Université de Montréal

Essays on Bayesian Analysis of State Space Models with Financial Applications

par

Samuel Gingras

Département de sciences économiques

Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)

en sciences économiques

Mai, 2021

Université de Montréal

Faculté des études supérieures et posdoctorales

Cette thèse intitulée :

Essays on Bayesian Analysis of State Space Models with Financial Applications

présentée par :

Samuel Gingras

a été évaluée par un jury composé des personnes suivantes :

Mathieu Marcoux	Président-rapporteur
William J. McCausland	Directeur de recherche
René Garcia	Membre du jury
Gregor Kastner	Examinateur externe
Louis Doray	Représentant du doyen de la FAS

Thèse acceptée le 27 Mai, 2021

à Claudia, Estelle, et mes parents Louise et Pierre

Remerciements

Acknowledgments

Je tiens tout d'abord à remercier mon directeur de recherche, William J. McCausland, pour ses innombrables conseils, sa grande disponibilité, sa gentillesse ainsi que pour son soutien durant toutes ces années. Je suis très heureux d'avoir choisi son cours, oui je l'avoue un peu par hasard, qui m'a amené à frapper à sa porte. Merci d'avoir si généreusement accepté d'encadrer mon travail et de m'avoir fait découvrir ces thèmes de recherche qui me passionnent tant.

Je remercie par ailleurs Nicolas Klein pour sa patience et sa compréhension. Toute ma gratitude à Vasia Panousi pour le soutien qu'elle a su m'apporté au cours des dernières années de mes études doctorales ainsi que de m'avoir poussé à en apprendre davantage sur l'analyse textuelle.

Je dois également exprimer ma reconnaissance à tous les doctorants en sciences économiques de l'Université de Montréal, en particulier à mes collègues de promotion Fatim, Marlène, Lucienne, Lionel, N'Golo, Idriss et Magnim ainsi qu'à Souleymane et Dominique. Une mention spéciale pour mes anciens collègues de bureau David et Jonathan avec qui j'ai eu grand nombre de discussions des plus enrichissantes. Merci à mes parents Louise et Pierre d'avoir toujours cru en moi et de m'avoir encouragé dans tous mes projets. Par-dessus tout, je serai éternellement reconnaissant envers mon irremplaçable partenaire de vie, Claudia, qui a accepté de venir vivre cette (longue) aventure que sont les études doctorales avec moi. Merci pour ta présence, pour ton soutien, et tout le reste.

Pour terminer, j'aimerais exprimer toute ma gratitude à tous les professeurs et le personnel administratif du département de sciences économiques et du CIREQ. Je n'aurais pu faire de si longues études sans le soutien financier des bourses du département et du CIREQ ainsi que les contrats d'enseignement.

Résumé

Cette thèse est organisée en trois chapitres où sont développées des méthodes de simulation à posteriori pour inférence Bayesienne dans des modèles espace-état ainsi que des modèles économétriques pour l'analyse de données financières.

Au chapitre 1, nous considérons le problème de simulation a posteriori dans les modèles espace-état univariés et non-Gaussiens. Nous proposons une nouvelle méthode de Monte-Carlo par chaînes de Markov (MCMC) mettant à jour le vecteur de paramètres de la dynamique d'état ainsi que la séquence de variables d'état conjointement dans un bloc unique. La proposition MCMC est tirée en deux étapes: la distribution marginale du vecteur de paramètres de la dynamique d'état est construite en utilisant une approximation du gradient et du Hessien du logarithme de sa densité a posteriori, pour laquelle le vecteur de variables d'état a été intégré. La distribution conditionnelle de la séquence de variables d'état, étant donné la proposition du vecteur de paramètres, est telle que décrite dans McCausland (2012). Le calcul du gradient et du Hessien approximatif combine des sous-produits de calcul du tirage d'état avec une quantité modeste de calculs supplémentaires. Nous comparons l'efficacité numérique de notre simulation a posteriori à celle de la méthode Ancillarity-Sufficiency Interweaving Strategy (ASIS) décrite dans Kastner & Frühwirth-Schnatter (2014), en utilisant un modèle de volatilité stochastique Gaussien et le même panel de 23 taux de change quotidiens utilisé dans ce même article. Pour calculer la moyenne a posteriori du paramètre de persistance de la volatilité, notre efficacité numérique est de 6 à 27 fois plus élevée; pour la volatilité du paramètre de volatilité, elle est de 18 à 53 fois plus élevée. Nous analysons dans un second exemple des données de compte de transaction avec un modèle Poisson et Gamma-Poisson dynamique. Malgré la nature non Gaussienne des données de compte, nous obtenons une efficacité numérique élevée, guère inférieure à celle rapportée dans McCausland (2012) pour une méthode d'échantillonnage impliquant un calcul préliminaire de la forme de la distribution a posteriori statique des paramètres.

Au chapitre 2, nous proposons un nouveau modèle de durée conditionnelle stochastique (SCD) pour l'analyse de données de transactions financières en haute fréquence. Nous identifions certaines caractéristiques indésirables des densités de durée conditionnelles paramétriques existantes et proposons une nouvelle famille de densités conditionnelles flexibles pouvant correspondre à une grande variété de distributions avec des fonctions de taux de probabilité modérément variable. Guidés par des considérations théoriques issues de la théorie des files d'attente, nous introduisons des déviations non-paramétriques autour d'une distribution exponentielle centrale, qui, selon nous, est un bon modèle de premier ordre pour les durées financières, en utilisant une densité de Bernstein. La densité résultante est non seulement flexible, dans le sens qu'elle peut s'approcher de n'importe quelle densité continue sur $[0, \infty)$ de manière arbitraire, à condition qu'elle se compose d'un nombre suffisamment grand de termes, mais également susceptible de rétrécissement vers la distribution exponentielle. Grâce aux tirages très efficaces des variables d'état, l'efficacité numérique de notre simulation a posteriori se compare très favorablement à celles obtenues dans les études précédentes. Nous illustrons nos méthodes à l'aide des données de cotation d'actions négociées à la Bourse de Toronto. Nous constatons que les modèles utilisant notre densité conditionnelle avec moins de quatre termes offrent le meilleur ajustement. La variation régulière trouvée dans les fonctions de taux de probabilité, ainsi que la possibilité qu'elle ne soit pas monotone, aurait été impossible à saisir avec une spécification paramétrique couramment utilisée.

Au chapitre 3, nous présentons un nouveau modèle de durée stochastique pour les temps de transaction dans les marchés d'actifs. Nous soutenons que les règles largement acceptées pour l'agrégation de transactions apparemment liées induisent une inférence erronée concernant les durées entre des transactions non liées: alors que deux transactions exécutées au cours de la même seconde sont probablement liées, il est extrêmement improbable que *toutes* paires de transactions le soient, dans un échantillon typique. En plaçant une incertitude sur les transactions liées dans notre modèle, nous améliorons l'inférence pour la distribution de la durée entre les transactions non liées, en particulier près de zéro. Nous proposons un modèle en temps discret pour les temps de transaction censurés permettant des valeurs nulles excessives résultant des durées entre les transactions liées. La distribution discrète des durées entre les transactions indépendantes découle d'une densité flexible susceptible de rétrécissement vers une distribution exponentielle. Dans un exemple empirique, nous constatons que la fonction de taux de probabilité conditionnelle sous-jacente pour des durées (non censurées) entre transactions non liées varie beaucoup moins que celles trouvées dans la plupart des études; une distribution discrète pour les transactions non liées basée sur une distribution expo-

nentielle fournit le meilleur ajustement pour les trois séries analysées. Nous prétendons que c'est parce que nous évitons les artefacts statistiques qui résultent de règles déterministes d'agrégation des échanges et d'une distribution paramétrique inadaptée.

Mots-clés: Modèle espace-état non linéaire non Gaussien; Méthodes Monte-Carlo par chaîne de Markov; Modèle avec variable latente; Fonction de taux de probabilité; Densité de Bernstein; Données de transaction; Durée financière; Données de taux de change

Abstract

This thesis is organized in three chapters which develop posterior simulation methods for Bayesian inference in state space models and econometrics models for the analysis of financial data.

In Chapter 1, we consider the problem of posterior simulation in state space models with non-linear non-Gaussian observables and univariate Gaussian states. We propose a new Markov Chain Monte Carlo (MCMC) method that updates the parameter vector of the state dynamics and the state sequence together as a single block. The MCMC proposal is drawn in two steps: the marginal proposal distribution for the parameter vector is constructed using an approximation of the gradient and Hessian of its log posterior density, with the state vector integrated out. The conditional proposal distribution for the state sequence given the proposal of the parameter vector is the one described in McCausland (2012). Computation of the approximate gradient and Hessian combines computational by-products of the state draw with a modest amount of additional computation. We compare the numerical efficiency of our posterior simulation with that of the Ancillarity-Sufficiency Interweaving Strategy (ASIS) described in Kastner & Frühwirth-Schnatter (2014), using the Gaussian stochastic volatility model and the panel of 23 daily exchange rates from that paper. For computing the posterior mean of the volatility persistence parameter, our numerical efficiency is 6-27 times higher; for the volatility of volatility parameter, 18-53 times higher. We analyse transaction counts in a second example using dynamic Poisson and Gamma-Poisson models. Despite non-Gaussianity of the count data, we obtain high numerical efficiency that is not much lower than that reported in McCausland (2012) for a sampler that involves pre-computing the shape of a static posterior distribution of parameters.

In Chapter 2, we propose a new stochastic conditional duration model (SCD) for the analysis of high-frequency financial transaction data. We identify undesirable features of existing parametric conditional duration densities and propose a new family of flexible conditional densities capable

of matching a wide variety of distributions with moderately varying hazard functions. Guided by theoretical consideration from queuing theory, we introduce nonparametric deviations around a central exponential distribution, which we argue is a sound first-order model for financial durations, using a Bernstein density. The resulting density is not only flexible, in the sense that it can approximate any continuous density on $[0, \infty)$ arbitrarily closely, provided it consists of a large enough number of terms, but also amenable to shrinkage towards the exponential distribution. Thank to highly efficiency draws of state variables, numerical efficiency of our posterior simulation compares very favourably with those obtained in previous studies. We illustrate our methods using quotation data on equities traded on the Toronto Stock Exchange. We find that models with our proposed conditional density having less than four terms provide the best fit. The smooth variation found in the hazard functions, together with the possibility of it being non-monotonic, would have been impossible to capture using commonly used parametric specification.

In Chapter 3, we introduce a new stochastic duration model for transaction times in asset markets. We argue that widely accepted rules for aggregating seemingly related trades mislead inference pertaining to durations between unrelated trades: while any two trades executed in the same second are probably related, it is extremely unlikely that *all* such pairs of trades are, in a typical sample. By placing uncertainty about which trades are related within our model, we improve inference for the distribution of duration between unrelated trades, especially near zero. We propose a discrete model for censored transaction times allowing for zero-inflation resulting from clusters of related trades. The discrete distribution of durations between unrelated trades arises from a flexible density amenable to shrinkage towards an exponential distribution. In an empirical example, we find that the underlying conditional hazard function for (uncensored) durations between unrelated trades varies much less than what most studies find; a discrete distribution for unrelated trades based on an exponential distribution provides a better fit for all three series analyzed. We claim that this is because we avoid statistical artifacts that arise from deterministic trade-aggregation rules and unsuitable parametric distribution.

Keywords: Non-linear non-Gaussian state space model; Markov chain Monte Carlo; Latent variable model; Hazard function; Bernstein density; Transaction data; Financial duration; Exchange rate data

Table des matières

Table of contents

Liste des figures

List of figures

Liste des tableaux

List of tables

Chapter 1

Joint sampling of states and parameters in state space models*

1.1 Introduction

State space models with non-linear non-Gaussian observations and univariate states, including various stochastic volatility and time-varying parameter models, have been applied widely. Bayesian analysis is commonly used, partly because the associated posterior simulation methods do not require evaluation of the likelihood function, which amounts to very high dimensional integration. Still, strong posterior dependence, both within the latent state series and between the latent state series and the parameters that govern its evolution, makes it difficult to simulate the posterior distribution with high numerical efficiency.

We provide here a new posterior simulation method that achieves high numerical efficiency through joint draws of the latent state series and its associated parameters, from their conditional posterior distribution. It is not model specific and it is robust to variation in the conditional posterior distribution as the conditioning information changes during posterior simulation. It extends the method described in McCausland (2012) for drawing the state series as a single block, taking advantage of computational by-products of that method so that there is little additional computation. McCausland (2012) proposed a highly efficient method for simulation smoothing in state space models with univariate Gaussian states and showed how to exploit it for drawing the full sequence of latent states

*This chapter is co-authored with my advisor William J. McCausland.

and the model parameters in a single step. The sampler considered in McCausland (2012) requires precomputing an approximation of the shape of the posterior distribution of the parameters with the states marginalized out. Even though this approach has been shown to work well for various models with low-dimensional parametric spaces, it does not scale particularly well to richer models due to the increasing pre-computation cost. In contrast, our method is based on new approximations of the gradient and Hessian of the likelihood function that does not rely on any pre-computation and only involves a modest amount of additional computation; these approximations may be of interest independently of their application in this paper.

We begin with a basic model, giving the joint distribution of a series $x = (x_1, \ldots, x_n)$ of unobserved latent state variables and a series $y = (y_1, \ldots, y_n)$ of observed dependent variables. The model may stand alone or it may be embedded in an encompassing model; we will see examples of both below. The scalar state variables evolve as a homogenous Gaussian first order Markov process:

$$\begin{aligned} (x_1 - \mu) &\sim \mathcal{N}(0, \sigma^2(1-\phi^2)^{-1}), \\ (x_t - \mu) &\sim \mathcal{N}(\phi(x_{t-1} - \mu), \sigma^2), \quad t = 2, \ldots, n. \end{aligned} \tag{1.1}$$

The dependent variable y_t may depend on its own lagged values but only on the contemporaneous value x_t of the latent state sequence. Thus, we can decompose $p(y|x)$ as

$$p(y|x) = \prod_{t=1}^{n} p(y_t|x_t, y_{1:t-1}), \tag{1.2}$$

where $y_{1:t-1} = (y_1, \ldots, y_{t-1})$. The conditional distribution of y_t given x_t and $y_{1:t-1}$ is in principle very flexible. The y_t may be scalars or vectors, as in factor models, and the dimension of y_t may vary with t, to accommodate mixed-frequency observations or missing data. Each element of y_t, considered separately, may be discrete, continuous or mixed. The $p(y_t|x_t, y_{1:t-1})$ may be functions of parameters other than σ, ϕ and μ. The only requirement is than the distribution of y_t depends on the contemporaneous value x_t and not on any other values of the latent state sequence.

The unknown parameters of the state dynamics are σ, ϕ and μ, although sometimes, due to normalization, μ does not appear. We organize these parameters in the parameter vector $\theta \equiv (\ln \omega, \tanh^{-1} \phi, \mu)$, where $\omega = 1/\sigma^2$ and $\tanh^{-1}(\cdot)$ is the inverse hyperbolic tangent. We also define the subvector $\vartheta \equiv (\ln \omega, \tanh^{-1} \phi)$. The reasons for these transformations will be made clear below.

The most common examples of these models are variations on the stochastic volatility (SV) model

introduced by Taylor (1982). Some of the variations add flexibility, typically by allowing excess kurtosis in the measurement equation or jumps. Often, SV models are embedded in more complicated models, such as multivariate stochastic volatility models. Other examples include models with time-varying counts or durations. Durbin & Koopman (1997) study counts of deaths of van drivers in Britain; Frühwirth-Schnatter & Wagner (2006), casualties of pedestrians in Linz; and Jung et al. (2006), admissions for asthma to a hospital in Sydney. Bauwens & Veredas (2004), Strickland et al. (2006) and Strickland et al. (2008) study durations between transactions in financial markets.

Several methods for Bayesian posterior simulation in such state state models have been proposed. Direct methods sample latent states from their conditional posterior distribution. Sampling may be done one-at-a-time as in Jacquier et al. (1994); in blocks, as in Shephard & Pitt (1997), Watanabe & Omori (2004), Strickland et al. (2006) or Omori & Watanabe (2008); or all at once, as in McCausland (2012) or Djegnene & McCausland (2015). Auxiliary mixture methods involve transforming the model into a linear Gaussian model, approximating any non-Gaussian distributions in the transformed model by finite Gaussian mixtures. Kim et al. (1998), Chib et al. (2002) and Omori et al. (2007) use auxiliary mixture sampling for various SV models. Stroud et al. (2003) use it for Gaussian, but non-linear, state space models with state dependent variances; Frühwirth-Schnatter & Wagner (2006) for state space models with Poisson counts; and Frühwirth-Schnatter & Frühwirth (2007) for logit and multinomial logit models.

Numerical efficiency varies greatly across posterior simulation methods. Since there is often strong posterior dependence both within x and between x and θ, it promotes efficiency to update x in a single Gibbs block and especially to update θ and x together. However, the larger the block, the more difficult it is to approximate its conditional posterior distribution, which is not a standard distribution. So, for example, a multivariate normal distribution is adequate for direct sampling of blocks of 20-50 state values, but not for the complete observed sequence x. McCausland (2012) and Djegnene & McCausland (2015) provide a non-Gaussian approximation of $p(x|\theta, y)$ for generic state space models that proved highly efficient for drawing x in a single block. Auxiliary mixture models yield conditionally Gaussian x when one conditions on discrete mixture component indicators, and x can be drawn as a single block here too.

There have been previous attempts to draw θ and x together in a single block. For the Taylor SV model, Kim et al. (1998) draw θ and x together, conditional on mixture component indicators, in what they call an "integration sampler" because x is marginalized out to draw parameters. Chib et al. (2002) analyse several SV models using such a sampler. McCausland (2012) draws parameters

and x together directly, as a single block, in many state space models. One application replicates the analysis in Chib et al. (2002) of a 6-parameter Student's t SV model, the model from that paper with the highest marginal likelihood for a sample of size $n = 8851$ of S&P 500 stock index returns. For the posterior sample mean of ϕ, Chib et al. (2002) achieve a numerical efficiency of 0.19 and McCausland (2012), 0.61. For the posterior sample mean of σ, the efficiencies are 0.10 and 0.87, respectively.

In all these cases, the SV model is a self-contained model for a single return series, not embedded in a larger model, and thus the posterior distribution of the parameters, with x marginalized out, does not change from draw to draw. One can precompute a close approximation of this target distribution, ensuring that it is thick-tailed relative to the target, and use the approximation as a proposal distribution.

When the SV model is embedded in a larger model, the shape of this posterior distribution is a moving target; each time we draw θ and x, the conditioning information, consisting of values of other unknown quantities in the encompassing model, is different. In part because of this issue, θ and x—and mixture component indicators, if any—are often updated in separate Gibbs blocks. Kastner & Frühwirth-Schnatter (2014) do this quite efficiently by interleaving draws of two different parameterizations of θ. In simulations not reported here, we found that the numerical efficiency they achieve is (slightly) better than the efficiency obtained using pure Gibbs draws of $\theta|x, y$ and $x|\theta, y$, despite the fact that they augment the state space to include mixture indicators. For this reason, we use their results as the benchmark against which we compare the efficiency results we obtain for our joint draws of θ and x.

Our contribution is a method to draw parameters and states together, but in a way that does not rely on any pre-computation. Instead, it is based on computations of the local shape of the posterior distribution of parameters, with x marginalized out. The shape of this target distribution can change from draw to draw. The result is a sampler that is much more numerically efficient than that of Kastner & Frühwirth-Schnatter (2014) and not much less efficient than that of McCausland (2012), for simple models where the posterior distribution of parameters is static and for which a close approximation can be pre-computed.

The simulation methods we provide in this paper can be applied to more general models which embed the basic model, in the sense that the basic model is a special case obtained by conditioning on some of the random elements of the larger model. The modular nature of Gibbs sampling for

posterior simulation exploits this kind of structure, and for the purposes of updating the conditional distribution of θ and x, we can ignore whether the model stands alone or is embedded in a larger model.

