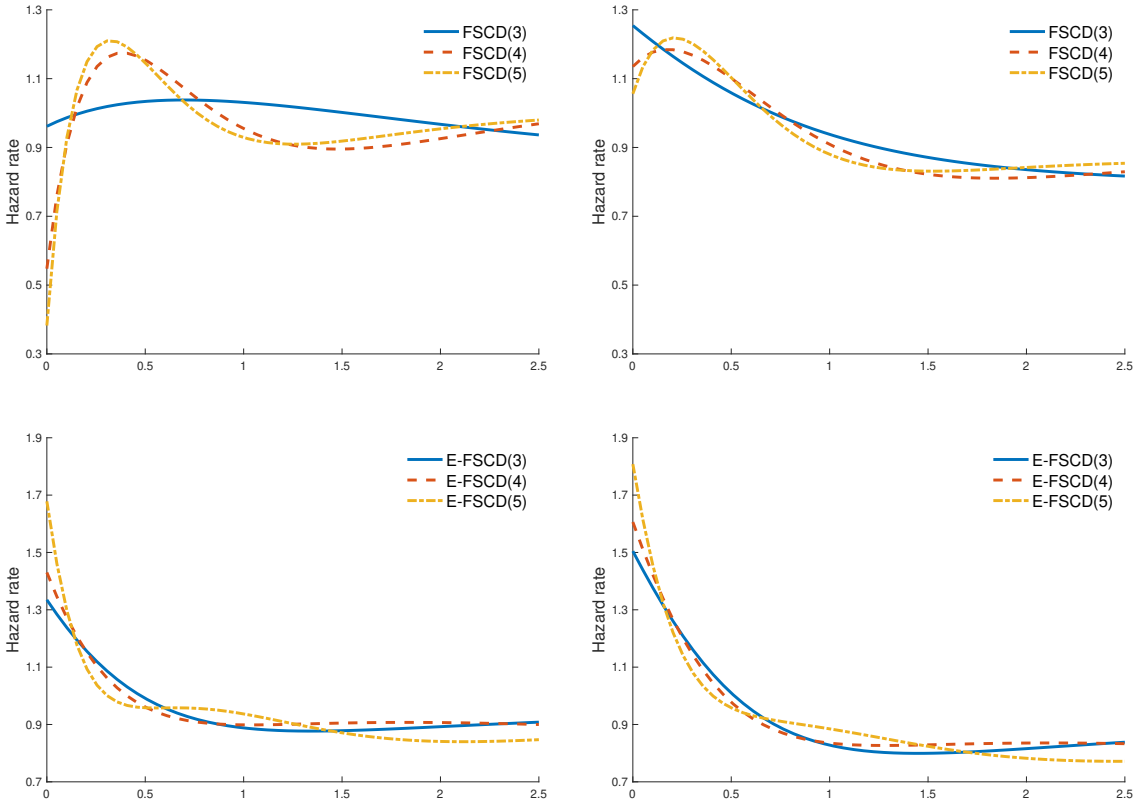



Figure 7: Hazard functions at posterior mean. Upper panels show hazard functions obtained with subsamples of regular duration classified using Grammig and Weller's rule, together with FSCD models; lower panels, the same for full samples of all durations and E-FSCD models. Panels on the right are for the RY series, panels on the left for POT.

| | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | -4.722 | 1.58e-01 | 1.94e-03 | 0.1334 | -4.508 | 1.40e-01 | 2.18e-03 | 0.0819 | -4.504 | 1.45e-01 | 2.62e-03 | 0.0611 |
| $\theta_2$ | -0.923 | 3.12e-02 | 3.11e-04 | 0.2003 | -0.917 | 2.92e-02 | 3.90e-04 | 0.1120 | -0.921 | 2.87e-02 | 4.25e-04 | 0.0913 |
| $\beta_1$ | 0.374 | 1.66e-02 | 2.29e-04 | 0.1056 | 0.129 | 1.05e-02 | 1.86e-04 | 0.0631 | 0.076 | 1.33e-02 | 3.42e-04 | 0.0303 |
| $\beta_2$ | 0.415 | 1.32e-02 | 1.39e-04 | 0.1800 | 0.512 | 1.44e-02 | 2.15e-04 | 0.0901 | 0.477 | 2.00e-02 | 5.72e-04 | 0.0245 |
| $\beta_3$ | 0.210 | 2.04e-02 | 2.22e-04 | 0.1683 | 0.019 | 1.35e-02 | 2.15e-04 | 0.0786 | 0.029 | 2.35e-02 | 7.85e-04 | 0.0180 |
| $\beta_4$ | | | | | 0.341 | 1.40e-02 | 2.85e-04 | 0.0485 | 0.201 | 2.16e-02 | 6.78e-04 | 0.0203 |
| $\beta_5$ | | | | | | | | | 0.217 | 2.14e-02 | 5.88e-04 | 0.0264 |
| $\delta_1$ | 1.393 | 8.89e-02 | 5.75e-04 | 0.4768 | 1.408 | 8.48e-02 | 5.59e-04 | 0.4605 | 1.407 | 8.45e-02 | 6.19e-04 | 0.3727 |
| $\delta_2$ | 1.842 | 7.43e-02 | 4.44e-04 | 0.5609 | 1.855 | 7.14e-02 | 4.91e-04 | 0.4232 | 1.855 | 7.18e-02 | 4.79e-04 | 0.4503 |
| $\delta_3$ | 1.860 | 6.56e-02 | 3.64e-04 | 0.6483 | 1.862 | 6.29e-02 | 4.16e-04 | 0.4573 | 1.861 | 6.30e-02 | 3.73e-04 | 0.5695 |
| $\delta_4$ | 2.022 | 5.96e-02 | 3.54e-04 | 0.5675 | 2.021 | 5.63e-02 | 3.64e-04 | 0.4790 | 2.024 | 5.66e-02 | 3.34e-04 | 0.5736 |
| $\delta_5$ | 2.312 | 6.14e-02 | 3.82e-04 | 0.5171 | 2.314 | 5.86e-02 | 3.90e-04 | 0.4517 | 2.313 | 5.86e-02 | 4.13e-04 | 0.4023 |
| $\delta_6$ | 2.269 | 5.98e-02 | 3.21e-04 | 0.6959 | 2.265 | 5.77e-02 | 3.65e-04 | 0.4981 | 2.263 | 5.70e-02 | 4.01e-04 | 0.4056 |
| $\delta_7$ | 2.367 | 5.95e-02 | 3.41e-04 | 0.6100 | 2.363 | 5.74e-02 | 3.87e-04 | 0.4401 | 2.362 | 5.77e-02 | 3.93e-04 | 0.4300 |
| $\delta_8$ | 2.512 | 6.09e-02 | 3.28e-04 | 0.6880 | 2.506 | 5.78e-02 | 4.17e-04 | 0.3842 | 2.503 | 5.81e-02 | 4.15e-04 | 0.3924 |
| $\delta_9$ | 2.615 | 6.18e-02 | 3.78e-04 | 0.5349 | 2.610 | 5.95e-02 | 4.45e-04 | 0.3580 | 2.608 | 5.93e-02 | 4.14e-04 | 0.4090 |
| $\delta_{10}$ | 2.488 | 6.08e-02 | 3.73e-04 | 0.5305 | 2.480 | 5.82e-02 | 4.16e-04 | 0.3917 | 2.479 | 5.81e-02 | 3.99e-04 | 0.4238 |
| $\delta_{11}$ | 2.429 | 6.03e-02 | 3.48e-04 | 0.6007 | 2.425 | 5.81e-02 | 4.06e-04 | 0.4099 | 2.422 | 5.76e-02 | 3.90e-04 | 0.4367 |
| $\delta_{12}$ | 2.410 | 6.04e-02 | 3.44e-04 | 0.6173 | 2.410 | 5.73e-02 | 3.68e-04 | 0.4830 | 2.409 | 5.74e-02 | 4.00e-04 | 0.4116 |
| $\delta_{13}$ | 2.409 | 6.12e-02 | 3.68e-04 | 0.5526 | 2.409 | 5.83e-02 | 4.06e-04 | 0.4126 | 2.408 | 5.81e-02 | 3.95e-04 | 0.4332 |
| $\delta_{14}$ | 2.147 | 6.59e-02 | 3.34e-04 | 0.7779 | 2.153 | 6.35e-02 | 4.16e-04 | 0.4664 | 2.157 | 6.30e-02 | 4.10e-04 | 0.4731 |