As an example, we can add to the model a series $z = (z_1, \ldots, z_n)$ of exogenous (specifically, predetermined) variables, provided that we make the conditional dependence assumptions implied by the factorization

$$p(x, y|z) = p(x_1|z_1)p(y_1|x_1, z_1)\prod_{t=2}^{n} p(x_t|x_{t-1}, z_{1:t})p(y_t|x_t, y_{1:t-1}, z_{1:t}), \tag{1.3}$$

with state dynamics generalized to

$$\begin{aligned} (x_1 - \eta_1(z_1))|z_1 &\sim \mathcal{N}(0, \sigma^2(1-\phi^2)^{-1}), \\ (x_t - \eta_t(z_{1:t}))|z_{1:t}, x_{t-1} &\sim \mathcal{N}(\phi(x_{t-1} - \eta_{t-1}(z_{1:t-1})), \sigma^2), \quad t = 2, \ldots, n, \end{aligned} \tag{1.4}$$

where the $\eta_t(z_{1:t})$, $t = 1, \ldots, n$, are functions that may depend on other parameters, but not σ or ϕ. Let $e = (x_1 - \eta_1(z_1), x_2 - \eta_2(z_{1:2}), \ldots, x_n - \eta_n(z_{1:n-1}))$. When we condition on z, the joint distribution of (e, y) is the same as the distribution of (x, y) in the basic model with $\mu = 0$. For instance, in the analysis of high-frequency data, it is well known that trading intensity varies systematically during the day; including time-of-day information in η_t allows one to capture this variation. We consider one such example below, where we analyse a time series consisting of counts of trades of IBM stock in 5 minute intervals. In this case, the variable y_t is the transaction count, the state variable x_t is a measure of trading intensity, and η_t linearly aggregates evaluations, at the time of day of the t'th observation, of a set of basis functions.

We will use local shape information to draw proposals of θ that target the distribution $\theta|y$, with x integrated out. Many methods that use local shape information to simulate a given target distribution use Metropolis-Hastings updates where the proposal distributions are, or combine, discretizations of stochastic differential equations whose stationary distribution is the target distribution. These include the Metropolis-adjusted Langevin Algorithm (MALA, Roberts & Rosenthal 1998, Roberts & Stramer 2003) and Hamiltonian Monte Carlo (HMC, Duane et al. 1987). MALA and HMC are well suited to high dimensional problems where small steps are all one can hope for and the gradient of the target distribution is cheap to evaluate. A single proposal in HMC, for example, might consist of many tens or hundreds of small steps, followed by an accept/reject of the final value. In our application, the dimension is much smaller and evaluation of the gradient is inexact and expensive.

For this reason, we construct a proposal based on a quadratic Taylor series approximation of the log target density, with some adjustments to accommodate changing curvature and to take advantage of special features of our target distribution.

We describe our new method in Section 1.2, with lengthly derivations relegated to the appendices. We demonstrate it in Section 1.3, comparing its numerical efficiency with those of competing methods. We conclude in Section 1.4.

1.2 A joint update of states and parameters

We describe here a new method for updating the posterior distribution $\theta, x|y$, useful for MCMC posterior simulation. It takes advantage of close approximations of the gradient $g_{\theta|y}(\theta)$ and Hessian $H_{\theta|y}(\theta)$ of the log posterior density $\ln p(\theta|y)$, where x has been marginalized out; $g_{\theta|y}(\theta)$ and $H_{\theta|y}(\theta)$ give a good idea of the shape of $p(\theta|y)$ at any point θ, allowing us to draw proposals that are adapted to the target distribution.

We will first motivate the parameterization $\theta = (\ln \omega, \tanh^{-1} \phi, \mu)$ of the dynamics of the latent state process x. We then describe how to compute approximations of $g_{\theta|y}(\theta)$ and $H_{\theta|y}(\theta)$. We then describe a three-step proposal of a candidate value (θ^*, x^*) given the current state (θ, x), and provide the Metropolis-Hastings acceptance probability for the joint draw (θ^*, x^*).

We will use the parameterization $\theta = (\ln \omega, \tanh^{-1}(\phi), \mu)$ of the latent state process dynamics, where $\omega = \sigma^{-2}$, the precision of the state innovation. The precision ω is more convenient to work with than σ^2 because it appears linearly in the term of $\ln p(x|\theta)$ that is quadratic in x. Taking the logarithm of ω and the inverse hyperbolic tangent of ϕ yields a parameter vector with the unrestricted support $\mathbb{R}^3$, and the log posterior density $\ln p(\theta|x)$ is better approximated by a quadratic function than is the log posterior density $\ln p(\omega, \phi, \mu|x)$, especially when ϕ has considerable posterior mass near the boundary of its support. We use the parametrization θ for posterior simulation, transforming θ back to $(\omega, \phi, \mu) = (e^{\theta_1}, \tanh \theta_2, \theta_3)$ when necessary.

We can write $g_{\theta|y}(\theta)$ and $H_{\theta|y}(\theta)$, the gradient and Hessian of $\ln p(\theta|y)$, as

$$g_{\theta|y}(\theta) = g_\theta(\theta) + g_{y|\theta}(\theta) \quad \text{and} \quad H_{\theta|y}(\theta) = H_\theta(\theta) + H_{y|\theta}(\theta),$$

where $g_\theta(\theta)$ and $H_\theta(\theta)$ are the gradient and Hessian of the log prior density $\ln p(\theta)$; and $g_{y|\theta}(\theta)$ and

$H_{y|\theta}(\theta)$ are the gradient and Hessian of $\ln p(y|\theta)$ with respect to θ. Typically, $g_\theta(\theta)$ and $H_\theta(\theta)$ are easily computed, although it will be necessary to compute derivatives of a log Jacobian term if the prior is not specified directly in terms of θ. Since $g_{y|\theta}(\theta)$ and $H_{y|\theta}(\theta)$ are not available in closed form, we compute approximations $\tilde{g}_{y|\theta}(\theta)$ and $\tilde{H}_{y|\theta}(\theta)$, described in Appendix A.1, and define

$$\tilde{g}_{\theta|y}(\theta) \equiv g_\theta(\theta) + \tilde{g}_{y|\theta}(\theta) \quad \text{and} \quad \tilde{H}_{\theta|y}(\theta) \equiv H_\theta(\theta) + \tilde{H}_{y|\theta}(\theta). \tag{1.5}$$

Our Metropolis-Hastings update of (θ, x) consists of a proposal of $(\theta^*, x^*) = (\vartheta^*, \mu^*, x^*)$ from a three-step transition kernel $q(\theta^*, x^*|\theta, x) = q(\vartheta^*|\vartheta)q(\mu^*|\vartheta^*, \vartheta)q(x^*|\theta^*)$, which is accepted with the usual Metropolis-Hastings acceptance probability which, here, is

$$\min\left[1, \frac{p(\theta^*)p(x^*|\theta^*)p(y|\theta^*, x^*)}{p(\theta)p(x|\theta)p(y|\theta, x)} \cdot \frac{q(\vartheta|\vartheta^*)q(\mu|\vartheta, \vartheta^*)q(x|\theta)}{q(\vartheta^*|\vartheta)q(\mu^*|\vartheta^*, \vartheta))q(x^*|\theta^*)}\right].$$

The three steps of the proposal are outlined in the next three sections.

1.2.1 The proposal density $q(\vartheta^*|\vartheta)$

The transition kernel $q(\vartheta^*|\vartheta)$ is based on the approximations $\tilde{g}_{\theta|y}(\theta)$ and $\tilde{H}_{\theta|y}(\theta)$ in (1.5) of the gradient and Hessian of the log posterior density $\log p(\theta|y)$, evaluated at the current value of θ. We use $\tilde{g}_{\theta|y}(\theta)$ and $\tilde{H}_{\theta|y}(\theta)$, together with the gradient $g_\theta(\theta)$ and Hessian $H_\theta(\theta)$ of the log prior density to construct a Gaussian distribution $\mathcal{N}(\bar{\vartheta}, \Sigma_\vartheta)$ with mean $\bar{\vartheta}$ equal to a guess of the conditional mode of ϑ given y, and variance Σ_ϑ a little larger than what we expect $\mathrm{Var}[\vartheta|y]$ to be. To determine $\bar{\vartheta}$ and Σ_ϑ, we first compute an approximation $\tilde{\vartheta}$ to the maximum likelihood value of ϑ and an approximation $\tilde{H}$ to the Hessian of the conditional log likelihood function there, as described in Appendix A.2. We then set $\bar{\vartheta} = \tilde{\vartheta} - (H_\vartheta(\vartheta) + \tilde{H})^{-1}g_\vartheta(\tilde{\vartheta})$ to adjust for the prior distribution and $\Sigma_\vartheta = -\lambda(H_\vartheta(\vartheta) + \tilde{H})^{-1}$ where $\lambda > 1$ to make sure the variance of our proposal distribution is larger than the target distribution. Then $q(\vartheta^*|\vartheta)$ is the Gaussian kernel with stationary distribution $\mathcal{N}(\bar{\vartheta}, \Sigma_\vartheta)$ and autocorrelation matrix Ψ chosen by the user to trade off step size against acceptance probability. That is, the proposal is $\vartheta^* \sim N((I-\Phi)\bar{\vartheta}+\Phi\vartheta, \Sigma_\vartheta - L\Psi\Psi^\top L^\top)$, where the autoregressive coefficient matrix is $\Phi = L\Psi L^\top$ and L is the lower Cholesky factor of Σ_ϑ. In our applications we use $\lambda = 1.2$ and $\Psi = 0.2I_2$, based on some experimentation.

1.2.2 The proposal density $q(\mu^*|\mu, \vartheta, \vartheta^*)$

The second step is a proposal of μ^*, also based on the approximations $\tilde{g}_{\theta|y}(\theta)$ and $\tilde{H}_{\theta|y}(\theta)$, but using the value of ϑ^* to obtain a better approximation of $H_{\theta|y}((\vartheta^{*\top}, \mu)^\top)$ than $\tilde{H}_{\theta|y}(\theta)$ is. Consider the following Gaussian approximation of the conditional distribution $\mu|\vartheta^*, y$, based on the trivariate Gaussian distribution whose log density at θ has gradient $\tilde{g}_{\theta|y}(\theta)$ and Hessian $\tilde{H}_{\theta|y}(\theta)$:

$$N(\mu - H_{\mu\mu}^{-1}(g_\mu + H_{\mu\vartheta}(\vartheta^* - \vartheta)), -H_{\mu\mu}^{-1}), \tag{1.6}$$

where g_μ, $H_{\mu\mu}$ and $H_{\vartheta\mu}$ are defined by the following partition of $\tilde{g}_{\theta|y}(\theta)$ and $\tilde{H}_{\theta|y}(\theta)$, where g_μ and $H_{\mu\mu}$ are scalar:

$$\tilde{g}_{\theta|y}(\theta) = \begin{bmatrix} g_\vartheta \\ g_\mu \end{bmatrix} \quad \tilde{H}_{\theta|y}(\theta) = \begin{bmatrix} H_{\vartheta\vartheta} & H_{\vartheta\mu} \\ H_{\mu\vartheta} & H_{\mu\mu} \end{bmatrix}.$$

We exploit the fact that $H_{\mu\mu}$ and $H_{\mu\vartheta}$ are approximately proportional to $\omega(1-\phi)^2$, based on the observation (see Appendix A.1) that for these elements of $\tilde{H}_{y|\theta}(\theta)$, the first term of (A.2) dominates the second, and that $E[e|\theta, y]$ is fairly insensitive to the values of ω and ϕ. We compute better local approximations $H_{\mu\mu}^*$ and $H_{\mu\mu}^{\text{mid}}$ of element $H_{\mu\mu}$ and a better local approximation $H_{\mu\vartheta}^{\text{mid}}$ of $H_{\mu\vartheta}$ as follows:

$$H_{\mu\mu}^* \equiv \frac{\omega^*(1-\phi^*)^2}{\omega(1-\phi)^2} H_{\mu\mu}$$

$$H_{\mu\mu}^{\text{mid}} \equiv \frac{\omega_{\text{mid}}(1-\phi_{\text{mid}})^2}{\omega(1-\phi)^2} H_{\mu\mu}, \qquad H_{\mu\vartheta}^{\text{mid}} \equiv \frac{\omega_{\text{mid}}(1-\phi_{\text{mid}})^2}{\omega(1-\phi)^2} H_{\mu\vartheta},$$

where

$$\omega_{\text{mid}} = \exp((\vartheta_1 + \vartheta_1^*)/2), \quad \phi_{\text{mid}} = \tanh((\vartheta_2 + \vartheta_2^*)/2)$$

Then the proposal density $q(\mu^*|\vartheta, \vartheta^*)$ for μ^* is the following modification of (1.6), based on the adjustments to $H_{\mu\mu}$ and $H_{\mu\vartheta}$:

$$\mu \sim N(\mu - (H_{\mu\mu}^{\text{mid}})^{-1}(g_\mu + H_{\mu\vartheta}^{\text{mid}}(\vartheta^* - \vartheta)), -(H_{\mu\mu}^*)^{-1}).$$

1.2.3 The proposal density $q(x^*|\theta^*, y)$

Given the proposal θ^*, we use the HESSIAN method to draw x^* from a close approximation $q(x^*|\theta^*, y)$ to the conditional posterior distribution $p(x^*|\theta^*, y)$ and to compute the proposal density

$q(x^*|\theta^*, y)$. Using a modest amount of additional computation, described in the appendices, we also compute $\tilde{g}_{y|\theta}(\theta^*)$ and $\tilde{H}_{y|\theta}(\theta^*)$, which we then use to evaluate the reverse proposal density $q(\theta|\theta^*)$. If the joint proposal (θ^*, x^*) is accepted, the values of $p(\theta^*)$, $\tilde{g}_{y|\theta}(\theta^*)$ and $\tilde{H}_{y|\theta}(\theta^*)$ can be kept for the next iteration.

1.3 Results

1.3.1 Exchange rates

We apply our methods to daily exchange rate data for 23 currencies, against the Euro. For each series, we observe 3140 consecutive trading days, from January 3rd, 2000 to April 4, 2012, and compute 3139 log returns. The data, from the European Central Bank, are those used by Kastner & Frühwirth-Schnatter (2014) in their empirical application.

As in Kastner & Frühwirth-Schnatter (2014), we analyse the de-meaned log returns. We use the Gaussian stochastic volatility model, without leverage, given by

$$y_t|x_t \sim \mathcal{N}\left(0, \exp(x_t)\right) \tag{1.7}$$

$$x_t|x_{t-1}, \mu, \phi, \sigma \sim \mathcal{N}\left(\mu + \phi\left(x_{t-1} - \mu\right), \sigma^2\right) \tag{1.8}$$

$$x_1|\mu, \phi, \sigma \sim \mathcal{N}\left(\mu, \sigma^2/\left(1-\phi^2\right)\right), \tag{1.9}$$

and select the following prior for $\theta = (\ln\omega, \tanh^{-1}(\phi), \mu)$:

$$\theta \sim N\left(\begin{bmatrix} 3.6 \\ 2.5 \\ -10.5 \end{bmatrix}, \begin{bmatrix} 1.25 & 0.5 & 0 \\ 0.5 & 0.25 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}\right),$$

where $\omega = \sigma^{-2}$. The prior is based on independent priors for $\tanh^{-1}(\phi)$ and $\ln\omega(1-\phi^2)$. We do this because in practice, the unconditional precision $\omega(1-\phi^2)$ of x_t covaries less with ϕ than does the conditional precision ω, across financial return series. Results (including efficiency) are fairly robust to setting the covariance $\mathrm{Cov}[\theta_1, \theta_2]$ to zero.

Table 1.1 illustrates the results, based on 45,000 posterior draws recored after a burn-in period of 5,000 draws. For each currency, and the three parameters σ, ϕ and μ, we report the posterior

Table 1.1: Posterior mean, standard deviation and numerical efficiency for the Gaussian stochastic volatility model and the European Central Bank exchange rates data.

Currency	$E[\sigma\|y]$	sd$[\sigma\|y]$	RNE	$\times$	$E[\phi\|y]$	sd$[\phi\|y]$	RNE	$\times$	$E[\mu\|y]$	sd$[\mu\|y]$	RNE
Australian dollar	0.155	0.021	0.42	40.3	0.981	0.006	0.28	19.1	-10.25	0.16	0.46
Canadian dollar	0.076	0.015	0.23	27.1	0.993	0.004	0.15	13.2	-10.11	0.26	0.36
Swiss franc	0.202	0.019	0.74	54.1	0.986	0.004	0.46	15.1	-12.00	0.27	0.49
Czech koruna	0.260	0.032	0.45	43.3	0.960	0.010	0.35	25.0	-11.50	0.12	0.59
Danish krone	0.409	0.040	0.24	17.1	0.912	0.016	0.25	14.4	-18.07	0.09	0.41
UK pound sterling	0.098	0.012	0.51	44.4	0.993	0.003	0.24	9.5	-10.84	0.28	0.47
Hong Kong dollar	0.064	0.010	0.40	30.3	0.996	0.002	0.22	8.0	-10.16	0.32	0.41
Indonesian rupiah	0.209	0.032	0.33	46.3	0.974	0.009	0.25	28.0	-9.86	0.15	0.43
Japanese yen	0.114	0.015	0.50	45.8	0.991	0.003	0.29	13.8	-9.95	0.26	0.44
Korean won	0.135	0.016	0.60	47.6	0.989	0.004	0.31	12.4	-10.03	0.24	0.45
Mexican peso	0.153	0.020	0.52	45.0	0.982	0.006	0.26	15.5	-9.76	0.17	0.50
Malaysian ringgit	0.074	0.012	0.33	29.6	0.994	0.003	0.17	9.7	-10.29	0.27	0.43
Norwegian krone	0.165	0.021	0.37	27.8	0.976	0.007	0.20	10.8	-11.14	0.14	0.47
New Zealand dollar	0.155	0.027	0.23	30.5	0.974	0.010	0.17	19.5	-10.01	0.12	0.48
Philippine peso	0.133	0.021	0.26	42.0	0.983	0.007	0.17	21.4	-10.11	0.16	0.41
Polish zloty	0.181	0.020	0.42	29.3	0.979	0.006	0.27	11.4	-10.42	0.17	0.44
Romanian leu	0.299	0.025	0.63	37.8	0.972	0.006	0.49	15.8	-11.08	0.20	0.43
Russian rouble	0.143	0.016	0.61	50.2	0.990	0.003	0.32	12.3	-10.61	0.27	0.43
Swedish krona	0.109	0.011	0.59	35.7	0.992	0.002	0.32	7.3	-11.33	0.28	0.44
Singapore dollar	0.066	0.010	0.45	44.5	0.996	0.002	0.25	11.8	-10.58	0.36	0.41
Thai baht	0.115	0.018	0.28	25.2	0.987	0.005	0.18	11.5	-10.17	0.18	0.45
Turkish lira	0.302	0.025	0.55	38.1	0.962	0.008	0.41	17.4	-9.78	0.15	0.52
US dollar	0.064	0.009	0.46	33.8	0.996	0.002	0.23	8.4	-10.13	0.33	0.39

Note. This table gives the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the mean, based on 45,000 posterior draws recorded after a burn-in period of 5,000 draws. The columns "$\times$" give the ratio of the numerical efficiency obtained with our method and the ones reported in Kastner & Frühwirth-Schnatter (2014) for the same data.

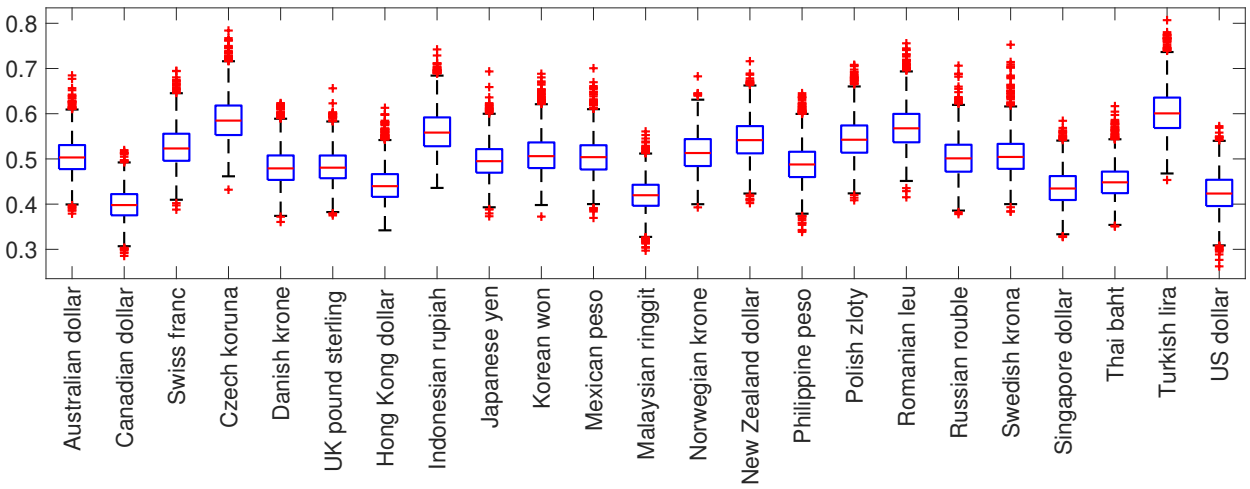

Figure 1.1: Relative numerical efficiency for the posterior mean of the state variables in a Gaussian stochastic volatility model without leverage apply to the European Central Bank daily exchange rate data. The box-plots report the relative numerical efficiency of the 3139 state variables computed for each individual currency against the Euro.

sample mean and standard deviation, as well as the relative numerical efficiency for the posterior sample mean. Figure 1.1 shows box-plots of the relative numerical efficiency for the state variables obtained for each currency. The relative numerical efficiency is the ratio of the numerical variance of the mean of an iid sample to the numerical variance of the posterior sample mean. It is the reciprocal of the inefficiency factor used by Kastner & Frühwirth-Schnatter (2014) and others. We estimate the numerical variance of our posterior sample means using the overlapping batch means method—see Flegal & Jones (2010). Simulations were performed in Matlab R2019a running on a MacBook Pro with a 2.9 GHz Quad-Core Intel Core i7 processor running OS X 10.16.7. The HESSIAN method and the approximations of the gradient and the Hessian of the log posterior density are coded in C and integrated to the Matlab interface via a MEX function. The average CPU time to perform 1000 iterations is about 3.25 seconds.

For σ and ϕ parameters, we also report the number of times more efficient the posterior sample means are, compared to those reported by Kastner & Frühwirth-Schnatter (2014). For computing the posterior mean of σ, our numerical efficiency is 18-53 times higher; for ϕ, 6-27 times higher. The relation between the numerical efficiency and the average runtime implies that our method is able to generate an effective sample of 1000 draws from $p(\sigma|y)$ in between 4.4 and 14.1 seconds and from $p(\phi|y)$ in between 6.6 and 21.7 seconds, depending on the currency.

For the μ parameter, numerical efficiency is comparable to that obtained by Kastner & Frühwirth-Schnatter (2014). The inefficiency factors reported by Kastner & Frühwirth-Schnatter (2014) for μ

are rounded to the nearest integer, one of 1, 2, 3 or 4. This makes it difficult to compare efficiencies for individual currencies. When we take the reciprocal of their average rounded inefficiency factor, we obtain 0.40, which is almost the same as the average efficiency of 0.43 across currencies that we estimate for our method. Given that the conditional posterior distribution of μ, (i.e. $\mu|\phi, \sigma, x, y$) is Gaussian, the numerical efficiency for μ and higher moments could easily be greatly improved through antithetic sampling or Rao-Blackwellization, for either method.

1.3.2 High-frequency counts with diurnal patterns

The next example illustrates the use of our method for dynamic count data, with an application to stock-market trading activity measured by the number of trades aggregated over time intervals with a fixed length. We use a data set of transactions of IBM stock from November 1, 1990 to January 31, 1991. Details are in Chapter 5 in Tsay (2002), and the raw data are kindly provided by Ruey Tsay at the website for his book.[1] Transactions made on November 23 and December 27 were removed because of market closings (see Engle & Russell 1998), leaving 61 trading days in the sample. We consider for analysis the number of trades in 5 minutes intervals between 9:30 am and 4:00 pm, giving 78 observations of transaction counts per day and a total of 4758 observations. The data are considerably non-Gaussian; the mean count is 12.41 and the numbers of intervals with 0, 1 and 2 transactions are 46, 75, and 107, respectively. They also show evidence of overdispersion; the sample standard deviation is 9.93, which gives a sample coefficient of variation of 1.25.

Dynamic count models have been successfully used in many applications where count intensity varies over time; examples using high-frequency financial data are Rydberg & Shephard (2003) and Liesenfeld et al. (2006) that analysed the absolute value of asset price changes as a multiple of the tick size. In the case of transaction data, it is well known that intra-day trading activity has a distinctive U-shape pattern (see, e.g., Tsay 2002). To capture this feature, we use time-of-day indexes as exogenous variables and specify the marginal mean μ_t of the state variable x_t via a flexible Fourier series approximation (Gallant 1981):

$$\mu_t = \mu + \alpha_0 \cdot \tau_t + \sum_{q=1}^{3} \alpha_{q,1} \cos(\tau_t \cdot 2q\pi) + \alpha_{q,2} \sin(\tau_t \cdot 2q\pi), \tag{1.10}$$

where $\tau_t \in [0, 1]$ is the normalized intra-day index of the t'th observation computed as the number

[1] https://faculty.chicagobooth.edu/ruey.tsay/teaching/fts/

of intervals between the opening of the market and the t'th observation, divided by the number of intervals in a day.

Define $z_t \equiv (1, \tau_t, \cos(\tau_t \cdot 2\pi), \sin(\tau_t \cdot 2\pi), \ldots, \cos(\tau_t \cdot 6\pi), \sin(\tau_t \cdot 6\pi))$ and $\beta \equiv (\mu, \alpha)$, so that $\mu_t = z_t\beta$. A simple parameter-driven model for transaction counts is given by

$$y_t | x_t \sim \text{Poisson}\left(\exp(x_t)\right) \tag{1.11}$$

$$x_t | x_{t-1}, \beta, \phi, \sigma \sim \mathcal{N}\left(z_t\beta + \phi(x_{t-1} - z_{t-1}\beta), \sigma^2\right) \tag{1.12}$$

$$x_1 | \beta, \phi, \sigma \sim \mathcal{N}\left(z_1\beta, \sigma^2/(1-\phi^2)\right). \tag{1.13}$$

In addition to the Poisson count model, we also consider the Gamma-Poisson count model given by the above state dynamics and

$$y_t | x_t, r \sim \text{GammaPoisson}(r, \exp(x_t)), \tag{1.14}$$

where $r > 0$ is a shape parameter. For both models, $\exp(x_t)$ gives the conditional mean of y_t. The Gamma-Poisson count model features overdispersion of the conditional count distribution relative to the Poisson count model. The dynamic models defined by the previous equations are typical examples of specifications using exogenous variables that satisfy the conditional dependence assumption implied by (1.3) and with states following the generalized dynamics given by (1.4).

We use a prior distribution analogous to the one used by McCausland (2012) for the same count data. The prior distribution for the state parameters $\theta = (\ln \omega, \tanh^{-1}(\phi))$ is more diffuse than for the exchange rate example:

$$\theta \sim \mathcal{N}\left(\begin{bmatrix} 3.0 \\ 1.5 \end{bmatrix}, \begin{bmatrix} 2.0 & 0.5 \\ 0.5 & 0.625 \end{bmatrix}\right),$$

where $\omega = \sigma^{-2}$. The parameters capturing the diurnal pattern are a *priori* independent and Gaussian: $\mu \sim \mathcal{N}(0.0, 25.0)$, $\alpha_0 \sim \mathcal{N}(0, 1)$ and $\alpha_{q,k} \sim \mathcal{N}(0, 0.25)$ for $q = 1, 2, 3$ and $k = 1, 2$. For the Gamma-Poisson model, the shape parameter r is also a *priori* independent of the other parameters with $\ln r \sim \mathcal{N}(2.5, 1)$.

The upper panel of Table 1.2 illustrates the results. 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. The posterior samples consist of 25,000 retained draws recorded after a burn-in period of 5,000 draws. Posterior simulation is by Metropolis-within-Gibbs,

Table 1.2: Posterior mean and standard deviation for the dynamic Poisson and Gamma-Poisson count model with and without diurnal pattern.

Panel A: Models with diurnal pattern

	Poisson				Gamma-Poisson			
	Mean	Std	NSE	RNE	Mean	Std	NSE	RNE
μ	2.6188	0.0409	0.00036	0.517	2.6715	0.0517	0.00075	0.189
ϕ	0.7640	0.0129	0.00008	0.968	0.9406	0.0087	0.00010	0.297
σ	0.3632	0.0076	0.00004	1.215	0.1645	0.0113	0.00015	0.236
r					10.5190	0.7059	0.01068	0.175
α_0	-0.5656	0.0676	0.00063	0.456	-0.5830	0.0605	0.00139	0.076
α_{11}	0.2709	0.0310	0.00018	1.132	0.2693	0.0352	0.00026	0.707
α_{12}	-0.0743	0.0375	0.00028	0.698	-0.0794	0.0407	0.00050	0.263
α_{21}	0.0680	0.0279	0.00019	0.898	0.0708	0.0223	0.00018	0.621
α_{22}	-0.0861	0.0299	0.00020	0.872	-0.0883	0.0242	0.00028	0.306
α_{31}	0.0670	0.0244	0.00016	0.951	0.0696	0.0167	0.00014	0.536
α_{32}	0.0092	0.0254	0.00018	0.839	0.0060	0.0179	0.00020	0.314

Panel B: Models without diurnal pattern

	Poisson				Gamma-Poisson			
	Mean	Std	NSE	RNE	Mean	Std	NSE	RNE
μ	2.3323	0.0268	0.00019	0.791	2.3725	0.0423	0.00033	0.664
ϕ	0.7931	0.0116	0.00008	0.819	0.9297	0.0086	0.00009	0.346
σ	0.3708	0.0076	0.00005	0.830	0.2002	0.0107	0.00013	0.265
r					11.5652	0.8178	0.01190	0.189

Note. The table gives posterior mean and standard deviation, and the numerical standard error (NSE) and relative numerical efficiency (RNE) for the posterior mean, based on 25,000 posterior draws recorded after a burn-in period of 5,000 draws.

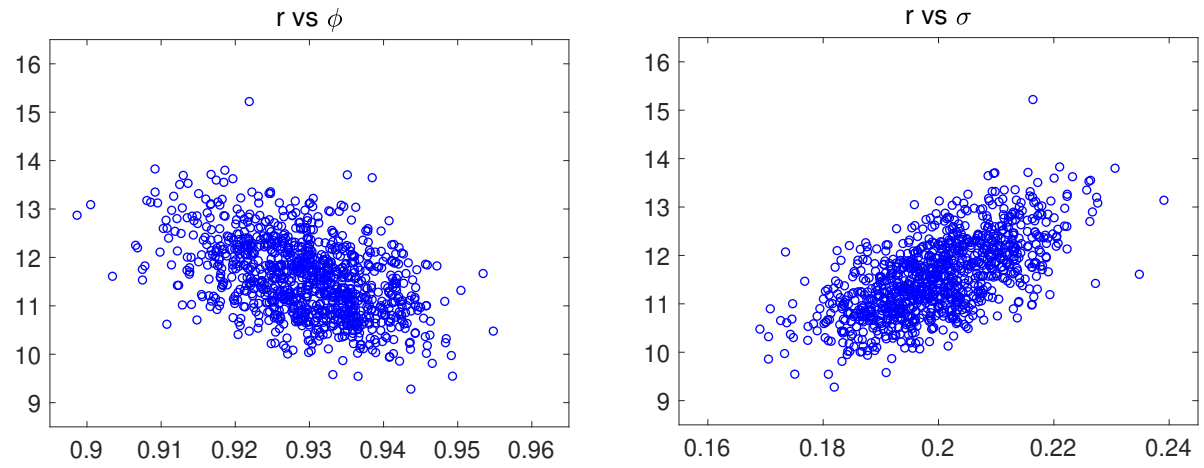

Figure 1.2: Bivariate scatter plots of 1000 thinned posterior draw for the dynamic Gamma-Poisson count model without diurnal pattern.

updating (ϕ, σ, x) in one block using our proposed method, and $\beta = (\mu, \alpha)$ in a second block, directly from its conditional posterior distribution. For the Gamma-Poisson model, we update the shape parameter r using a Metropolis-Hastings step. The proposal is a Student's t distribution, $t_\nu(\hat{r}, \hat{s})$, where $\hat{r}$ is the approximate mode of $\log p(r|x, y)$ and $\hat{s}$ is minus the inverse of its second derivative; the degrees of freedom ν is set to 15. Again, numerical efficiency is quite high for all parameters, especially for the Poisson model.

We also estimates both dynamic count model without diurnal pattern, i.e. by fixing $\alpha = 0$ in equation (1.10), to compare our method with the sampler of McCausland (2012) which involve pre-computing the shape of $p(\mu, \phi, \sigma|y)$ or $p(\mu, \phi, \sigma, r|y)$. The results are reported in the lower panel of Table 1.2. For the Poisson count model, we obtain numerical efficiency for ϕ and σ that is very similar to that reported in McCausland (2012); for the Gamma-Poisson model, the numerical efficiency is about 1.8 times lower for ϕ and σ and about 2.5 times lower for r. Lower efficiencies for the Gamma-Poisson model can be explained by the presence of posterior correlations among ϕ, σ and r and the fact that our sampler updates r conditionally, given all other parameters and state variables, rather than jointly, as in McCausland (2012); Figure 1.2 shows two-way scatter plots for 1000 thinned posterior draws of ϕ, σ and r. Numerically efficiency is still considerably high, especially for non-Gaussian data.

1.4 Concluding remarks

With modest additional computation, we can compute approximations of the gradient and Hessian of $\ln p(\theta|y)$ in univariate non-linear non-Gaussian state space models. This allows us to construct a one-block posterior sampler for (θ, x) that can be used in a stand-alone basic model or in a more general model which embeds the basic one.

In an empirical application using a Gaussian stochastic volatility model with no additional parameters, we show that the approximation is good enough to achieve high numerical efficiency for the 23 return series we investigate. As a second example, we analyse a data set of transaction counts using a dynamic Poisson and a dynamic Gamma-Poisson count model with and without a diurnal pattern. Again, we achieve high numerical efficiency for both the state parameters and the parameters that describe the diurnal pattern. Despite the posterior correlation observed between θ and the shape parameter of the Gamma-Poisson distribution, the high numerical efficiency shows that our method provides a good approximation of the target, even as its shape changes from draw to draw.

Chapter 2

A flexible stochastic conditional duration model*

2.1 Introduction

An important aspect of financial data analysis is the modelling of durations between events. The ever increasing availability of intra-daily data at the highest possible frequency allows researchers to analyse the occurrence of almost any event of interest, and thus to shed light on various market features. For instance, the time between transactions (trade duration) and the length of time it takes for a cumulative volume of trade to reach a fixed threshold (volume duration) are known to mirror market liquidity (Gourieroux et al. 1999). The length of time it takes for the price of an asset to change by at least a fixed threshold (price duration) has been shown to be linked to instantaneous volatility (Engle & Russell 1998). The ability to correctly model financial duration is thus of crucial importance for empirical market microstructure analysis.

Financial durations are known to exhibit strong serial correlation. To capture this time dependence, they are usually modelled using multiplicative error models (Engle 2002) where the scale of the conditional duration distribution depends on the history of the process, as well as other observables such as time of day. We will call the conditional distribution divided by its scale the *normalized* conditional distribution. Two basic models are the autoregressive conditional duration model (ACD, Engle & Russell 1998), and the stochastic conditional duration model (SCD, Bauwens & Veredas

*This chapter is co-authored with my advisor William J. McCausland.

2004). The ACD model, like (G)ARCH models of market volatility, is data driven, where the scale depends deterministically on past durations.[1] The SCD model, like stochastic volatility models, is parameter driven, and the scale is a latent stochastic process.[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).

Our paper makes two main contributions. First, we propose a new SCD model for financial durations. We identify some undesirable features of existing parametric conditional duration densities and propose a new family of conditional distributions that is flexible, but also amenable to shrinkage towards the exponential distribution, which we will argue is a theoretically appealing first-order model. Our second contribution is computational, and promotes highly numerically efficient posterior simulation. We use the HESSIAN method introduced by McCausland (2012), a numerically efficient simulation smoothing method, that we combine with the adaptive algorithm proposed by Vihola (2012) to draw, in a single Gibbs block, the full sequence of state variable describing the realization intensity of the financial event under analysis, together with various parameters. To date, the HESSIAN method has only been applied to non-Gaussian state space models with parametric distributions for observed variables and homogenous state transitions. Here we use it in a model with a flexible distribution for observables and heterogenous state transitions. The numerical efficiency we achieve compares very favourably to that achieved using auxiliary mixture methods, which rely on special features of parametric distributions.

Two other features of our model are appealing and uncommon, but not original. First, we estimate the regular diurnal (time of day) pattern of trading intensity jointly with other features of the model.[3] Second, the latent state process is an irregularly sampled Ornstein-Uhlenbeck (OU) process, rather than a homogenous autoregressive process; autocorrelations depend on the elapsed time, rather than the number of intervening trades, between two durations.[4]

We will be introducing a flexible conditional distribution and so we will take some time to motivate

[1] 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.

[2] In the case of ACD models, "conditional" refers to conditioning on the history of observables; in the case of SCD models, it refers to conditioning on the current value of the latent state variable. Since the states in SCD models are unobserved, the conditional duration density given only past observations 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).

[3] 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.

[4] For instance, Koopman et al. (2008) proposed a multi-state latent intensity model in which they modelled the dynamic process of the latent factor as an OU process.

it. We will not just point out, as others have done, that any parametric distribution is inflexible, although that is a good point in itself. We will also identify particular weaknesses of these distributions.

The most commonly used conditional distributions in ACD and SCD models are the exponential, and two generalizations: 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 others conclude that the exponential is too inflexible and favour the Weibull or gamma. Except for the knife-edge special case where they reduce to an exponential, both the gamma and the Weibull distribution have the property that their hazard function is either zero at a duration of zero and strictly increasing; or infinite at a duration of zero and strictly decreasing.

The generalized gamma and Burr distributions generalize the gamma and the Weibull distribution, respectively.[5] They were proposed as conditional distributions for ACD models by Lunde (1999) and Grammig & Maurer (2000). These distributions, unlike the gamma and the Weibull, allow for non-monotonic hazard functions, but they retain the property that their hazard function is bounded away from zero and infinity only for very special cases.

We argue that extreme variation of the hazard function near zero is implausible. Rather, hazard functions should be bounded away from zero and infinity, and the ratio of supremum to infimum of the hazard function should not be too large. Our argument applies most naturally to trade durations or to any type of durations between events that take a single transaction to occur, like a change in the best ask or bid price available. Price and volume durations are somewhat different, as it might take several transactions before the price change or cumulative volume reaches the required threshold. We first note that this makes unbounded hazards at zero even less plausible for price and volume durations. However, it does suggest that the hazard function at zero might be quite small and increasing; empirical evidence in Grammig & Maurer (2000) and elsewhere suggests that this is indeed the case. Even so, we consider it implausible that the hazard function should equal zero at a duration of zero: price changes and volumes both have long right tails and the probability of a single trade crossing the price or volume threshold within the first second should not be only a tiny fraction of the probability of the threshold being crossed with a trade occurring, say, between seconds nine and ten.