| $\delta_{15}$ | 1.752 | 7.20e-02 | 3.59e-04 | 0.8057 | 1.770 | 6.90e-02 | 4.28e-04 | 0.5204 | 1.774 | 6.91e-02 | 4.26e-04 | 0.5264 |
| $\delta_{16}$ | 1.074 | 9.51e-02 | 6.42e-04 | 0.4390 | 1.088 | 9.11e-02 | 6.68e-04 | 0.3722 | 1.088 | 9.06e-02 | 7.34e-04 | 0.3046 |

Table 6: Posterior mean and standard deviation obtained for the RY series with the subsample of regular durations classified using Grammig and Weller's rule, together with FSCD models. It is based on 50,000 posterior draw recorded after a burn-in period of 15,000 draws. Are reported respectively form left to right three, four and five component normalized duration specification with a cubic B-spline function defined for knots fix each half hour to capture the diurnal pattern. The columns gives posterior mean and standard deviation, and the numerical standard error (NSE) and relative numerical efficiency (RNE) for the posterior mean. The first two row are the log transformation of the mean reversion parameter $\phi$ and the log transformation of the the marginal standard deviation $\sigma$ of the state $x_{d,i}$; the following five are the mixture weights $\beta$ characterizing the normalized duration density; and the last sixteen are the coefficients $\delta$ of the diurnal pattern B-splines basis expansion.

| | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | -5.630 | 1.78e-01 | 1.56e-03 | 0.2633 | -5.471 | 1.88e-01 | 2.84e-03 | 0.0873 | -5.484 | 1.88e-01 | 2.40e-03 | 0.1233 |
| $\theta_2$ | -0.956 | 4.15e-02 | 3.28e-04 | 0.3209 | -0.938 | 4.11e-02 | 4.68e-04 | 0.1543 | -0.938 | 4.07e-02 | 4.64e-04 | 0.1537 |
| $\beta_1$ | 0.526 | 1.64e-02 | 1.75e-04 | 0.1752 | 0.312 | 2.26e-02 | 5.98e-04 | 0.0285 | 0.248 | 2.23e-02 | 4.78e-04 | 0.0437 |
| $\beta_2$ | 0.250 | 1.39e-02 | 1.27e-04 | 0.2395 | 0.350 | 2.66e-02 | 5.64e-04 | 0.0444 | 0.363 | 2.14e-02 | 4.01e-04 | 0.0569 |
| $\beta_3$ | 0.224 | 1.82e-02 | 1.77e-04 | 0.2125 | 0.082 | 2.87e-02 | 6.52e-04 | 0.0389 | 0.039 | 2.54e-02 | 4.97e-04 | 0.0524 |
| $\beta_4$ | | | | | 0.257 | 2.61e-02 | 7.02e-04 | 0.0276 | 0.198 | 2.22e-02 | 4.15e-04 | 0.0574 |
| $\beta_5$ | | | | | | | | | 0.151 | 2.62e-02 | 5.27e-04 | 0.0493 |
| $\delta_1$ | 1.619 | 1.04e-01 | 6.67e-04 | 0.4887 | 1.631 | 1.02e-01 | 6.81e-04 | 0.4456 | 1.631 | 1.02e-01 | 6.88e-04 | 0.4364 |
| $\delta_2$ | 1.956 | 8.62e-02 | 4.64e-04 | 0.6909 | 1.959 | 8.36e-02 | 4.36e-04 | 0.7346 | 1.961 | 8.43e-02 | 4.42e-04 | 0.7269 |
| $\delta_3$ | 2.193 | 7.87e-02 | 3.91e-04 | 0.8123 | 2.195 | 7.62e-02 | 4.43e-04 | 0.5905 | 2.196 | 7.69e-02 | 4.27e-04 | 0.6496 |
| $\delta_4$ | 2.499 | 7.30e-02 | 4.15e-04 | 0.6184 | 2.501 | 7.03e-02 | 3.93e-04 | 0.6399 | 2.500 | 7.06e-02 | 3.84e-04 | 0.6748 |