In queueing theory, a simple model for arrival times (of, say, customers at an ATM) is the Poisson

[5] The Weibull distribution is also a special case of the generalized gamma distribution.

process. It is reasonable when there are a large number of potential customers, acting independently and homogenously 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 that interval. In a Poisson process, durations between arrivals are exponentially distributed. The constant hazard function of the exponential makes it *memoryless*: the probability of an arrival in the next minute does not depend on how long you have been waiting.

Of course, activity in financial markets varies over time. But after conditioning on relevant predictors and latent states measuring realization intensity, we would expect the distribution of durations to be not too far from an exponential—its hazard rate a function of this conditioning information—due to the large number of unrelated potential traders, most of whom are small and anonymous. 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, it should have the flexibility to capture distributions whose hazard function fluctuates more widely, if the data support this strongly enough.

The hazard functions of mixtures of exponentials are bounded away from zero and infinity,[6] but they are decreasing—see Barlow et al. (1963)—which is a restrictive feature. Moreover, not all decreasing hazard functions can be easily captured by mixtures of exponentials.[7] In simulations we do not report, we find that adding mixture components after the second yields little: only components with the largest and smallest hazards are important, in the sense that the posterior distributions of the weights of other components are highly concentrated near zero. This suggests that if one allowed the density of *any* linear combination of exponential distributions the posterior distribution would assign high probability to the region where at least one coefficient is negative.

In this paper, we propose a SCD model that features a flexible normalized conditional density capable of matching a wide variety of distributions with moderately varying hazard functions. The normalized conditional distribution takes the form of a perturbed version of an exponential distribution. We use a Bernstein density, a mixture of Beta densities whose parameter values are integers and depend only on the number of components, to capture deviations from an exponential. This approach allows us to easily introduce nonparametric deviations, and to center a prior distribution

[6] DeLuca & Gallo (2004) used a mixture of two exponentials 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 exponentials but allow mixture weights to depend on observable market activity.

[7] 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 two-component mixtures: two exponentials, two Weibulls and two gammas.

around what we argue is a reasonable first order model.[8]

The rest of the paper is organized as follows. We describe our SCD model in Section 2.2 and our methods for posterior simulation in Section 2.3. In Section 2.4, we conduct an artificial data experiment to test for the correctness of our posterior simulator and then illustrate our methods in an application featuring two equities traded on the Toronto Stock Exchange. We conclude in Section 2.5.

2.2 A Stochastic Conditional Duration Model

2.2.1 The Data Generating Process

We observe realization times of a financial event of interest over D trading days in the interval $[t_{\text{open}}, t_{\text{close}}]$, where t_{open} and t_{close} are the opening and closing times. All times of day are measured in seconds after midnight. For each day $d = 1, \ldots, D$, denote the sequence of realization times by $t_{d,0} < t_{d,1} < \cdots < t_{d,n_d}$ and construct the durations between consecutive events $y_{d,i} \equiv t_{d,i} - t_{d,i-1}$, $i = 1, \ldots, n_d$. Following Bauwens & Veredas (2004), our model gives durations as

$$y_{d,i} = e^{x_{d,i}} \epsilon_{d,i}, \tag{2.1}$$

where $x_{d,i} \in \mathbb{R}$ is a latent state process giving the realization intensity of the financial event at $t_{d,i-1}$, and $\epsilon_{d,i}$ is a positive iid process with $\mathrm{E}[\epsilon_{d,i}] = 1$. The two processes are independent which assumes away any leverage-like effect (see Feng et al. 2004, Xu et al. 2011, Men et al. 2015).

At each day d, the latent intensity process $x_d(t)$, $t \in [t_{\text{open}}, t_{\text{close}}]$ is the sum of a daily random effect component ψ_d capturing daily heterogeneity in the mean of durations, a common deterministic function $m(t)$ (constrained to be zero on average) describing a diurnal pattern, and a zero-mean OU process. Sampling the $x_d(t)$ process at all realization times gives $x_{d,i} \equiv x_d(t_{d,i-1})$; then the discrete time process $x_{d,i}$ is first order autoregressive, but not homogeneous due to the irregularly spaced realization times:

$$x_{d,i+1} \,|\, x_{d,i} \sim \mathcal{N}\left((1 - e^{-\rho y_{d,i}})\psi_d + m(t_{d,i}) + e^{-\rho y_{d,i}}(x_{d,i} - m(t_{d,i-1})), \sigma^2(1 - e^{-2\rho y_{d,i}})\right), \tag{2.2}$$

$$x_{d,1} \sim \mathcal{N}\left(\psi_d + m(t_{d,0}), \sigma^2\right), \tag{2.3}$$

[8] See Chen et al. (2014) for an application of Bernstein densities to accelerated hazards model.

where $\rho > 0$ is the mean reversion parameter, and σ is the marginal standard deviation parameter of the OU process. The daily components follow a stationary Gaussian AR(1) process with mean μ, first order autocorrelation $\phi \in (-1, 1)$ and innovation variance ξ^2:

$$\psi_1 \sim \mathcal{N}\left(\mu, \xi^2/(1-\phi^2)\right), \qquad \psi_d \,|\, \psi_{d-1} \sim \mathcal{N}\left(\mu + \phi(\gamma_{d-1} - \mu), \xi^2\right). \tag{2.4}$$

To model the diurnal pattern, we specify $m(t)$ as a cubic B-spline function, a piecewise polynomial indexed by a set of predetermined knots and expressed as a linear combination of B-spline basis functions:

$$m(t) = \sum_{l=1}^{L} \delta_l B_l(t), \tag{2.5}$$

where $B_l(\cdot)$ denotes the l-th B-spline basis function, a local cubic polynomial, and δ_l its coefficient. The basis functions depend on the location and the multiplicity of the knots. Following Eilers & Marx (1996), we consider equally spaced knots over $[t_{\text{open}}, t_{\text{close}}]$. The first and last knots have multiplicity 4 and the rest have multiplicity 1, which makes $m(t_{\text{open}}) = \delta_1$ and $m(t_{\text{close}}) = \delta_L$. B-spline basis functions have the property that $\sum_l B_l(t) = 1$ for all $t \in [t_{\text{open}}, t_{\text{close}}]$. Hence, identification of the model requires that $\sum_l \delta_l = 0$. See de Boor (1978) and Dierckx (1993) for more on B-splines.

We denote the density of $\epsilon_{d,i}$ by $p_\epsilon(\cdot)$ and call it the normalized duration density. Thus, the conditional density of $y_{d,i}$ given the contemporaneous value $x_{d,i}$ of the latent intensity state is

$$p(y_{d,i} \,|\, x_{d,i}) = e^{-x_{d,i}} p_\epsilon(e^{-x_{d,i}} y_{d,i}). \tag{2.6}$$

The quantity $e^{x_{d,i}}$ is the conditional mean of $y_{d,i}$. It gives only the scale of the distribution, the shape being determined by $p_\epsilon(\cdot)$. We argued that some commonly used duration distributions are unsuitable because of their restrictive or implausible hazard functions. We will now propose a new flexible family of normalized densities capable of matching a wider variety of hazard functions than a default parametric distribution.[9]

[9] In SCD models, the conditional density given past observations, analogous to the conditional density in ACD models, is a mixture distribution. Hence, flexibility can be added to SCD models through the normalized density or the mixing distribution (i.e. the filtering distribution of the latent intensity states). In this paper, we propose a more flexible normalized density and remain parametric for the mixing distribution. Relaxing the usual Gaussian assumption for the dynamics of the latent states would introduce additional computational challenges in the estimation of the model. A potential feasible extension could to replace the Gaussian innovation in the state dynamic by a parametric skewed and heavy-tailed innovation distribution that can be represented as a Gaussian mixture. We leave this possibility for future research.

2.2.2 A Normalized Density for Durations

We adopt an approach similar to the one described in Ferreira & Steel (2006), where the cdf $P_\epsilon(\cdot)$ of a flexible univariate distribution is constructed as $P_\epsilon(\epsilon) \equiv G(F(\epsilon))$, where $F(\cdot)$ is a parametric continuous cdf with density $f(\cdot)$ having the same support as $P_\epsilon(\cdot)$; and $G(\cdot)$ is a flexible continuous cdf on $[0,1]$ with density $g(\cdot)$. The cdf $P_\epsilon(\cdot)$ can be viewed as a perturbed version, depending on $G(\cdot)$, of the original parametric cdf $F(\cdot)$. When $G(\cdot)$ is uniform on $[0,1]$, there is no perturbation and $P_\epsilon(\cdot) = F(\cdot)$. As noted by Ferreira & Steel (2006), distributions defined in this way cover the entire class of continuous distributions, since any such $P_\epsilon(\cdot)$ can be constructed for a suitable choice of $G(\cdot)$. This construction implies the normalized density

$$p_\epsilon(\epsilon) = f(\epsilon)g\big(F(\epsilon)\big). \tag{2.7}$$

We argued that an exponential distribution was a theoretically promising first order approximation of a conditional duration distribution, and so we specify the original distribution $F(\cdot)$ as an exponential in the hope of capturing realistic hazard functions using a parsimonious distortion distribution $G(\cdot)$. The hazard rate of $F(\cdot)$, which we denote λ, will be substituted out using the scale normalization condition $\mathrm{E}[\epsilon] = 1$. At the same time, we want $P_\epsilon(\cdot)$ to be flexible, allowing for large departures from the exponential if the data warrant it. For this reason, we choose a flexible functional form for the distortion distribution and specify $g(\cdot)$ as a J'th order Bernstein density.[10] This is a J-component mixture of beta densities, each with two integer-valued shape parameter adding to $J+1$; coefficients of the first and last component determine $g(0)$ and $g(1)$, respectively. Specifically,

$$g(z) = \sum_{j=1}^{J} \beta_j \,\mathrm{Beta}(z \,|\, j, J-j+1), \qquad 0 \le z \le 1, \tag{2.8}$$

where $\sum_{j=1}^{J} \beta_j = 1$, $\beta \equiv (\beta_1, \ldots, \beta_J) \ge 0$, and $\mathrm{Beta}(z \,|\, a, b)$ denotes the beta density with shape parameters a and b, for $a, b > 0$. The J components of the J'th order Bernstein density form a partition of unity, so that if $\beta = (1/J, \ldots, 1/J)$, then $g(\cdot)$ is uniform and we get back the original exponential density, $p_e(\cdot) = f(\cdot)$. This makes it easy to choose a prior distribution for β so as to centre the induced prior for $p_\epsilon(\cdot)$ around the exponential distribution and to control the amount of shrinkage towards it. The order J governs the range of possible deviations from the exponential

[10] A Bernstein density can approximate any continuous density on $[0,1]$ arbitrarily closely (in sup norm) for J sufficiently large (cf. Lorentz 1953).

distribution. As J gets larger, the Beta densities are more "spiked", which allows more precise detail in the overall shape of $p_\epsilon(\cdot)$. We do not treat the order J of the Bernstein density as a parameter to estimate *per se*, but rather as a fixed model indicator. Instead of estimating J jointly with all other parameters using trans-dimensional methods, we will compare results over different values of J and compute the posterior probability of each specification.[11] For a discussion of Bayesian nonparametric density estimation using Bernstein densities with unknown J, see Petrone (1999*a*,*b*) and Petrone & Wasserman (2002).

Substituting the exponential distribution function $F(\epsilon) = 1 - e^{-\lambda\epsilon}$ and the above expression for $g(\cdot)$ into equation (2.7) gives $p_\epsilon(\cdot)$ as a polynomial in $e^{-\lambda\epsilon}$, which we write explicitly as the following linear combination of exponential densities:

$$p_\epsilon(\epsilon) = \sum_{j=1}^{J} \beta_j \left[\sum_{k=1}^{j} a_{j,k}(J-j+k)\lambda e^{-(J-j+k)\lambda\epsilon} \right], \tag{2.9}$$

where

$$a_{j,k} = \binom{j-1}{k-1} \frac{\Gamma(J+1)}{\Gamma(j)\Gamma(J-j+1)} \frac{(-1)^{k+1}}{J-j+k}. \tag{2.10}$$

Non-negativity of β ensures the non-negativity of $p_\epsilon(\cdot)$, but as the $a_{j,k}$ are not necessarily all non-negative, $p_\epsilon(\cdot)$ is not necessarily a mixture of exponentials. It is easy to check that the hazard function for $p_\epsilon(\cdot)$ is bounded away from zero and infinity, and its limiting value as ϵ goes to infinity is the hazard parameter λ of the exponential distribution with cdf $F(\cdot)$. The scale normalization condition $\mathrm{E}[\epsilon] = 1$ gives $\lambda = \sum_{j=1}^{J} \beta_j \sum_{k=1}^{j} a_{j,k}/(J-j+k)$, which we use to substitute out λ from the expression for $p_\epsilon(\cdot)$, freeing us from having to impose restriction on β.

We now illustrate the flexibility we achieve with just a few terms. Figure 2.1 shows examples of decreasing, increasing and non-monotonic hazard functions that can be captured with $J = 3$ components. Taking into account the adding-up constraint, there are two degrees of freedom, the same as a unit-mean mixture of two exponentials. The solid lines are the density and (constant) hazard functions of an exponential with mean equal to one. The dashed lines show pairs of density and hazard functions where the hazard is monotonic; dash-dotted lines, pairs where it is not monotonic. Recall that all mixtures of exponentials have a decreasing hazard function.

[11] See Quintana et al. (2009) for an example of inference using reversible-jump Markov chain Monte-Carlo methods in a similar application of Bernstein densities, but in models without state variables.

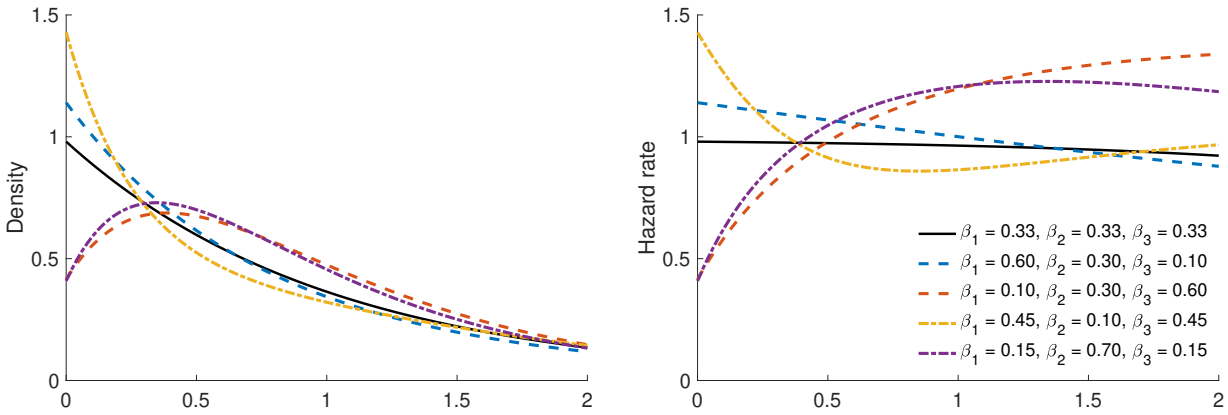

Figure 2.1: Density and hazard functions for a unit-mean exponential and five normalized distributions $p_\epsilon(\cdot)$ with $J = 3$ terms.

2.2.3 Prior Distributions

To complete the model, we describe a prior distribution for the parameters of our SCD model $\Theta \equiv (\sigma, \rho, \mu, \phi, \omega, \delta, \beta)$. All parameters are *a priori* independent. In the following, the overbar notation $\bar{\cdot}$ is used to denote prior hyperparameters whose values will be specified in Section 2.4.

We specify log-normal prior distributions for the marginal standard deviation σ and the mean reversion parameter ρ of the OU process: $\log \sigma \sim \mathcal{N}(\bar{\mu}_\sigma, \bar{h}_\sigma^{-1})$ and $\log \rho \sim \mathcal{N}(\bar{\mu}_\rho, \bar{h}_\rho^{-1})$. The transformation eliminates the need for parameter restrictions. We select commonly used prior distributions for the parameters of the daily component dynamics: we specify a Normal distribution for the mean, $\mu \sim \mathcal{N}(\bar{\mu}_\mu, \bar{h}_\mu^{-1})$, a Normal distribution truncated to the stationarity region for the first order autocorrelation, $\phi \sim \mathcal{N}(\bar{\mu}_\phi, \bar{h}_\phi^{-1})\mathbf{1}\{\phi \in (-1, 1)\}$, and a scaled chi-square for the inverse innovation variance, $\bar{s}_\xi \xi^{-2} \sim \chi^2(\bar{\nu}_\xi)$.

Following Lang & Brezger (2004), we specify a first order Gaussian random walk prior for the adjacent coefficients of the vector δ defining the diurnal pattern:

$$\nabla \delta \sim \mathcal{N}(0_{L-1}, \tau^{-1} I_{L-1}), \tag{2.11}$$

where ∇ is the first order backward difference operator with dimension $L \times (L-1)$, 0_{L-1} is an $(L-1) \times 1$ vector of zeros, I_{L-1} is the $(L-1) \times (L-1)$ identity matrix and τ is a scaler precision parameter. Thus, the first order differences $\delta_l - \delta_{l-1}$ are iid $\mathcal{N}(0, \tau^{-1})$. This prior favour smoothness and is agnostic with respect to the signs of the derivatives; since the derivative of B-splines are linear combinations of the first difference $\delta_l - \delta_{l-1}$, we can interpret τ as a smoothing parameter for the

diurnal pattern. Higher values of τ imply more shrinkage towards a flat diurnal pattern. Following common practices, we will estimate τ, and specify its prior as the following scaled chi-square: $\bar{s}_\tau \tau \sim \chi^2(\bar{\nu}_\tau)$.

Combined with the identification restriction $\sum_{l=1}^{L} \delta_l = 0$, equation (2.11) induces a proper prior distribution for the vector δ with the symmetry property that $(\delta_1, \ldots, \delta_L)$ and $(\delta_L, \ldots, \delta_1)$ have the same distribution, and prior uncertainty about the realization intensity of the event of interest is the same at the opening and at the closing of the market. This implies the conditional prior $\tilde{\delta} \,|\, \tau \sim \mathcal{N}(0_{L-1}, (\tau \nabla T)^{-1})$, where $\delta \equiv T\tilde{\delta}$ and T is a linear transformation that enforces identification. We will refer to this specification in Section 2.3.

For the beta mixture weights indexing the normalized duration density $p_\epsilon(\cdot)$, we specify a Dirichlet distribution: $\beta \sim \text{Dirichlet}(\bar{M}\bar{\beta})$, where $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_J) > 0$, $\sum \bar{\beta}_j = 1$, and $\bar{M} > 0$. The prior mean of β is $\bar{\beta}$; $\bar{M}$ is a concentration parameter. When $\bar{\beta} = (1/J, \ldots, 1/J)$, the prior mean of β corresponds to $g(\cdot)$ being uniform on $[0, 1]$ and therefore $p_\epsilon(\cdot) = f(\cdot)$ being an exponential distribution. In this sense, $\bar{\beta} = (1/J, \ldots, 1/J)$ centres the prior distribution for $p_\epsilon(\cdot)$ around an exponential density, with $\bar{M}$ controlling the amount of shrinkage towards it; higher values of $\bar{M}$ imply more shrinkage.

2.2.4 Joint Density

We conclude the exposition of the model by giving the joint density of all parameters, latent variables and observations, making explicit all conditional independence relationships. We refer to the model as the flexible SCD model (FSCD). Let x and y be the flat vectors of all states and durations, respectively. Then the joint density is

$$p(\sigma, \rho, \mu, \phi, \xi, \tau, \delta, \beta, \psi, x, y) = p(\sigma)\, p(\rho)\, p(\mu)\, p(\phi)\, p(\xi)\, p(\tau)\, p(\delta \,|\, \tau)\, p(\beta)$$
$$\prod_{d=1}^{D} \left[p(\psi_d \,|\, \psi_{d-1}, \mu, \phi, \xi) \prod_{i=1}^{n_d} p(x_{d,i} \,|\, x_{d,i-1}, t_{d,i-1}, t_{d,i-2}, \sigma, \rho, \delta, \psi_d)\, p(y_{d,i} \,|\, x_{d,i}, \beta) \right]. \quad (2.12)$$

The densities for the initial daily component ψ_1 and values $x_{d,1}$ on each day d are understood to be $p(\psi_1 \,|\, \psi_0, \mu, \phi, \xi) \equiv p(\psi_1 \,|\, \mu, \phi, \xi)$ and $p(x_{d,1} \,|\, x_{d,-1}, t_{d,0}, t_{d,-1}, \sigma, \rho, \delta, \psi_d) \equiv p(x_{d,1} \,|\, t_{d,0}, \sigma, \delta, \psi_d)$.

2.3 Bayesian Inference

In this section, we first describe posterior simulation methods for Bayesian inference in our flexible SCD model followed by how to compute the marginal likelihood for models for a fixed value of J, the number of terms in the normalized duration density.

2.3.1 Posterior simulation

We use Markov chain Monte Carlo (MCMC) to sample the joint posterior distribution of parameters and state variables. We break the posterior distribution into five blocks, updating (σ, ρ, x), ψ, (μ, ϕ, ω), (τ, δ) and β. We now describe each of the Gibbs blocks in turn.

Drawing from $p(\sigma, \rho, x \,|\, \delta, \psi, z, y)$

It is widely known that when there is strong posterior dependence between state variables (here, x) and the parameters of their dynamics (here, σ and ρ), updating both in a single block improves numerical efficiency. However, in non-linear non-Gaussian state space models, it is difficult to draw the state sequence as a single block, even when conditioning on the parameters of the state dynamics. Several methods have been proposed, and many of them have been applied to draw latent volatilities in stochastic volatility models (see, e.g., Kim et al. 1998, Chib et al. 2002, Richard & Zhang 2007). Here, an additional difficulty is the nonparametric nature of the measurement distribution, which appears to rule out methods based on auxiliary mixture models. In this paper, we use the HESSIAN method of McCausland (2012), a procedure to draw a state sequence as a single block, to construct a sampler to jointly draw the state sequence and its associated parameters.[12]

We update (σ, ρ, x) in two steps. We first draw component indicators $z_{d,i} \in \{1, \ldots, J\}$ for each duration. The indicators are conditionally independent, with probability mass function given, up to a multiplicative factor, by

$$\Pr[z_{d,i} = j \,|\, \beta, x, y] \propto \beta_j \left[\sum_{k=1}^{j} a_{j,k} (J - j + k) \lambda e^{-(J-j+k)\lambda e^{-x_{d,i}} y_{d,i}} \right],$$

[12] McCausland (2012) shows how to adapt the HESSIAN method to draw, in a single block, all latent states and all parameters in several univariate state space models by pre-computing the shape of the posterior distribution using Laplace-like approximation of the likelihood. However, a similar strategy would have been difficult to implement in our case in part due to the dimension of the model parametric space, but also due to the flexible duration density.

for $d = 1, \ldots, D$ and $i = 1, \ldots, n_d$, where the coefficients $a_{j,k}$ are given by (2.10). Recall that β determines the hazard parameter λ of the exponential distribution $F(\cdot)$ through the unit-mean normalization of the duration density $p_\epsilon(\cdot)$. We then update (σ, ρ, x) given the component indicators $z = (z_{1,1}, \ldots, z_{1,n_1}, \ldots, z_{D,1}, \ldots, z_{D,n_D})$ in a single block. Our joint proposal (σ^*, ρ^*, x^*) consists of a random walk proposal for $(\log \sigma^*, \log \rho^*)$ followed by a conditional proposal $x^* \,|\, \sigma^*, \rho^*$ drawn from a proposal density $q(x \,|\, \sigma^*, \rho^*, \delta, \psi, z, y)$. We accept the triple (σ^*, ρ^*, x^*) with probability

$$\min \left\{ 1, \frac{p(x^*, y \,|\, \sigma^*, \rho^*, \delta, \psi, z) p(\sigma^*) p(\rho^*)}{p(x, y \,|\, \sigma, \rho^*, \delta, \psi, z) p(\sigma) p(\rho)} \times \frac{q(x \,|\, \sigma, \rho, \delta, \psi, z, y)}{q(x^* \,|\, \sigma^*, \rho^*, \delta, \psi, z, y)} \right\}.$$

The random walk proposal is Gaussian with covariance Σ. During the burn-in period, we use the approach described in Vihola (2012) to tune the proposal distribution by adapting Σ to track an acceptance probability of 0.352, the optimal acceptance rate for 2 dimensional settings—see Gelman et al. (1996). We use the final value of Σ at the end of the burn-in period for subsequent draws to ensures that our posterior simulator is Markov.

We draw $x \,|\, \sigma^*, \rho^*, \delta, \psi, z, y$ using the HESSIAN method. In general terms, and suppressing notation for any parameters there may be, this gives a close approximation $q(x \,|\, y)$ of the conditional density $p(x \,|\, y)$ of the state sequence x given the observed sequence y, for state space models with univariate states 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 homogenous. The method is generic, as the only model-specific code required consists of a routine to evaluate $\log p(y_i \,|\, x_i)$, and its first five derivatives with respect to x_i, at a given point. For our stochastic duration model, we compute exact values of these derivatives without deriving analytic expressions for them; instead, we exploit automatic routines to combine evaluations of derivatives of primitive functions using Faá di Bruno's rule, which is much easier. Details are provided in Appendix B.1.

Drawing from $p(\psi \,|\, \sigma, \rho, \mu, \phi, \xi, \delta, x)$

Notice that equations (2.2) and (2.4) describe a univariate and linear Gaussian state space model. Hence, conditionally on (μ, ϕ, ξ, x), the vector of daily components ψ can be drawn in a single step from its conditional posterior distribution using computationally efficient precision sampling methods (cf. Chan & Jeliazkov 2009, McCausland et al. 2011). The equation (2.4) can be written such that $\psi \,|\, \mu, \phi, \xi \sim \mathcal{N}(\bar{\psi}, \bar{H}_\psi^{-1})$, where the conditional prior mean is $\bar{\psi} = (\mu, \ldots, \mu)$ and the

conditional prior precision is $\bar{H}_\psi = \Phi'\Omega\Phi$ with $\Omega = diag(\xi^{-2}(1-\phi^2), \xi^{-2}, \ldots, \xi^{-2})$ and

$$\Phi = \begin{bmatrix} 1 & & & \\ -\phi & 1 & & \\ & \ddots & \ddots & \\ & & -\phi & 1 \end{bmatrix}.$$

Using this representation, the conditional posterior distribution of ψ is $\psi \,|\, \mu, \phi, \xi, x \sim \mathcal{N}(\bar{\bar{\psi}}, \bar{\bar{H}}_\psi^{-1})$, where the posterior precision is $\bar{\bar{H}}_\psi = \bar{H}_\psi + K'K$ and the posterior mean is $\bar{\bar{\psi}} = \bar{\bar{H}}_\psi^{-1}(\bar{H}_\psi\bar{\psi} + K'w)$. The vector u and matrix K come from writing the state equation (2.2) as $u \sim \mathcal{N}(K\psi, I_N)$ where $u = (u_1, \ldots, u_D)'$ and $K = diag(K_{11}, \ldots, K_{DD})$ is block diagonal with, for $d = 1, \ldots, D$,

$$u_d = \begin{bmatrix} (x_{d,1} - m(t_{d,0}))/\sigma \\ (x_{d,2} - m(t_{d,1}) - \exp(-\rho y_{d,1})(x_{d,1} - m(t_{d,0}))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,1}))} \\ \vdots \\ (x_{d,n_d} - m(t_{d,n_d-1}) - \exp(-\rho y_{d,n_d-1})(x_{d,n_d-1} - m(t_{d,n_d-2}))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,n_{d-1}}))} \end{bmatrix}$$

and

$$K_{dd} = \begin{bmatrix} 1/\sigma \\ (1-\exp(-\rho y_{d,1}))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,1}))} \\ \vdots \\ (1-\exp(-\rho y_{d,n_{d-1}}))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,n_{d-1}}))} \end{bmatrix}.$$

Drawing from $p(\mu, \phi, \xi \,|\, \psi)$

We update (μ, ϕ, ξ) using three sub-blocks. We first draw μ from its conditional posterior distribution: $\mu \,|\, \phi, \xi, \psi \sim \mathcal{N}(\bar{\bar{\mu}}_\mu, \bar{\bar{h}}_\mu^{-1})$, where

$$\bar{\bar{h}}_\mu = \bar{h}_\mu + \xi^{-2}((1-\phi^2) + (D-1)(1-\phi)^2), \qquad \bar{\bar{\mu}} = \bar{\bar{h}}_\mu^{-1}(\bar{h}_\mu\bar{\mu}_\mu + \xi^{-2}((1-\phi^2)\psi_1 + (1-\phi)\sum_{d=2}^{D}(\psi_d - \phi\psi_{d-1})).$$

We next update ϕ using a Metropolis step: we draw a proposal $\phi^* \sim \mathcal{N}(\bar{\bar{\mu}}_\phi, \bar{\bar{h}}_\phi^{-1})$, where

$$\bar{\bar{h}}_\phi = \bar{h}_\phi + \xi^{-2}\sum_{d=2}^{D}(\psi_{d-1} - \mu)^2, \qquad \bar{\bar{\mu}}_\phi = \bar{\bar{h}}_\phi^{-1}\left(\bar{h}_\phi\bar{\phi} + \xi^{-2}\sum_{d=2}^{D}(\psi_d - \mu)(\psi_{d-1} - \mu)\right),$$

and accept the proposal with probability

$$\left\{1, \left(\frac{1-\phi^{*2}}{1-\phi^2}\right)^{1/2} \exp\left(-\frac{\omega}{2}(\psi_1-\mu)^2(\phi^2-\phi^{*2})\right)\right\}.$$

We then draw ξ^{-2} from its conditional posterior distribution: $\bar{\bar{s}}_\xi \xi^{-2} \,|\, \mu, \phi, \psi \sim \chi^2(\bar{\bar{\nu}}_\xi)$, where

$$\bar{\bar{\nu}}_\xi = \bar{\nu}_\xi + D, \qquad \bar{\bar{s}}_\xi = \bar{s}_\xi + (1-\phi^2)(\psi_1-\mu)^2 + \sum_{d=2}^{D}(\psi_d - \phi\psi_{d-1} - (1-\phi)\mu))^2.$$

Drawing from $p(\tau, \delta \,|\, \sigma, \rho, \psi, x)$

We update (τ, δ) using two sub-blocks. We first draw τ from its conditional posterior distribution: $\bar{\bar{s}}_\tau \tau \,|\, \delta \sim \chi^2(\bar{\bar{\nu}}_\tau)$, where $\bar{\bar{s}}_\tau = \bar{s}_\tau + \delta'\nabla'\nabla\delta$ and $\bar{\bar{\nu}}_\tau = \bar{\nu}_\tau + L - 1$. We then draw $\delta \equiv T\tilde{\delta}$ from its conditional posterior distribution: $\tilde{\delta} \,|\, \sigma, \rho, \tau, \psi, x \sim \mathcal{N}(\bar{\bar{\tilde{\delta}}}, \bar{\bar{H}}_\delta^{-1})$, where the posterior precision is $\bar{\bar{H}}_\delta = \tau\nabla T + W'T'TW$ and the posterior mean is $\bar{\bar{\tilde{\delta}}} = \bar{\bar{H}}_\delta^{-1} W'T'v$. The vector v and matrix W come from writing the state equation (2.2) as $v \sim \mathcal{N}(W\delta, I_N)$ where v and W are organized in blocks

$$\begin{bmatrix} v_1 \\ \vdots \\ v_D \end{bmatrix}, \quad \begin{bmatrix} W_{11} & \dots & W_{1L} \\ \vdots & \ddots & \vdots \\ W_{D1} & \dots & W_{DL} \end{bmatrix},$$

with, for $d = 1, \dots, D$ and $l = 1, \dots, L$,

$$v_d = \begin{bmatrix} (x_{d,1} - \psi_d)/\sigma \\ (x_{d,2} - \psi_d - \exp(-\rho y_{d,1})(x_{d,1} - \psi_d))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,1}))} \\ \vdots \\ (x_{d,n_d} - \psi_d - \exp(-\rho y_{d,n_{d-1}})(x_{d,n_{d-1}} - \psi_d))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,n_{d-1}}))} \end{bmatrix}$$

and

$$W_{dl} = \begin{bmatrix} B_l(t_{d,1})/\sigma \\ (B_l(t_{d,2}) - \exp(-\rho y_{d,1})B_l(t_{d,1}))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,1}))} \\ \vdots \\ (B_l(t_{d,n_d}) - \exp(-\rho y_{d,n_{d-1}})B_l(t_{d,n_d-1}))/\sqrt{\sigma^2(1-\exp(-2\rho y_{d,n_{d-1}}))} \end{bmatrix}.$$

Drawing from $p(\beta \mid x, y)$

We update β using a multiple-try algorithm that generalizes the random walk algorithm by considering multiple candidates values at each iteration (Liu 2004). For the purposes of drawing proposals, we use the logistic transformation $\vartheta(\beta) = \log(\beta_1/\beta_J), \ldots, \log(\beta_{J-1}/\beta_J))$, which maps the $J-1$ dimensional simplex to $\mathbb{R}^{J-1}$. The absence of positivity and adding-up constraints is convenient, as is the fact that the posterior distribution of ϑ is more nearly Gaussian than that of β, especially when some elements of β have high posterior density close to zero. The inverse transformation is $\beta(\vartheta) = (1 + \sum_{j=1}^{J-1} \exp(\vartheta_j))^{-1}(\exp(\vartheta_1, \ldots, \exp(\vartheta_{J-1}), 1)$ and the induced prior on ϑ is

$$p(\vartheta) = p(\beta(\vartheta)) \prod_{j=1}^{J} \beta_j(\vartheta).$$

The multiple-try algorithm goes as follows: we first generate R candidates $\vartheta_1^*, \ldots, \vartheta_R^* \sim \mathcal{N}(\vartheta, \Xi)$ and draw a proposal $\vartheta_{\tilde{r}}^* \in \{\vartheta_1^*, \ldots, \vartheta_R^*\}$ with probability proportional to $p(y \mid \vartheta_r^*, x)p(\vartheta_r^*)$. We then generate $R-1$ reference points $\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_{R-1} \sim \mathcal{N}(\vartheta_{\tilde{r}}^*, \Xi)$ and accept the proposal $\vartheta_{\tilde{r}}^*$ with probability

$$\min \left\{ 1, \frac{\sum_{r=1}^{R} p(y \mid \vartheta_r^*, x)p(\vartheta_r^*)}{p(y \mid \vartheta, x)p(\vartheta) + \sum_{r=1}^{R-1} p(y \mid \tilde{\vartheta}_r, x)p(\tilde{\vartheta}_r)} \right\}.$$

The proposal covariance Ξ is a rescaled approximation of the marginal posterior distribution of ϑ, computed during the burn-in period. During the burn-in period, we update Ξ using the adaptive method described in Haario et al. (2001). At the end of the burn-in period, we compute Ξ as the sample covariance of the burn-in period, rescaled by a factor of $2.38(J-1)^{-1/2}$, the optimal scaling factor for multi-dimensional settings—see Gelman et al. (1996). For posterior simulation, we set the number of candidates generated per iteration to $R = 5$.

In computational experiments not reported here, we considered several values of the tuning parameter R and found that the above choice offers a good trade-off between the additional computational cost and the gain in numerical efficiency for the range of values considered in our empirical illustration for J, the number of term in the normalized density. We also compared the multiple-try algorithm with the version of the adaptive random walk metropolis algorithm described in Vihola (2012); both approaches deliver similar performances for lower values of J, but we obtained better results with the former for higher values of J.[13]

[13] Examples of adaptive MCMC algorithms can be found in Roberts & Rosenthal (2009).

2.3.2 Marginal likelihood

In a Bayesian framework, a standard practice for model comparison is to compute for all possible models the marginal likelihood and to apply Bayes' rule to obtain the posterior probabilities of each competing model. This approach can be used to deal with model specification uncertainty for a finite mixture model when the number of admissible components is not too large. See Frühwrith-Schnatter (2006) for a textbook treatment on finite mixture models.

The marginal likelihood is defined as the integral of the likelihood with respect to the prior density of the parameters. Following Chib (1995), we estimate the marginal likelihood $p(y)$ on the log scale as

$$\log p(y) = \log p(y \mid \Theta) + \log p(\Theta) - \log p(\Theta \mid y), \tag{2.13}$$

where Θ denotes all parameters in the model, $p(y \mid \Theta)$ is the likelihood, $p(\Theta)$ is the prior density, and $p(\Theta \mid y)$ is the posterior density. This equality holds for any value of Θ, but it is usually evaluated at a high density point Θ^* (say) to obtain a stable estimate.

The prior density is available directly although the likelihood and posterior ordinate must be evaluated by simulation. We compute the likelihood

$$p(y \mid \Theta^*) = \int p(\psi, x, y \mid \Theta^*)\, d\psi dx$$

using importance sampling and estimate the posterior ordinate through posterior decomposition and additional *reduced* MCMC runs. More specifically, we decompose the posterior ordinate as

$$p(\Theta^* \mid y) = p(\vartheta(\beta^*) \mid y)\, p(\mu^*, \phi^*, \xi^* \mid \beta^*, y)\, p(\delta^* \mid \mu^*, \phi^*, \xi^*, \beta^*, y)\, p(\tau^* \mid \delta^*)\, p(\sigma^*, \rho^* \mid \mu^*, \phi^*, \xi^*, \delta^*, \beta^*, y)$$

where $\vartheta(\beta^*)$ is the logistic transformation of β^*, and estimate each ordinate in turn. We estimate the first term by kernel smoothing applied to the draws from the full MCMC run. We then fix β and do a reduced run, sampling all parameters and state variables except β, and estimate the second term by kernel smoothing applied to the draws from the reduced run. To estimate $p(\delta^* \mid \mu^*, \phi^*, \xi^*, \beta^*, y)$, we fix the parameters (μ, ϕ, ξ, β) and perform a second reduced MCMC run. The required ordinate then follows by averaging the Gaussian density of δ. The ordinate $p(\tau^* \mid \delta^*)$ can be evaluated directly. We then do a third reduced run where all parameters except (σ, ρ) are fixed and we estimate the last ordinate by kernel smoothing applied to the resulting draws.

2.4 Results

Here, we first report results from a pre-data simulation experiment involving only artificial data meant to test for the correctness of our posterior simulation methods. We then illustrate the use of our flexible SCD model with an empirical example using quotation data for equities traded on the Toronto Stock Exchange.

2.4.1 Getting it right (GIR)

The following tests for program correctness are similar to those described in Geweke (2004) and the title of this subsection comes from the title of that paper. The idea of the exercise is to simulate a Markov chain whose stationary distribution is the *joint* distribution of parameters, latent state variables *and data*, making use of the same simulation methods that will be used later for posterior simulation, together with an additional Gibbs block to draw data from their conditional distribution given parameters and state variables.[14] If the posterior simulation methods are correct in concept and implementation, the marginal distribution of the parameter vector, with respect to this stationary distribution, is identical to its (known) prior distribution. This is a strong condition with many easily testable implications.

We will need to simplify part of our model in order to proceed. The problem is that we record values of the underlying continuous-time OU process $x(t)$ only at the times of realization of the event of interest. Redrawing duration data changes the realization times, which requires conditioning on the entire path of continuous-time process, which is impractical. For the GIR simulations only, we modify the latent state process described in (2.2), replacing the OU process with a homogeneous autoregressive process where the transition distribution from $x_{d,i}$ to $x_{d,i+1}$ depends only on $x_{d,i}$ and not on the duration $y_{d,i}$. We parameterize the process in a way that resembles the sampled OU process, giving

$$x_{d,i+1} \,|\, x_{d,i} \sim \mathcal{N}\left((1-e^{-\rho})\psi_d + m(t_{d,i}) + e^{-\rho}(x_{d,i} - m(t_{d,i-1})), \sigma^2(1-e^{-2\rho})\right). \tag{2.14}$$

Sampling from the posterior distribution also requires some minor modifications. Since we draw a new sample of artificial data at each iteration, the adaptive schemes implemented during the burn-in

[14] The additional Gibbs block updating the data y from its conditional distribution given parameters and latent variables for the GIR simulation is described in Appendix B.2.