| $\delta_5$ | 2.730 | 7.44e-02 | 4.12e-04 | 0.6536 | 2.732 | 7.22e-02 | 4.07e-04 | 0.6297 | 2.729 | 7.23e-02 | 4.40e-04 | 0.5405 |
| $\delta_6$ | 2.792 | 7.34e-02 | 4.06e-04 | 0.6538 | 2.794 | 7.14e-02 | 4.13e-04 | 0.5963 | 2.790 | 7.18e-02 | 4.17e-04 | 0.5927 |
| $\delta_7$ | 2.847 | 7.35e-02 | 4.18e-04 | 0.6181 | 2.848 | 7.16e-02 | 4.45e-04 | 0.5171 | 2.847 | 7.21e-02 | 4.63e-04 | 0.4837 |
| $\delta_8$ | 2.979 | 7.46e-02 | 4.14e-04 | 0.6510 | 2.983 | 7.20e-02 | 4.50e-04 | 0.5128 | 2.982 | 7.23e-02 | 4.31e-04 | 0.5620 |
| $\delta_9$ | 3.057 | 7.56e-02 | 4.56e-04 | 0.5495 | 3.056 | 7.35e-02 | 4.88e-04 | 0.4533 | 3.055 | 7.38e-02 | 4.66e-04 | 0.5013 |
| $\delta_{10}$ | 2.948 | 7.44e-02 | 4.34e-04 | 0.5867 | 2.942 | 7.25e-02 | 4.60e-04 | 0.4969 | 2.941 | 7.25e-02 | 4.39e-04 | 0.5446 |
| $\delta_{11}$ | 2.866 | 7.41e-02 | 4.13e-04 | 0.6427 | 2.858 | 7.24e-02 | 4.44e-04 | 0.5317 | 2.860 | 7.22e-02 | 4.33e-04 | 0.5578 |
| $\delta_{12}$ | 2.836 | 7.41e-02 | 4.14e-04 | 0.6392 | 2.829 | 7.14e-02 | 4.29e-04 | 0.5537 | 2.829 | 7.17e-02 | 4.29e-04 | 0.5587 |
| $\delta_{13}$ | 2.847 | 7.57e-02 | 4.31e-04 | 0.6171 | 2.848 | 7.33e-02 | 4.33e-04 | 0.5723 | 2.844 | 7.29e-02 | 4.78e-04 | 0.4652 |
| $\delta_{14}$ | 2.630 | 8.10e-02 | 4.40e-04 | 0.6783 | 2.630 | 7.94e-02 | 5.20e-04 | 0.4664 | 2.630 | 7.98e-02 | 4.78e-04 | 0.5587 |
| $\delta_{15}$ | 2.224 | 8.77e-02 | 4.60e-04 | 0.7275 | 2.223 | 8.46e-02 | 4.94e-04 | 0.5860 | 2.222 | 8.48e-02 | 4.88e-04 | 0.6051 |
| $\delta_{16}$ | 1.494 | 1.18e-01 | 8.35e-04 | 0.3965 | 1.516 | 1.15e-01 | 9.05e-04 | 0.3236 | 1.514 | 1.14e-01 | 9.27e-04 | 0.3043 |

Table 7: Posterior mean and standard deviation obtained for the POT series with the subsample of regular durations classified using Grammig and Weller's rule, together with FSCD models. It is based on 50,000 posterior draw recorded after a burn-in period of 15,000 draws. Are reported respectively form left to right three, four and five component normalized duration specification with a cubic B-spline function defined for knots fix each half hour to capture the diurnal pattern. The columns gives posterior mean and standard deviation, and the numerical standard error (NSE) and relative numerical efficiency (RNE) for the posterior mean. The first two row are the log transformation of the mean reversion parameter $\phi$ and the log transformation of the the marginal standard deviation $\sigma$ of the state $x_{d,i}$; the following five are the mixture weights $\beta$ characterizing the normalized duration density; and the last sixteen are the coefficients $\delta$ of the diurnal pattern B-splines basis expansion.

| | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | -4.653 | 1.62e-01 | 2.70e-03 | 0.0718 | -4.726 | 1.62e-01 | 2.72e-03 | 0.0707 | -4.855 | 1.57e-01 | 2.75e-03 | 0.0655 |
| $\theta_2$ | -0.857 | 3.13e-02 | 4.32e-04 | 0.1048 | -0.862 | 3.16e-02 | 4.57e-04 | 0.0952 | -0.884 | 3.14e-02 | 4.45e-04 | 0.0995 |
| $\beta_1$ | 0.471 | 1.03e-02 | 1.82e-04 | 0.0634 | 0.418 | 1.91e-02 | 3.23e-04 | 0.0696 | 0.357 | 1.08e-02 | 2.77e-04 | 0.0304 |
| $\beta_2$ | 0.183 | 1.67e-02 | 2.59e-04 | 0.0832 | 0.126 | 2.18e-02 | 3.45e-04 | 0.0798 | 0.013 | 1.12e-02 | 2.24e-04 | 0.0503 |
| $\beta_3$ | 0.347 | 1.87e-02 | 3.55e-04 | 0.0556 | 0.275 | 2.29e-02 | 3.37e-04 | 0.0924 | 0.350 | 1.90e-02 | 4.22e-04 | 0.0406 |
| $\beta_4$ | | | | | 0.182 | 2.48e-02 | 4.25e-04 | 0.0679 | 0.059 | 2.02e-02 | 4.51e-04 | 0.0399 |
| $\beta_5$ | | | | | | | | | 0.220 | 1.78e-02 | 4.57e-04 | 0.0303 |
| $\delta_1$ | 1.326 | 9.28e-02 | 6.77e-04 | 0.3761 | 1.317 | 9.44e-02 | 6.80e-04 | 0.3848 | 1.312 | 9.56e-02 | 6.62e-04 | 0.4166 |
| $\delta_2$ | 1.774 | 7.82e-02 | 5.11e-04 | 0.4691 | 1.766 | 7.88e-02 | 5.51e-04 | 0.4094 | 1.756 | 7.95e-02 | 5.22e-04 | 0.4646 |
| $\delta_3$ | 1.772 | 6.87e-02 | 4.02e-04 | 0.5830 | 1.771 | 6.94e-02 | 4.00e-04 | 0.6014 | 1.769 | 7.02e-02 | 4.17e-04 | 0.5676 |
| $\delta_4$ | 1.947 | 6.20e-02 | 3.50e-04 | 0.6271 | 1.945 | 6.29e-02 | 3.75e-04 | 0.5631 | 1.940 | 6.39e-02 | 3.92e-04 | 0.5319 |
| $\delta_5$ | 2.262 | 6.40e-02 | 4.35e-04 | 0.4327 | 2.256 | 6.44e-02 | 4.27e-04 | 0.4551 | 2.247 | 6.60e-02 | 4.17e-04 | 0.4995 |
| $\delta_6$ | 2.220 | 6.25e-02 | 3.70e-04 | 0.5719 | 2.218 | 6.35e-02 | 3.75e-04 | 0.5713 | 2.214 | 6.44e-02 | 4.15e-04 | 0.4809 |
| $\delta_7$ | 2.321 | 6.25e-02 | 3.89e-04 | 0.5166 | 2.319 | 6.34e-02 | 3.86e-04 | 0.5392 | 2.310 | 6.45e-02 | 4.28e-04 | 0.4531 |
| $\delta_8$ | 2.467 | 6.32e-02 | 3.73e-04 | 0.5746 | 2.465 | 6.45e-02 | 4.05e-04 | 0.5061 | 2.458 | 6.48e-02 | 4.20e-04 | 0.4748 |
| $\delta_9$ | 2.568 | 6.42e-02 | 4.42e-04 | 0.4221 | 2.565 | 6.54e-02 | 4.05e-04 | 0.5199 | 2.554 | 6.63e-02 | 4.51e-04 | 0.4322 |