Table 2.1: Difference between prior and simulation sample first and second moments in the Getting it right experiment

θ	$\mathrm{E}[\theta] - \bar{\theta}$	$\hat{\sigma}_{\text{nse}}$	t-stat	$\mathrm{E}[\theta^2] - \bar{\theta}^2$	$\hat{\sigma}_{\text{nse}}$	t-stat
$\log \sigma$	0.0000076	0.0000485	0.156	-0.0000152	0.0000970	-0.156
$\log \rho$	0.0000357	0.0000490	0.728	-0.0001633	0.0002255	-0.724
μ	0.0000912	0.0003194	0.285	0.0006175	0.0015960	0.387
ϕ	0.0002621	0.0002848	0.920	0.0000444	0.0000236	1.883
ξ^{-2}	-0.0402189	0.0205170	-1.960	-8.2016201	4.1217279	-1.990
τ	-0.0021692	0.0137267	-0.158	-0.1221362	1.3769431	-0.089
β_1	-0.0002561	0.0001820	-1.408	-0.0002527	0.0001822	-1.388
β_2	0.0001376	0.0001499	0.918	0.0000891	0.0000902	0.988
β_3	0.0001185	0.0001363	0.869	0.0000529	0.0000556	0.951

period to tune the proposal distributions used in the Gibbs block updating (σ, ρ, x) and β do not work well. Instead, we fix the proposal covariance matrices to the prior covariance of the parameter vector in question.

We set the number of components of the normalized density to $J = 3$ and use a B-spline function defined on two knots, t_{open} and t_{close}, giving a diurnal pattern that is an expansion with $L = 4$ cubic polynomials. To avoid realization times in simulations occurring after t_{close}, where the diurnal pattern is undefined, we choose the sample size n (the number of durations, not the size of the simulation sample) and prior distributions such that the probability that the last transaction of the day occurs after t_{close} is extremely small. We fix $D = 10$ and choose sample size of $n_d = 10$ observations by day. For each day, the initial transaction time is $t_{d,0} = t_{\text{open}}$ and the length of the trading session is 1800 seconds. We used the following hyperparameter values:

$$
\begin{aligned}
&(\bar{\mu}_\sigma, \bar{h}_\sigma) = (-1.0, 1000), \qquad (\bar{\mu}_\mu, \bar{h}_\mu) = (2.5, 250), \qquad (\bar{s}_\tau, \bar{\nu}_\tau) = (10, 500), \qquad \bar{M} = 200, \\
&(\bar{\mu}_\rho, \bar{h}_\rho) = (-2.3, 1000), \qquad (\bar{\mu}_\phi, \bar{h}_\phi) = (0.0, 250), \qquad (\bar{s}_\xi, \bar{\nu}_\xi) = (10, 1000), \qquad \bar{\beta} = (0.5, 0.3, 0.2).
\end{aligned}
$$

The tighter prior distribution, and much smaller number of observations, compared with the empirical example, ensure high numerical precision with a moderate amount of computation.

We generate a sample of size 5×10^5 and for analysis, we use a subsample of size $50,000$ consisting of every 10'th draw. Table 2.1 shows the results of the comparison of first (columns 2-5) and second (6-8) moments between the prior population and the simulation sample. Columns 2 and 6 give, for selected parameters θ, the difference between the prior population moment $\mathrm{E}[\theta^k]$ and the simulation sample moment $\bar{\theta}^k$ $(k = 1, 2)$; columns 3 and 7, the numerical standard error (i.e. the

simulation standard deviation quantifying error in finite simulations) of the sample moment; and columns 3 and 8, the t-statistic for the test of the hypothesis that the simulation population moment of the parameter (from the stationary distribution of the Markov chain) equals the (known) prior population moment. Each hypothesis is a necessary condition for the correctness of the posterior simulation methods. Numerical standard errors are computed using the overlapping batch means method (Flegal & Jones 2010). Sample moments are close to the true prior moments, relative to the numerical standard error. The results fail to cast doubt on the correctness of the posterior simulator: one hypothesis (out of 18) is rejected at the 10% level and none at the 5% level.

2.4.2 An empirical example

Data

We demonstrate our flexible SCD model and posterior simulation methods using quotation data for three equities traded on the Toronto Stock Exchange: the Potash Corporation (POT), the Royal Bank of Canada (RY) and the Toronto-Dominion Bank (TD).[15] The data comes from the TickData database, where it is freely available.[16] The sample covers one month of activity for a total of 21 consecutive trading days, from March 3 to 31, 2014. We follow standard practice and remove observations that are obviously erroneous or with aberrant prices, following the procedure detailed in Barndorff-Nielsen et al. (2009). Price durations are then computed using the mid-price (bid+ask)/2 of the the most recently posted quote in each second. We construct three duration series for each equity by selecting observations involving an absolute cumulative mid-price change of at least \$0.01, \$0.02 and \$0.03, respectively. A change of \$0.01 correspond to the smallest possible price change (or *ticksize*). We select the observations recorded between 10:00 am and 4:00 pm for the analysis. Descriptive statistics of the price durations are reported in Table 2.2. For each series, we report the number of durations, followed by the sample mean, standard deviation, kurtosis and coefficient of variation. The last four columns report the 10%, 50% and 90% quantiles and the maximum value.

Model specification and prior distributions

For each series, we report results for models with a fixed value of J, the number of terms in the normalized density, going from $J = 1$ to $J = 10$. In all models, the diurnal pattern is specified as a

[15] The Potash Corporation merged with Agrium on January 1, 2018 to form Nutrien Ltd (NTR).
[16] https://www.tickdata.com/equity-data/

Table 2.2: Descriptive statistics of price durations from March 3 to 31, 2014 for various price thresholds.

Symbol	ΔPrice	Obs.	Mean	Std.	Kurt.	C.V.	$\mathcal{Q}_{0.1}$	$\mathcal{Q}_{0.5}$	$\mathcal{Q}_{0.9}$	Max.
POT	\$ 0.01	8964	50.5	76.3	44.8	1.51	3	24	128	1510
	\$ 0.02	3186	141.1	194.2	22.2	1.38	11	78	337	2317
	\$ 0.03	2061	217.4	290.9	23.6	1.34	21	123	490	3548
RY	\$ 0.01	14412	31.5	47.8	45.5	1.52	2	15	78	888
	\$ 0.02	4852	93.3	132.2	25.7	1.42	8	49	224	1686
	\$ 0.03	2813	160.8	215.2	26.0	1.34	15	90	397	2630
TD	\$ 0.01	6153	73.5	120.3	81.0	1.64	3	34	180	2892
	\$ 0.02	2168	207.4	298.6	22.4	1.44	16	106	483	3141
	\$ 0.03	1292	346.7	466.5	17.5	1.35	33	194	824	3861

The first column gives the number of price durations followed by the mean, standard deviation, kurtosis, and coefficient of variation. The last four columns give the 10%, 50% and 90% quantiles and the maximum value.

B-spline function defined on knots set at 30 minutes intervals between 10:00 am to 4:00 pm, giving an expansion with $L = 15$ piecewise polynomials.

We use the same prior distribution for each specification. For each series, we set the hyperparameters of the OU process at $(\bar{\mu}_\sigma, \bar{h}_\sigma) = (-0.7, 4)$ and $(\bar{\mu}_\rho, \bar{h}_\rho) = (\bar{y}/10, 4)$, where $\bar{y}$ is the sample average. This gives an autocorrelation of approximately 0.9 for durations equal to $\bar{y}$ at prior mean. The smoothing hyperparameters for the vector of coefficients describing a diurnal pattern are fixed at $(\bar{s}_\tau, \bar{\nu}_\tau) = (1, 200)$. We fix the hyperparameters for the mean, the first order autocorrelation and the innovation variance of the daily component dynamics at $(\bar{\mu}_\mu, \bar{h}_\mu) = (\hat{a}/(1-\hat{b}), 1)$, $(\bar{\mu}_\phi, \bar{h}_\phi) = (\hat{b}, 200)$, and $(\bar{s}_\xi, \bar{\nu}_\xi) = (0.04/\hat{s}, 0.04/\hat{s}^2)$, respectively, where $(\hat{a}, \hat{b})$ and $\hat{s}^2$ are the OLS and the unbiased variance estimate for the linear regression $\log \bar{y}_d = a + b \log \bar{y}_{d-1} + e_d$, where $\bar{y}_d$ is the average duration on day d. The location and the concentration parameter for the mixture weight indexing the normalized density are set respectively at $\bar{\beta} = (1/J, \ldots, 1/J)$ and $\bar{M} = 2J$, which centres the prior distribution for $p_\epsilon(\cdot)$ around the exponential distribution and implies moderate shrinkage towards it. For the values of J considered, this choice of hyperparameters has been shown to provide a good balance in terms of number of parameters and flexibility; it allows for approximately 20 % of the mass to be moved around the first-order model (Chen et al. 2014).

Model comparison

An important property of Bernstein polynomials is that polynomials of lower degree are nested within polynomials of larger degree; all Bernstein polynomials of degree $1, 2, \ldots, J-1$ are included in the Bernstein polynomial of degree J. This implies that after some J, we expect little change in the shape of the normalized density.

Figure 2.2 shows the normalized density functions $p_\epsilon(\cdot)$ obtained using $J = 1, 3, 6$ and 9 terms at the posterior mean of the β coefficients for the POT equity. The left figure is for the price durations constructed using a threshold of \$ 0.01; the one in the middle, a threshold of \$ 0.02; the one on the right, a threshold of \$ 0.03. We see that the normalized density varies much more with J at smaller values than it does at larger values and the overall variation remains small between each densities. For the series with the \$ 0.02 and \$ 0.03 thresholds, the mean densities obtained using 6 and 9 terms are very similar; for the series with the \$ 0.01 threshold, the same observation applies, but for the densities obtained using 3 and 6. However, in this last case, the density obtained using 9 terms has a different shape that the ones obtained with lower values of J, having a small mode close to zero rather than being monotonically decreasing. This preliminary graphical analysis suggest that a specification with $J \leq 9$ should be able to capture well the conditional duration density for each series. The same comments apply for the RY and TD equity and for this reason we do not show similar illustrations for the other series.

Table 2.3 reports the posterior probabilities of each specification computed based on the prior assumption that the specifications are equally probable. In this case, the posterior probabilities are proportional to the marginal likelihood. We see that a model with a normalized density with 3 terms is the preferred specification for most series. A model using a more flexible normalized density with 4 terms is preferred only once. The posterior probabilities for models using a normalized density with more than 4 terms are all zeros. Hence, specifications with 2, 3 and 4 terms are able to capture well the shape of the conditional duration density.

Parameter estimates

Table 2.4 shows posterior summaries for the specifications having the highest posterior probability for each equity and price threshold. The upper panel is for the POT equity; the one in the middle for the RY equity; the bottom one for the TD equity. The three groups of columns are for the

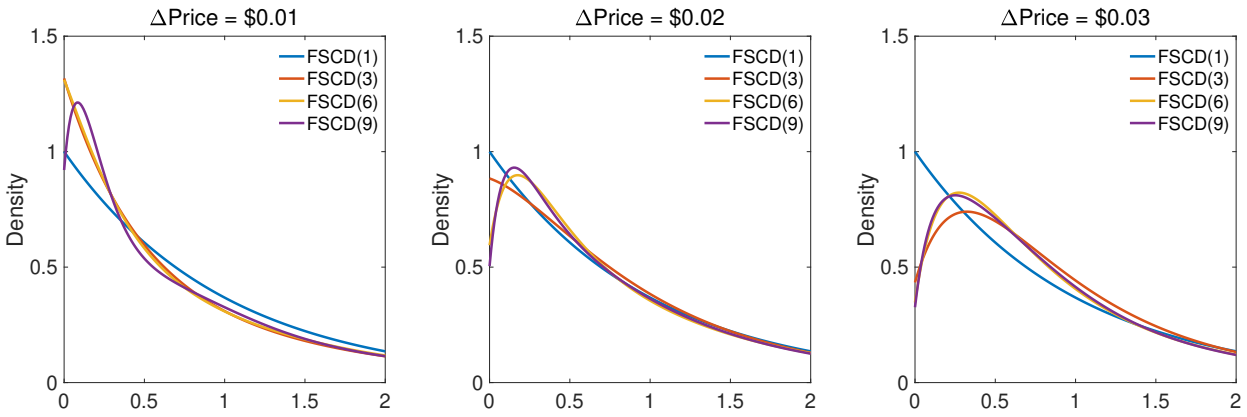

Figure 2.2: Normalized density functions at the posterior mean of β obtained for the POT equity. The figure on the left is for the price durations constructed using a threshold of $ 0.01; the one in the middle, a threshold $ 0.02; the one on the right, a threshold of $ 0.03.

Table 2.3: Posterior probability for the different values of J, the number of term in the normalized density.

Symbol	ΔPrice	FSCD(1)	FSCD(2)	FSCD(3)	FSCD(4)	FSCD(5)
POT	$ 0.01	0.000	0.096	0.904	0.000	0.000
	$ 0.02	0.115	0.004	0.881	0.000	0.000
	$ 0.03	0.000	0.011	0.871	0.118	0.000
RY	$ 0.01	0.000	0.000	1.000	0.000	0.000
	$ 0.02	0.012	0.234	0.742	0.012	0.000
	$ 0.03	0.000	0.000	0.015	0.985	0.000
TD	$ 0.01	0.000	0.509	0.491	0.000	0.000
	$ 0.02	0.177	0.777	0.044	0.002	0.000
	$ 0.03	0.000	0.030	0.612	0.356	0.001

price duration series constructed using thresholds of $ 0.01, $ 0.02 and $ 0.03, respectively. For each parameter, we report the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the posterior mean. Defined in Geweke (1989), the relative numerical efficiency is a variance ratio that quantifies the numerical precision of the sample mean of a ergodic process, relative to that of a (hypothetical) iid sample. RNE times sample size gives the size of an iid sample with the same numerical standard error. Numerical standard errors are computed using the overlapping batch mean method (Flegal & Jones 2010). The posterior samples consist of 50,000 retained draws recorded after a burn-in period of 15,000 draws.

It is well known that it is difficult to sample efficiently the persistence and variance parameters of latent states in non-Gaussian state space models. We obtain a numerical efficiency for ρ and σ that is considerably higher than that reported for the analogous parameters using the block sampling

Table 2.4: Posterior mean and standard deviation of various parameters for the POT, RY and TD equity of price durations and the flexible SCD model with the preferred specification.

Panel A: Potash Corporation (POT)

	ΔPrice = \$ 0.01			ΔPrice = \$ 0.02			ΔPrice = \$ 0.03		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
σ	0.5137	0.0237	0.0684	0.5286	0.0337	0.0566	0.5846	0.0326	0.0682
$\rho^{\dagger}$	0.0827	0.0139	0.0399	0.0553	0.0151	0.0407	0.1051	0.0257	0.0625
μ	4.0163	0.0564	0.2999	5.0563	0.0669	0.2678	5.4795	0.0675	0.2431
ϕ	-0.2325	0.0684	0.8064	-0.2127	0.0692	0.8440	-0.1179	0.0689	0.8146
ξ	0.2512	0.0412	0.4373	0.2757	0.0543	0.2926	0.2699	0.0500	0.3286
β_1	0.5834	0.0273	0.0657	0.3413	0.0450	0.0726	0.1413	0.0290	0.1095
β_2	0.2056	0.0182	0.0958	0.4581	0.0339	0.0944	0.5983	0.0615	0.1328
β_3	0.2111	0.0275	0.0762	0.2006	0.0534	0.1086	0.2604	0.0720	0.1330

Panel B: Royal Bank of Canada (RY)

	ΔPrice = \$ 0.01			ΔPrice = \$ 0.02			ΔPrice = \$ 0.03		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
σ	0.4332	0.0196	0.0569	0.4938	0.0338	0.0504	0.4929	0.0431	0.0478
$\rho^{\dagger}$	0.0942	0.0183	0.0315	0.0432	0.0113	0.0384	0.0289	0.0096	0.0325
μ	3.4943	0.0745	0.5441	4.6030	0.1074	0.5861	5.1323	0.1111	0.4302
ϕ	0.3103	0.0692	0.6948	0.3317	0.0696	0.7774	0.3566	0.0697	0.7171
ξ	0.2202	0.0309	0.5143	0.3007	0.0595	0.2547	0.2865	0.0588	0.2716
β_1	0.6302	0.0250	0.0786	0.3803	0.0411	0.0912	0.1407	0.0266	0.0896
β_2	0.2066	0.0156	0.0654	0.4591	0.0251	0.0993	0.4946	0.0434	0.0922
β_3	0.1632	0.0179	0.1156	0.1606	0.0395	0.1323	0.1438	0.0392	0.0911
β_4							0.2209	0.0454	0.0823

Panel C: Toronto Dominion Bank (TD)

	ΔPrice = \$ 0.01			ΔPrice = \$ 0.02			ΔPrice = \$ 0.03		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
σ	0.6286	0.0232	0.0848	0.5164	0.0471	0.0780	0.5809	0.0464	0.0569
$\rho^{\dagger}$	0.1933	0.0394	0.0477	0.0253	0.0085	0.0531	0.0541	0.0204	0.0454
μ	4.3807	0.1150	0.5970	5.3884	0.1496	0.5300	5.9410	0.1496	0.4490
ϕ	0.6293	0.0690	0.5601	0.6413	0.0701	0.5657	0.5841	0.0700	0.6043
ξ	0.1923	0.0221	0.5592	0.2320	0.0394	0.3790	0.2722	0.0557	0.3183
β_1	0.7472	0.0351	0.1329	0.7880	0.1047	0.0682	0.1654	0.0395	0.0713
β_2	0.2528	0.0351	0.1329	0.2120	0.1047	0.0682	0.5651	0.0713	0.1174
β_3							0.2694	0.0838	0.1061

†: Rescale $\times 60$ (mean reversion in minute).

Note. This table gives the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the mean, based on 50,000 posterior draws recorded after a burn-in period of 15,000 draws. The three groups of these columns are for the price duration series constructed using a threshold of \$ 0.01, \$ 0.02 and \$ 0.03.

Table 2.5: Posterior quantiles and moments of the half-life $t_{1/2}$, measured in seconds, of the latent intensity state OU process $x_d(t)$.

Symbol	ΔPrice	Mean	Std	$\mathcal{Q}_{0.01}$	$\mathcal{Q}_{0.25}$	$\mathcal{Q}_{0.50}$	$\mathcal{Q}_{0.75}$	$\mathcal{Q}_{0.99}$
POT	$0.01	517	87	346	455	510	571	756
	$0.02	809	229	421	646	775	936	1503
	$0.03	419	103	232	347	407	479	716
RY	$0.01	459	92	293	393	449	512	723
	$0.02	1029	271	547	838	994	1177	1850
	$0.03	1601	551	718	1216	1502	1876	3337
TD	$0.01	224	46	138	192	219	251	351
	$0.02	1538	545	571	1160	1468	1839	3146
	$0.03	874	317	334	651	822	1038	1867

method in Strickland et al. (2006).[17] The lowest numerical efficiency we obtain for ρ is 2-17 times higher, and for σ is 4-36 times higher than the numerical efficiencies they obtained for analogous parameters, using parametric duration distribution. Posterior sample means of the mixture weights have numerical efficiency ranging from 0.0654 to 0.1330, regardless of the number of terms used in the normalized duration density. We see that the numerical efficiency appears to be slightly lower for the first mixture weights, where we observe the most variation in the shape of the normalized density. Nevertheless, the numerical efficiencies in general are quite good as far as methods for non-linear non-Gaussian state space models go.

Table 2.5 reports posterior quantiles and moments of the half-life $t_{1/2}$, measured in seconds, of the OU process $x_d(t)$: the length of time it takes for $x_d(t)$ and $x_d(t + t_{1/2})$ to have a correlation of $1/2$ between them. This quantity is more easily interpretable than the mean reversion parameter ρ. The results reported are those obtained with the specifications having the highest posterior probability (see Table 2.3). Persistence of the latent intensity process is fairly high for the POT and RY equities, especially for the $0.01 threshold, where the half-life is more around 10 and 14 times, respectively, the average duration of each series (see Table 2.2). Still for the POT and RY equities, the persistence decreases with the increase of the price threshold used to construct the durations; for POT, the half-life is around 6 and 2 times the average duration, for the $0.02 and $0.03 thresholds, respectively. For RY, the variation is less severe; the half-life is around 11 and 10 times the average duration. However, we do not observe this pattern for the TD equity where the

[17] The latent intensity state in their analysis follows a Gaussian AR(1) process with fixed autocorrelation and fixed innovation variance rather than an OU process sampled at irregular intervals.

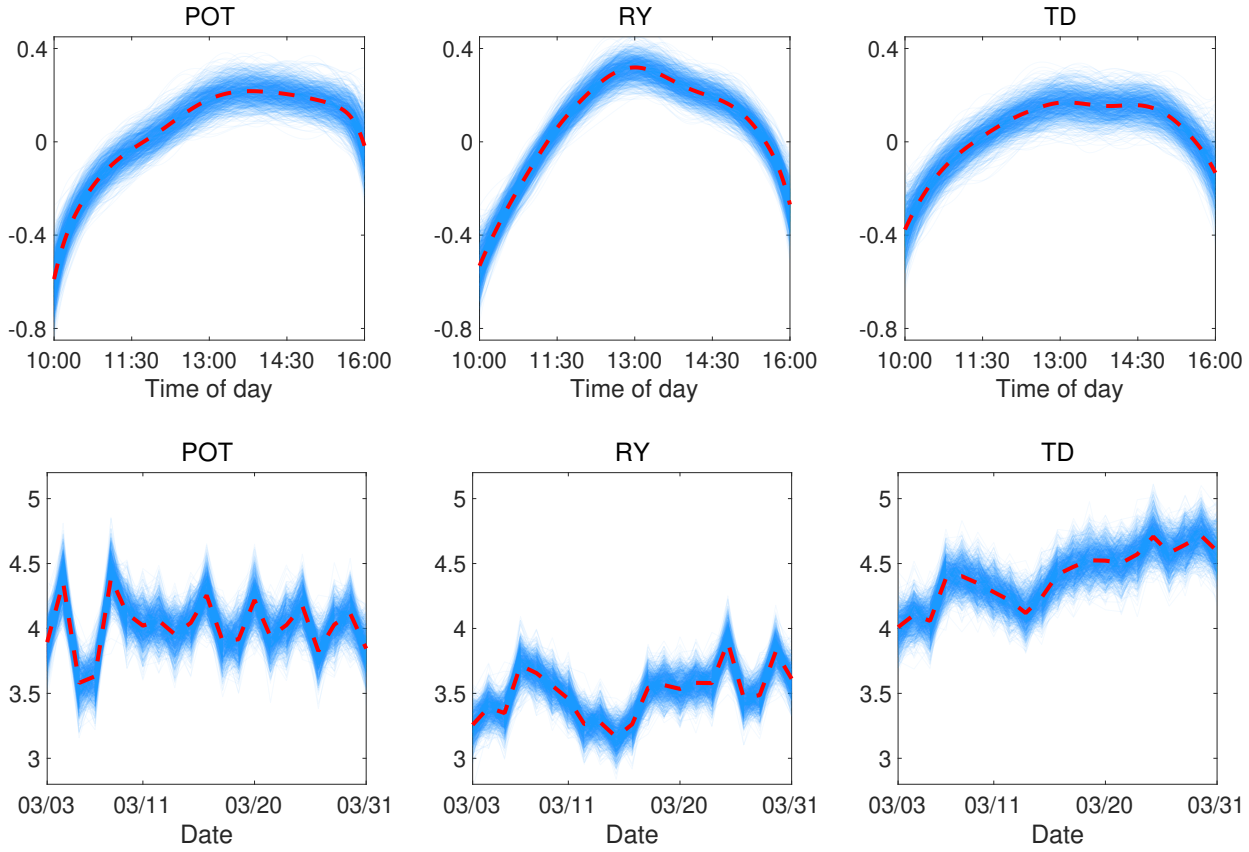

Figure 2.3: Diurnal patterns and vector of daily components at the posterior mean (dashed line) and for 1000 posterior draws (solid lines) obtained for the price durations series constructed using a price threshold of $ 0.01.

persistence is higher for the $ 0.02 series than for the $ 0.01 and $ 0.03 series. Overall, there is a fair degree of posterior uncertainty about $t_{1/2}$.

Figure 2.3 shows the posterior mean (dashed line) and 1000 posterior draws (solid lines) of the diurnal pattern $m(t)$ in the upper panel, and of the vector of daily components ψ in the lower panel, for the duration series constructed using a price threshold of $ 0.01. The figures on the left are for the POT equity; the ones in the middle, the RY equity; the ones on the left, the TD equity. The results reported are those obtained with the specification having the highest posterior probability (see Table 2.3). In each case, we obtain the usual inverted U-shaped diurnal pattern found in most studies, with more variation in price change near opening and closing times. However, we see that the intensity is considerably higher near opening than near closing times, and more so for POT. The diurnal pattern for RY is more pronounced, indicating more predictable variation in price change intensity. For the POT and TD equities, the daily intensity is more stable and smoothly varying around the sample mean; for the TD equity, we see a slow decrease of intensity in price change

between the first and last day of the sample. However, the number of days in the sample doesn't allow us to conclude much more about the long-run behaviour of inter-daily variation. The posterior variation in the diurnal patterns and the daily components is fairly small compared to the variation in average intensity through the day and between days, respectively. The posterior distribution for δ is not very sensitive to the price threshold used. Although changing in level, the daily components ψ are also very similar for each equity between the different duration series. For this reason, we do not show illustrations similar to Figure 2.3 for the other series.

Hazard functions

Figure 2.4 illustrates the posterior distribution of the unconditional hazard function for the normalized duration density obtained for the POT equity and models with $J = 3$ terms. The dashed line shows the hazard function at the posterior mean. Solid lines give 1000 posterior draws. For details on the calculation of the unconditional hazard function, see Bauwens & Veredas (2004). There is not much posterior uncertainty in the hazard functions. Most of the uncertainty in the hazard rate is observed for the $\$\,0.02$ series (in the middle) for values near zero. In this case, the hazard function is non-monotonic with high posterior probability. Although posterior precision is fairly high, there is a non-negligible posterior probability that the hazard function is monotonic and decreasing. For the $\$\,0.01$ series (on the left), the hazard has a smoothly decreasing shape and for the $\$\,0.03$ series (on the right), it is initially increasing and non-monotonic. The smooth variation in the hazard function, together with the possibility of it being non-monotonic, would have been impossible to capture with commonly used parametric conditional duration densities. We observe very similar hazard shapes for the other equity and for this reason the same figures for the RY and the TD series are not reported.

2.5 Concluding remarks

Conditional duration distributions in ACD and SCD models should be flexible, in the sense of being able to approximate any distribution on $[0, \infty)$. At the same time, it should be possible to approximate realistic distributions with a small number of terms.

Appealing to queueing theory, we argued that a good first-order model for durations is an exponential distribution. We introduced a normalized conditional distribution for financial durations that is

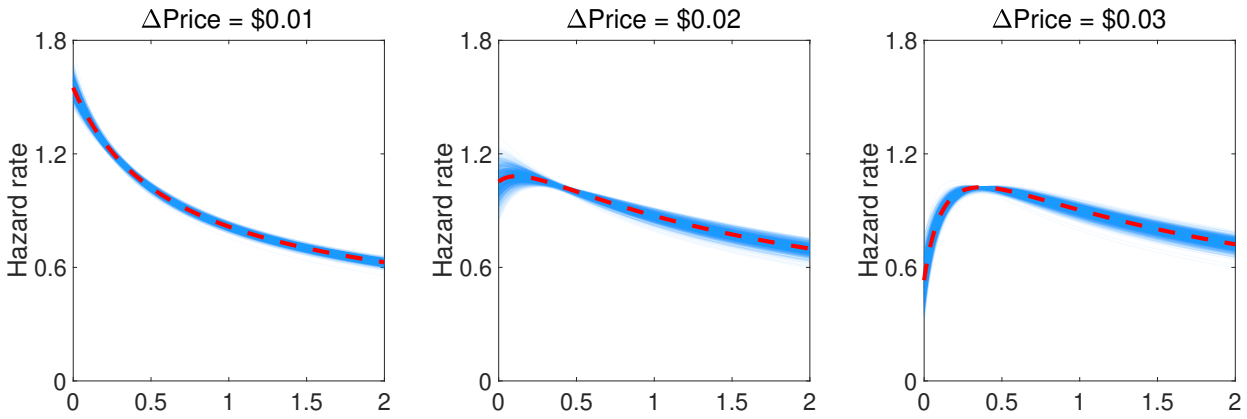

Figure 2.4: Unconditional hazard functions for the normalized duration density at the posterior mean (dashed lines) and for 1000 posterior draws (solid lines) obtained for the POT equity and models with $J = 3$ terms. The figure on the left is for the price durations constructed using a threshold of $\$\,0.01$; the one in the middle, a threshold $\$\,0.02$; the one on the right, a threshold of $\$\,0.03$.

flexible, and also expressible as a perturbed exponential distribution. This allows us to shrink towards the exponential distribution.

We introduce posterior simulation methods for Bayesian inference in SCD models featuring our proposed conditional duration density. Due in part to efficient draws of the latent trade intensity state, and despite the flexible distribution, numerical efficiency of posterior simulation is considerably better than that of previous studies where duration distributions are parametric.

We illustrated our methods using quotation data for three equities traded on the Toronto Stock Exchange. For each equity, we analyse the durations built for transactions involving a cumulative mid-price change of at least $\$\,0.01$, $\$\,0.02$ and $\$\,0.03$. For each series, we found that a model featuring our proposed conditional duration density with more than four terms has a posterior probability of nearly zero, and that, out of the nine series analysed, a conditional density with three terms provides the best fit 6 times. Despite the small number of terms, this conditional normalized density is capable of matching distribution functions with various hazard shapes. In each case, the ratio of the maximum of the hazard function to the minimum is less than three. This suggests that our proposed conditional duration distribution is empirically relevant and avoids problems associated with other distributions used in the literature.

Chapter 3

A censored stochastic conditional duration model for discrete trade durations with zero inflation*

3.1 Introduction

Duration models for financial transactions describe the irregular timing of trades, or other events such as price changes. They are useful because trading intensity is one measure of market liquidity. And unlike count models, duration models use all the information in trading times. They also shed light on market microstructure phenomena.

Some stylized facts about trade 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 our definition of clusters in our model will be precise,

*This chapter is co-authored with my advisor William J. McCausland.

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.

Modelers have to confront this fact that some trades are related to others and that related trades are nearly simultaneous. We will argue that widely accepted rules for aggregating seemingly related trades into clusters, together with unsuitable parametric duration distributions, mislead inference in important ways. The most commonly used rule treats two trades as related if they are executed, to within recording precision, simultaneously. In this paper, as in many studies, we consider transaction times truncated to the second, and in this context we will call it the *same-second* aggregation rule.[1] The rule is usually correct, case by case, but in a reasonably sized sample from a liquid market, it is likely that many pairs of unrelated trades occur within the same clock second by happenstance.

Our paper makes two main contributions. First, we propose a new model for discrete-valued trade durations, defined as the difference between consecutive trading times, truncated to the second. Traditional durations models are based on continuous distribution. Financial durations are, however, inherently discrete, whether they are recorded with lower or finer precision. As discreteness is a property of the selected truncation precision, and not the underlying process, we treat the data as a censored continuous-time process. The second contribution is to propose a new way to classify trades as being part or a cluster or not. Rather than apply a deterministic rule prior to inference, we put into our model the uncertainty about whether each duration is a cluster duration—one between related trades—or a regular duration—one between unrelated trades—and develop a mixture model, featuring a latent indicator variable for every duration. Not surprisingly, the posterior probability that a discrete $0s$ duration (zero second) is due to coincidence, and therefore regular, is always low. But it is never zero, and the posterior mean of the number of these coincidences is much larger than its posterior standard deviation. We find that the conditional hazard function of the underlying continuous-time process for regular durations varies smoothly near zero, in contrast to the abrupt changes found in most studies, which we argue are artifacts arising from neglecting the discrete nature of data and classifying *all* discrete $0s$ durations as cluster durations.

We now illustrate in detail the problem with treating all discrete $0s$ durations as cluster durations,

[1] Many datasets, especially more recent ones, feature millisecond or finer recording precision. The issues raised in this paper apply widely, but since the unfortunate consequences of deterministic aggregation rules are qualitatively different when recording precision is finer, we leave this case for future research.

a practice that goes back to the seminal paper of Engle & Russell (1998). To do so, we will refer to the sample used in the illustrative empirical example presented in Section 3.4.2 that contains transaction data for one week on trading of March 2014 for three equities traded on the Toronto Stock Exchange; the Potash Corporation (POT), the Royal Bank of Canada (RY) and the Toronto-Dominion Bank (TD). Table 3.3 shows descriptive statistics of the constructed duration series. The important information related to the current discussion is given by the last six columns, which give the percentage of discrete durations recorded as $0s$, $1s$, $2s$, $3s$, $4s$, and $5s$. As we can see, the values from $1s$ to $5s$ vary smoothly, but far more discrete durations are recorded as $0s$ than as $1s$.

This zero inflation clearly needs to be addressed. As mention earlier, it arises because of nearly simultaneous related trades. These often occur when a market order is matched against, and filled with, several limit orders on the opposite side of the market, which is commonly referred to as a split-transaction. It may also occur if many traders submit limit orders to be 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. Again, we emphasize that we do not try to distinguish different causes of related trades in our model but generally refer to the resulting durations as cluster durations.

In fact, the amount of zero inflation is greater than it might appear: the duration between two trades will be recorded as a discrete $0s$ duration if the second trade occurs during the *remainder* of the same clock second as the first, but as a discrete $1s$ duration if it occurs at *any* time during the next clock second. Extrapolating percentages from $1s$ to $5s$ back to $0s$ and dividing by two gives us a rough idea of the percentage of all durations that are both regular and recorded as $0s$: about 7% for the RY and the TD series and about 5% for the POT series. Of course, these represent a much larger percentage of regular durations, around 22% and 14%, respectively. While it is likely that a large majority of discrete $0s$ durations are cluster durations, it would be very surprising if all were.

The same-second aggregation rule amounts to removing all discrete $0s$ durations from the sample. The result is a truncated sample of regular durations. The fact that the truncation is at both the mode and lower bound of the distribution is particularly unfortunate. Trading intensity is understated, and the understatement varies with trading intensity: when it is high, there are more short regular durations and more spurious aggregation of unrelated trades. Moreover, distinguishing cluster and regular transactions is not always easy, as regular durations will often be $0s$ durations at times of intense trading activity. For instance, for the data considered in this paper, we observe that over 50 transactions were recorded within the same second on many occasions. It would be

surprising if every time this happens all transactions were linked to a single cluster of transactions, as implied by the same-second aggregation rule, and especially when that many transactions are recorded in a one second interval.

Another aggregation rule, proposed by Grammig & Wellner (2002), aggregates any sequence of transactions within the same clock second where prices are non-decreasing or non-increasing. We will call this the *GW* aggregation rule. Figure 3.1 shows histograms of durations that are classified as regular by this rule. Not enough of the $0s$ durations are classified as regular for compatibility with a smooth density near zero. This is not surprising, as many pairs of unrelated trades will feature price changes of the same sign by coincidence.

While the spurious aggregation of unrelated trades is our main concern with these rules, we also question whether even error-free aggregation would be desirable. If a market order matches nine limit orders, five times as many traders are getting their orders filled than if it matches a single limit order. Treating these cases as equivalent may understate liquidity, as perceived by traders.

Discreteness of duration data is traditionally ignored, and financial durations are usually modelled using multiplicative error models featuring continuous distribution where the scale of the conditional duration distribution depends on the history of the process (Engle 2002).[2] The two general class of multiplicative error models are the autoregressive conditional duration model (ACD, Engle & Russell 1998), where the scale depends deterministically on past durations, and the stochastic conditional duration model (SCD, Bauwens & Veredas 2004), where the scale is a latent stochastic process.[3] It was only recently that Blasques et al. (2020) proposed an alternative ACD model based on parametric discrete distributions, with and without zero inflation.[4] In this paper, we also adopt a discrete approach, but based on a censored distribution related to an underlying continuous-time process agreeing with the usual multiplicative error assumption. Its additional innovative feature is to distinguish between regular and cluster durations by using a model where only the distribution of the regular durations depends on a latent trade intensity indicator.

Models for regular durations were initially specified using an exponential distribution, but it was quickly concluded that models featuring a Weibull or a gamma distribution provide a better fit for

[2] In high-frequency data analysis, transaction price dynamics are modelled using discrete distributions (see, e.g. Russell & Engle 2005).

[3] For surveys on the analysis of high-frequency financial durations, see Pacurar (2008), Hautsch (2012) and Bhogal & Variyam (2019).

[4] Yatigammana et al. (2019) used a dynamic mixture of two continuous distributions in an ACD model for modelling trade durations with high density close to zero. However, their approach cannot indicate whether close to zero durations are related or not.

various dataset (Engle & Russell 1998, Bauwens & Giot 2000, Bauwens & Veredas 2004, Feng et al. 2004, Strickland et al. 2006, Men et al. 2015). However, these conclusions favouring the Weibull or gamma come largely from studies in which $0s$ durations are removed from the sample, following the same-second aggregation rule, and without adjusting for the discreteness of the observed data.[5] Moreover, except for the special case where they reduce to an exponential, which has a constant hazard function, the hazard function is either zero at a duration of zero and increasing; or infinite at a duration of zero and decreasing. We argued in Chapter 2 that such extreme variation of the hazard function near zero is implausible for regular, unrelated, durations. A reasonable simple model for transaction times in financial market, where there is a large number of independent potential traders, is the Poisson process where durations between events are exponentially distributed. After conditioning on relevant predictors measuring the variation of trading intensity, the underlying distribution of regular durations should not be too far from an exponential, due to the fact that a regular duration is, by our definition, the time interval between unrelated trades. We found that this is indeed the case for the three duration series analyzed in an illustrative empirical example.

The rest of the paper is organized as follows. We describe our censored SCD model in Section 3.2 and our methods for posterior simulation in Section 3.3. In Section 3.4, we conduct an artificial data experiment to test for the correctness of our posterior simulator and then illustrate our methods in an application featuring three equities traded on the Toronto Stock Exchange. We conclude in Section 3.5.

3.2 A Censored Stochastic Conditional Duration Model

Our model builds on the stochastic conditional duration model of Bauwens & Veredas (2004). It differs in two important ways. First, it is for discrete-valued data. Transaction data are inherently discrete, whether they are recorded with low or high precision. Our model takes into account this discreteness by treating the data as a censored continuous-time process. More precisely, it is designed for transaction times truncated to the second. Second, it has a second state variable, one which indicates which discrete durations are cluster durations and which are regular. Since this replaces the usual practice of aggregating trades into clusters before analysis, our model is intended for unaggregated data. For the purpose of comparison, we also define a special case of our model

[5] Examples of continuous-time models incorporating the zero observations are the dynamic zero augmented model of Hautsch et al. (2014) and the dynamic censored model of Harvey & Ito (2020).

where cluster durations have probability zero, for samples in which durations have been aggregated into clusters. We will call this special case *regular-duration* model and the general model the *all-duration* model. Placing classification of trades within the model makes it possible to infer that the number of regular durations that happen to be a $0s$ duration, while small, is not zero. We will see that this overcomes the problems arising from the spurious aggregation of unrelated trades.