| $\delta_{10}$ | 2.446 | 6.30e-02 | 3.80e-04 | 0.5513 | 2.445 | 6.41e-02 | 4.35e-04 | 0.4357 | 2.435 | 6.47e-02 | 4.33e-04 | 0.4469 |
| $\delta_{11}$ | 2.385 | 6.28e-02 | 4.09e-04 | 0.4703 | 2.381 | 6.35e-02 | 4.19e-04 | 0.4596 | 2.377 | 6.47e-02 | 4.25e-04 | 0.4633 |
| $\delta_{12}$ | 2.369 | 6.29e-02 | 4.01e-04 | 0.4914 | 2.363 | 6.34e-02 | 4.12e-04 | 0.4748 | 2.357 | 6.45e-02 | 3.82e-04 | 0.5693 |
| $\delta_{13}$ | 2.367 | 6.34e-02 | 4.04e-04 | 0.4928 | 2.363 | 6.41e-02 | 3.99e-04 | 0.5150 | 2.351 | 6.54e-02 | 4.40e-04 | 0.4417 |
| $\delta_{14}$ | 2.094 | 6.92e-02 | 4.22e-04 | 0.5377 | 2.088 | 7.02e-02 | 4.02e-04 | 0.6098 | 2.075 | 7.12e-02 | 4.25e-04 | 0.5628 |
| $\delta_{15}$ | 1.653 | 7.53e-02 | 4.38e-04 | 0.5905 | 1.646 | 7.59e-02 | 4.45e-04 | 0.5825 | 1.639 | 7.74e-02 | 4.74e-04 | 0.5336 |
| $\delta_{16}$ | 0.880 | 1.02e-01 | 7.79e-04 | 0.3420 | 0.876 | 1.03e-01 | 8.23e-04 | 0.3142 | 0.868 | 1.05e-01 | 8.53e-04 | 0.3042 |
| $\xi_{00}$ | 0.744 | 2.03e-03 | 1.18e-05 | 0.5901 | 0.743 | 2.04e-03 | 1.20e-05 | 0.5721 | 0.743 | 2.06e-03 | 1.12e-05 | 0.6759 |
| $\xi_{11}$ | 0.514 | 3.54e-03 | 2.48e-05 | 0.4067 | 0.516 | 3.57e-03 | 2.61e-05 | 0.3730 | 0.518 | 3.59e-03 | 2.91e-05 | 0.3035 |

Table 8: Posterior mean and standard deviation obtained for the RY series with the full sample of all duration, together with E-FSCD models. It is based on 50,000 posterior draws recorded after a burn-in period of 15,000 draws. Are reported respectively form left to right three, four and five component normalized duration specification with a cubic B-spline function defined for knots fix each half hour to capture the diurnal pattern. The columns gives posterior mean and standard deviation, and the numerical standard error (NSE) and relative numerical efficiency (RNE) for the posterior mean. The first two row are the log transformation of the mean reversion parameter $\phi$ and the log transformation of the the marginal standard deviation $\sigma$ of the state $x_{d,i}$; the following five are the mixture weights $\beta$ characterizing the normalized duration density; and the last sixteen are the coefficients $\delta$ of the diurnal pattern B-splines basis expansion.

| | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE | Mean | Std | NSE | RNE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\theta_1$ | -5.237 | 2.06e-01 | 3.65e-03 | 0.0637 | -5.419 | 1.98e-01 | 2.81e-03 | 0.0991 | -5.481 | 1.92e-01 | 3.24e-03 | 0.0703 |
| $\theta_2$ | -0.843 | 3.96e-02 | 5.56e-04 | 0.1014 | -0.862 | 4.01e-02 | 4.31e-04 | 0.1731 | -0.875 | 4.12e-02 | 5.26e-04 | 0.1227 |
| $\beta_1$ | 0.559 | 1.24e-02 | 2.21e-04 | 0.0631 | 0.517 | 2.31e-02 | 3.63e-04 | 0.0810 | 0.426 | 1.56e-02 | 3.70e-04 | 0.0355 |
| $\beta_2$ | 0.099 | 1.65e-02 | 2.39e-04 | 0.0949 | 0.065 | 2.39e-02 | 3.78e-04 | 0.0799 | 0.019 | 1.49e-02 | 3.19e-04 | 0.0438 |
| $\beta_3$ | 0.343 | 1.81e-02 | 3.46e-04 | 0.0549 | 0.266 | 2.36e-02 | 3.67e-04 | 0.0828 | 0.305 | 2.36e-02 | 5.56e-04 | 0.0360 |
| $\beta_4$ | | | | | 0.152 | 2.46e-02 | 3.79e-04 | 0.0840 | 0.065 | 2.37e-02 | 5.62e-04 | 0.0356 |
| $\beta_5$ | | | | | | | | | 0.185 | 2.19e-02 | 5.43e-04 | 0.0324 |
| $\delta_1$ | 1.613 | 1.04e-01 | 7.73e-04 | 0.3593 | 1.602 | 1.07e-01 | 7.83e-04 | 0.3718 | 1.593 | 1.08e-01 | 7.77e-04 | 0.3840 |
| $\delta_2$ | 1.896 | 8.40e-02 | 4.66e-04 | 0.6508 | 1.891 | 8.73e-02 | 5.16e-04 | 0.5726 | 1.884 | 8.79e-02 | 4.94e-04 | 0.6320 |
| $\delta_3$ | 2.133 | 7.67e-02 | 4.41e-04 | 0.6042 | 2.128 | 7.89e-02 | 4.43e-04 | 0.6335 | 2.120 | 7.95e-02 | 4.53e-04 | 0.6178 |
| $\delta_4$ | 2.460 | 7.10e-02 | 4.28e-04 | 0.5501 | 2.453 | 7.32e-02 | 4.04e-04 | 0.6570 | 2.445 | 7.34e-02 | 4.46e-04 | 0.5413 |
| $\delta_5$ | 2.695 | 7.25e-02 | 4.89e-04 | 0.4389 | 2.689 | 7.43e-02 | 4.29e-04 | 0.6003 | 2.684 | 7.47e-02 | 5.09e-04 | 0.4314 |
| $\delta_6$ | 2.764 | 7.14e-02 | 4.22e-04 | 0.5718 | 2.760 | 7.42e-02 | 4.65e-04 | 0.5098 | 2.757 | 7.46e-02 | 4.50e-04 | 0.5505 |
| $\delta_7$ | 2.822 | 7.19e-02 | 4.61e-04 | 0.4865 | 2.819 | 7.48e-02 | 4.67e-04 | 0.5137 | 2.812 | 7.49e-02 | 5.06e-04 | 0.4394 |
| $\delta_8$ | 2.952 | 7.33e-02 | 5.31e-04 | 0.3807 | 2.947 | 7.47e-02 | 4.68e-04 | 0.5084 | 2.938 | 7.53e-02 | 4.71e-04 | 0.5115 |
| $\delta_9$ | 3.027 | 7.39e-02 | 4.91e-04 | 0.4537 | 3.022 | 7.57e-02 | 4.99e-04 | 0.4611 | 3.016 | 7.65e-02 | 5.40e-04 | 0.4007 |
| $\delta_{10}$ | 2.905 | 7.22e-02 | 4.47e-04 | 0.5210 | 2.907 | 7.41e-02 | 4.36e-04 | 0.5772 | 2.906 | 7.57e-02 | 4.98e-04 | 0.4610 |
| $\delta_{11}$ | 2.825 | 7.23e-02 | 4.89e-04 | 0.4372 | 2.829 | 7.47e-02 | 4.64e-04 | 0.5187 | 2.824 | 7.49e-02 | 4.78e-04 | 0.4921 |
| $\delta_{12}$ | 2.806 | 7.22e-02 | 4.47e-04 | 0.5216 | 2.806 | 7.35e-02 | 4.40e-04 | 0.5585 | 2.802 | 7.50e-02 | 4.71e-04 | 0.5065 |
| $\delta_{13}$ | 2.814 | 7.33e-02 | 5.03e-04 | 0.4252 | 2.807 | 7.59e-02 | 4.91e-04 | 0.4784 | 2.805 | 7.66e-02 | 5.36e-04 | 0.4078 |
| $\delta_{14}$ | 2.586 | 7.94e-02 | 5.37e-04 | 0.4367 | 2.577 | 8.20e-02 | 5.03e-04 | 0.5317 | 2.572 | 8.18e-02 | 5.48e-04 | 0.4448 |
| $\delta_{15}$ | 2.152 | 8.47e-02 | 5.02e-04 | 0.5707 | 2.146 | 8.76e-02 | 5.21e-04 | 0.5669 | 2.144 | 8.78e-02 | 5.15e-04 | 0.5813 |
| $\delta_{16}$ | 1.425 | 1.19e-01 | 1.01e-03 | 0.2801 | 1.406 | 1.22e-01 | 1.05e-03 | 0.2667 | 1.403 | 1.22e-01 | 1.08e-03 | 0.2555 |
| $\xi_{00}$ | 0.779 | 2.00e-03 | 1.10e-05 | 0.6658 | 0.778 | 2.00e-03 | 1.07e-05 | 0.6945 | 0.778 | 2.01e-03 | 1.11e-05 | 0.6568 |
| $\xi_{11}$ | 0.426 | 4.08e-03 | 2.37e-05 | 0.5917 | 0.427 | 4.09e-03 | 2.48e-05 | 0.5448 | 0.429 | 4.12e-03 | 2.72e-05 | 0.4591 |

Table 9: Posterior mean and standard deviation obtained for the POT series with the full sample of all duration, together with E-FSCD models. It is based on 50,000 posterior draws recorded after a burn-in period of 15,000 draws. Are reported respectively form left to right three, four and five component normalized duration specification with a cubic B-spline function defined for knots fix each half hour to capture the diurnal pattern. The columns gives posterior mean and standard deviation, and the numerical standard error (NSE) and relative numerical efficiency (RNE) for the posterior mean. The first two row are the log transformation of the mean reversion parameter $\phi$ and the log transformation of the the marginal standard deviation $\sigma$ of the state $x_{d,i}$; the following five are the mixture weights $\beta$ characterizing the normalized duration density; and the last sixteen are the coefficients $\delta$ of the diurnal pattern B-splines basis expansion.