3.2.1 The Data Generating Process

We observe transaction times over D trading days in the interval $[t_{\text{open}}, t_{\text{close}}]$, where t_{open} and t_{close} are the opening and closing times. All times of day are measured in seconds after midnight and recorded with limited precision. For each day $d = 1, \ldots, D$, denote the sequence of *observed* transaction times by $\tilde{t}_{d,0}, \tilde{t}_{d,1}, \ldots, \tilde{t}_{d,n_d}$ and construct the *censored* transaction times $t_{d,i} \equiv \lfloor \tilde{t}_{d,i} \rfloor$ and the corresponding discrete durations $y_{d,i} \equiv t_{d,i} - t_{d,i-1}$, $i = 1, \ldots, n_d$, where $\lfloor \cdot \rfloor$ is the floor operator that truncated the transaction times to second precision.

The conditional distribution of each $y_{d,i}$ depends on the values $s_{d,i} \in \{0, 1\}$ and $x_{d,i} \in \mathbb{R}$, at time $t_{d,i-1}$, of two latent states. The state $s_{d,i}$ is a mixture component indicator: $s_{d,i} = 0$ indicates that $y_{d,i}$ is a cluster duration; $s_{d,i} = 1$, a regular duration. The state $x_{d,i}$ gives the trading intensity at $t_{d,i-1}$, but only if the duration is regular. The conditional distribution of $y_{d,i}$, given $s_{d,i}$ and $x_{d,i}$, is

$$p(y_{d,i} \mid s_{d,i}, x_{d,i}) = \begin{cases} p_0(y_{d,i}) & s_{d,i} = 0, \\ p_1(y_{d,i} \mid x_{d,i}) & s_{d,i} = 1, \end{cases} \tag{3.1}$$

where $p_0(\cdot)$ and $p_1(\cdot)$ are probability mass functions that we will be specified in Section 3.2.2.

At each day d, the indicator process $s_{d,i}$ is first-order Markov and stationary, with

$$\Pr[s_{d,i+1} = l \mid s_{d,i} = k] = \xi_{kl} \quad \text{and} \quad \Pr[s_{d,1} = k] = \xi_k, \tag{3.2}$$

where $\xi_0 \equiv (1 - \xi_{11})/(2 - \xi_{00} - \xi_{11})$ and $\xi_1 \equiv (1 - \xi_{00})/(2 - \xi_{00} - \xi_{11})$ by stationarity. The marginal probabilities (ξ_0, ξ_1) give, respectively, the proportion of cluster and regular durations in the population. The trading intensity process $x_d(t)$, $t \in [t_{\text{open}}, t_{\text{close}}]$ is the sum of a deterministic function $m(t)$ describing a diurnal pattern and a zero-mean OU process. Sampling the $x_d(t)$ process at all irregularly spaced *censored* transaction times gives the values $x_{d,i} \equiv x_d(t_{d,i-1})$, which follow

an inhomogeneous first order autoregressive discrete process:

$$x_{d,i+1} \,|\, x_{d,i}, t_{d,i}, t_{d,i-1} \sim \mathcal{N}\left(m(t_{d,i}) + e^{-\rho y_{d,i}}(x_{d,i} - m(t_{d,i-1})), \sigma^2(1 - e^{-2\rho y_{d,i}})\right), \tag{3.3}$$

$$x_{d,1} \,|\, t_{d,0} \sim \mathcal{N}\left(m(t_{d,0}), \sigma^2\right), \tag{3.4}$$

where $\rho > 0$ is the mean reversion parameter, and σ is the marginal standard deviation parameter of the OU process. Remark that $y_{d,i} = 0$ gives $x_{d,i+1} = x_{d,i}$; thus, our censored SCD model implies a reduced number of latent intensity states equal to the number of nonzero discrete durations. The deterministic function $m(t)$ describing a diurnal pattern is a cubic B-spline linear basis expansion with respect to a set of equally spaced knots (Eilers & Marx 1996):

$$m(t) = \sum_{l=1}^{L} \delta_l B_l(t), \tag{3.5}$$

where $B_l(\cdot)$ denotes the l-th B-spline basis function and δ_l its coefficient. The first and last knots have multiplicity 4 and the rest have multiplicity 1, which makes $m(t_{\text{open}}) = \delta_1$ and $m(t_{\text{close}}) = \delta_L$. See de Boor (1978) or Dierckx (1993) for more on B-splines.

3.2.2 The Conditional Duration Process

In accordance with the GW-aggregation rule proposed by Grammig & Wellner (2002), we assume that discrete durations of more than $1s$ are regular by definition and specify the probability mass function of cluster durations as a Bernoulli distribution,

$$p_0(y_{d,i}) = \pi^{1-y_{d,i}}(1-\pi)^{y_{d,i}}\mathbf{1}\{y_{d,i} \leq 1\}, \tag{3.6}$$

where π gives the conditional probability of a cluster duration being equal to zero. Allowing for cluster durations of $1s$ takes into account that censoring transaction times to the second implies that transactions recorded within the same clock second will result in a duration of $1s$ rather than $0s$ with positive probability. However, this probability should be small given our definition of cluster durations and $p_0(\cdot)$ should concentrate most of its probability on zero. The prior distribution for π must be chosen accordingly.

Unlike cluster durations, regular durations depend on market conditions, as captured by the intensity state $x_{d,i}$. Our specification of $p_1(\cdot)$ follows from censoring a continuous-time process conforms to

the usual multiplicative error structure of SCD models (Bauwens & Veredas 2004), which gives durations as $y_{d,i} = \exp(x_{d,i})\epsilon_{d,i}$, where $\epsilon_{d,i}$ is the contemporaneous value of a positive iid process with $E[\epsilon_{d,i}] = 1$ and independent of $x_{d,i}$. We denote by $p_\epsilon(\cdot)$ the normalized density function of $\epsilon_{d,i}$. Thus, the conditional probability mass function of regular duration $y_{d,i}$ is obtained by integrating over the range of possible uncensored values:

$$p_1(y_{d,i} \,|\, x_{d,i}) = \frac{1}{2} \int_{y_{d,i}-1}^{y_{d,i}+1} e^{-x_{d,i}} p_\epsilon(y e^{-x_{d,i}})\, dy. \tag{3.7}$$

For the probability of $y_{d,i} = 0$, the integral is from zero to one. Given $s_{d,i} = 1$, the quantity $e^{-x_{d,i}}$ is the conditional mean of $y_{d,i}$, when viewed as a continuous-time process.

For reasons mentioned in the introduction, we select a *perturbed* exponential density, as specified in Chapter 2, as normalized density; $p_\epsilon(\epsilon) \equiv f(\epsilon) g\left(F(\epsilon)\right)$, where $F(\cdot)$ the cdf of an exponential distribution, with density $f(\cdot)$, and $g(\cdot)$ is a J'th order Bernstein density; a J-component mixture of beta densities, each with integer-valued shape parameters adding to $J+1$. This gives

$$p_\epsilon(\epsilon) = \lambda \exp(-\lambda\epsilon) \sum_{j=1}^{J} \beta_j \,\mathrm{Beta}\left(1 - \exp(-\lambda\epsilon) \,|\, j, J-j+1\right), \tag{3.8}$$

where λ is a scale parameter, $\sum_{j=1}^{J} \beta_j = 1$, $\beta \equiv (\beta_1, \ldots, \beta_J) \geq 0$, and $\mathrm{Beta}(z \,|\, a, b)$ denotes the beta density with shape parameters a and b, for $a, b > 0$. Notice that if $\beta = (1/J, \ldots, 1/J)$ then the Bernstein density $g(\cdot)$ is uniform and there is no perturbation. In this paper, we treat the order J of the Bernstein density as fixed, not as a parameter to estimate, and will compare results over different values of J. Notice that we can write the above expression as

$$p_\epsilon(\epsilon) = \sum_{j=1}^{J} \beta_j \left[\sum_{k=1}^{j} a_{j,k} \lambda_{j,k} \exp(-\lambda_{j,k}\epsilon) \right],$$

where $\lambda_{j,k} = (J-j+k)\lambda$ and

$$a_{j,k} = \binom{j-1}{k-1} \frac{\Gamma(J+1)}{\Gamma(j)\Gamma(J-j+1)} \frac{(-1)^{k+1}}{J-j+k}.$$

This allows to easily evaluate the integral in (3.7) to obtain a convenient closed-form expression as

a mixture distribution for the probability mass function of regular durations;

$$p_1(y_{d,i} \mid x_{d,i}) = \frac{1}{2} \sum_{j=1}^{J} \beta_j \sum_{k=1}^{j} \alpha_{j,k} \Big[\exp\left(-\lambda_{j,k} e^{-x_{d,i}} (y_{d,i} - 1)\right) - \exp\left(-\lambda_{j,k} e^{-x_{d,i}} (y_{d,i} + 1)\right) \Big]. \quad (3.9)$$

We will exploit this expression for posterior inference. Also, remark that the scale normalization condition $\mathrm{E}[\epsilon] = 1$ gives $\lambda = \sum_{j=1}^{J} \beta_j \sum_{k=1}^{j} a_{j,k}/(J - j + k)$, which we use to substitute out λ from the expression for $p_\epsilon(\cdot)$, freeing us from having to impose restriction on β.

3.2.3 Prior Distributions

To complete the model, we describe a prior distribution for the parameter vector of our censored SCD model $\Theta = (\sigma, \rho, \delta, \beta, \xi, \pi)$. All parameters are *a priori* independent. In the following, the overbar notation $\bar{\cdot}$ is used to denote prior hyperparameters. The different values used for the simulation exercise and the empirical example are available in Table 3.1.

We specify log-Normal prior distributions for the marginal standard deviation σ and the mean reversion parameter ρ of the OU process: $\log \sigma \sim \mathcal{N}(\bar{\mu}_\sigma, \bar{h}_\sigma^{-1})$ and $\log \rho \sim \mathcal{N}(\bar{\mu}_\rho, \bar{h}_\rho^{-1})$. This eliminates the need for parameter restriction. We induce a Gaussian prior for the vector δ of coefficients defining the diurnal pattern by specifying the following Gaussian prior on an invertible linear function of δ:

$$\begin{bmatrix} u' \\ \nabla \end{bmatrix} \delta \sim \mathcal{N}\left(\begin{bmatrix} \bar{\mu}_\delta \\ 0_{L-1} \end{bmatrix}, \begin{bmatrix} \bar{h}_\delta & 0'_{L-1} \\ 0_{L-1} & \tau I_{L-1} \end{bmatrix}^{-1} \right), \quad (3.10)$$

where u and ∇ are the mean and first backward difference operators, 0_{L-1} is an $(L-1) \times 1$ vector of zeros, I_{L-1} is the $(L-1) \times (L-1)$ identity matrix, $\bar{\mu}$ is a scalar location parameter and τ and $\bar{h}$ are scalar precision parameters. Thus, first differences $\delta_l - \delta_{l-1}$ are iid $\mathcal{N}(0, \tau^{-1})$ and the arithmetic mean is independent of the first differences, with $u'\delta \sim \mathcal{N}(\bar{\mu}_\delta, \bar{h}_\delta^{-1})$. This is the same as the random walk prior proposed by Lang & Brezger (2004), except that they use $u = (1, 0, \ldots, 0)$. Both priors favour smoothness and are agnostic with respect to the signs of derivatives; since the derivatives of B-splines are linear combinations of the first differences $\delta_l - \delta_{l-1}$, we can interpret τ as a smoothing parameter for the diurnal pattern. Higher values of τ imply more shrinkage towards a flat diurnal pattern. Unlike the prior proposed by Lang & Brezger (2004), where the prior variances of the δ_l increase with l, our prior has the symmetry property that $(\delta_1, \ldots, \delta_L)$ and $(\delta_L, \ldots, \delta_1)$ have the same distribution. Following common practices, we will estimate τ, and specify its prior as the following

scaled chi-square: $\bar{s}\tau \sim \chi^2(\bar{\nu})$. Equation (3.10) induces the conditional prior $\delta \,|\, \tau \sim N(\bar{\delta}, \bar{H}^{-1})$, where $\bar{\delta} = \bar{\mu}_\delta Lu$ and $\bar{H} = \bar{h}_\delta uu' + \tau \nabla'\nabla$.

For the mixture weights indexing the normalized density $p_\epsilon(\cdot)$, we specify a Dirichlet distribution: $\beta \sim \text{Dirichlet}(\bar{M}\bar{\beta})$, parametrized in term of a mean parameter $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_J) > 0$, $\sum \bar{\beta}_j = 1$, and a concentration parameter $\bar{M} > 0$. If $\bar{\beta} = (1/J, \ldots, 1/J)$, the prior mean of β corresponds to $g(\cdot)$ being uniform on $[0, 1]$ and therefore $p_\epsilon(\cdot)$ being an exponential density. The diagonal elements of the Markov transition matrix ξ of the latent indicator process are independent and beta distributed: $\xi_{kk} \sim \text{Beta}(\bar{a}_k, \bar{b}_k)$, $k = 0, 1$. The conditional probability that a cluster duration been $0s$ duration is beta, with $\pi \sim \text{Beta}(\bar{a}_\pi, \bar{b}_\pi)$.

3.2.4 Joint Density

We conclude the exposition of the model by giving the joint density of all parameters, latent variables and observations, making explicit all conditional independence relationships. We refer to the model as the all-duration censored SCD model (A-CSCD). Let s, x and y be the flat vectors of all indicators, states and durations, respectively. Then the joint density is

$$p(\sigma, \rho, \tau, \delta, \beta, \xi, \pi, s, x, y) = p(\sigma)\, p(\rho)\, p(\tau)\, p(\delta \,|\, \tau)\, p(\beta)\, p(\xi)\, p(\pi)$$
$$\prod_{d=1}^{D} \prod_{i=1}^{n_d} p(s_{d,i} \,|\, s_{d,i-1}, \xi) p(x_{d,i} \,|\, x_{d,i-1}, t_{d,i-1}, t_{d,i-2}, \sigma, \rho, \delta) p(y_{d,i} \,|\, s_{d,i}, x_{d,i}, \pi, \beta). \quad (3.11)$$

The densities for the initial values $s_{d,1}$ and $x_{d,1}$ on each day d are $p(s_{d,1} \,|\, s_{d,0}, \xi) \equiv p(s_{d,1} \,|\, \xi)$ and $p(x_{d,1} \,|\, x_{d,0}, t_{d,0}, t_{d,-1}, \sigma, \rho, \delta) \equiv p(x_{d,1} \,|\, t_{d,0}, \sigma, \delta)$.

We also define the regular-duration model (R-CSCD), the special case of our model where all durations are regular, which is suitable for use with data where trades are aggregated into clusters. The joint density of the regular-duration model is obtained by removing factors related to ξ, π, and s in the above expressions.

3.3 Bayesian Inference

Here we describe posterior simulation methods for Bayesian inference in our censored SCD model. We do this for the all-duration model; methods for the regular-duration model require only straight-

forward modifications. We use Markov chain Monte Carlo (MCMC) to sample the joint posterior distribution of parameters, latent indicators and state variables. For the all-duration model, there are six Gibbs blocks, updating (σ, ρ, x), (τ, δ), β, s, ξ and π. For the regular-duration model, we only require the first three blocks. The first three blocks are similar to the posterior simulation methods proposed in Chapter 2, but adapted for censored durations; the last three implement standard methods for Bayesian inference in Markov-Switching models (cf. Frühwrith-Schnatter 2006). We now describe each of the Gibbs blocks in turn.

Drawing from $p(\sigma, \rho, x \,|\, \delta, s, y)$

We update (σ, ρ, x) in two sub-blocks. First, we perform data augmentation and draw component indicators $z_{d,i} \in \{1, \ldots, J\}$ for each (d, i) where $y_{d,i}$ is a regular duration; that is $s_{d,i} = 1$. The indicators are conditionally independent, with probability mass function given, up to a multiplicative factor, by

$$\Pr[z_{d,i} = j \,|\, \beta, s, x, y] \propto \beta_j \sum_{k=1}^{j} \alpha_{j,k} \Big[\exp\left(-\lambda_{j,k} e^{-x_{d,i}} (y_{d,i} - 1)\right) - \exp\left(-\lambda_{j,k} e^{-x_{d,i}} (y_{d,i} + 1)\right) \Big], \tag{3.12}$$

for $j = 1, \ldots, J$. We then update (σ, ρ, x) given all $z_{d,i}$ as a single block using a Metropolis-Hastings step that consist of a joint accept-reject for the state variables x and the parameters describing their dynamics (σ, ρ). Our proposal rely on the HESSIAN method (McCausland 2012), a generic procedure to draw the complete state sequence as a single block that required, as only model specific elements, routines to evaluate $\log p(y_{d,i} \,|\, s_{d,i}, x_{d,i}, z_{d,i})$, and its first five derivatives with respect to $x_{d,i}$, at a given point. We compute exact values of this conditional density function and its derivatives using automatic routines to combine evaluations of derivatives of primitive functions using Faá di Bruno's rule. Details are provided in Appendix C.1.

Our joint proposal (σ^*, ρ^*, x^*) consists of a Gaussian random walk proposal for $(\log \sigma^*, \log \rho^*)$ followed by a conditional proposal $x^* \,|\, \sigma^*, \rho^*$ drawn from the proposal density $q(x \,|\, \sigma^*, \rho^*, \delta, z, s, y)$ constructed using the HESSIAN method. We accept the triple (σ^*, ρ^*, x^*) with probability

$$\min \left\{ 1, \frac{p(x^*, y \,|\, \sigma^*, \rho^*, \delta, \pi, z, s) p(\sigma^*, \rho^*)}{p(x, y \,|\, \sigma, \rho, \delta, \pi, z, s) p(\sigma, \rho)} \times \frac{q(x \,|\, \sigma, \rho, \delta, z, s, y)}{q(x^* \,|\, \sigma^*, \rho^*, \delta, z, s, y)} \right\}.$$

Recall that we specified a prior for $(\log \sigma, \log \rho)$, thus, the Hastings-ratio does not need to be adjusted by the Jacobian of the logarithmic transformation. During the burn-in period, we tune the

covariance of the random walk proposal to track an acceptance probability of 0.352—the optimal acceptance rate for 2 dimensional parameter given in Gelman et al. (1996)—by following the approach described in Vihola (2012). We stop the adaptation after the burn-in period to ensure that our posterior simulator is Markov.

Drawing from $p(\tau, \delta \,|\, \sigma, \rho, x, y)$

We update (τ, δ) using two sub-blocks. We first draw τ from its conditional posterior distribution: $\bar{\bar{s}}\tau \mid \delta \sim \chi^2(\bar{\bar{\nu}})$, where $\bar{\bar{s}} = \bar{s} + \delta'\nabla'\nabla\delta$ and $\bar{\bar{\nu}} = \bar{\nu} + L - 1$. We then draw δ from its conditional posterior distribution: $\delta \,|\, \tau, \sigma, \rho, x, y \sim N(\bar{\bar{\delta}}, \bar{\bar{H}}^{-1})$, where the posterior precision is $\bar{\bar{H}} = \bar{H} + W'W$, the posterior mean is $\bar{\bar{\delta}} = \bar{\bar{H}}^{-1}(\bar{H}\bar{\delta} + W'v)$. The vector v and matrix W come from writing the state equation (3.3) as $v \sim \mathcal{N}(W\delta, I_N)$ where v and W are organized in blocks

$$\begin{bmatrix} v_1 \\ \vdots \\ v_D \end{bmatrix}, \quad \begin{bmatrix} W_{11} & \dots & W_{1L} \\ \vdots & \ddots & \vdots \\ W_{D1} & \dots & W_{DL} \end{bmatrix},$$

with, for $d = 1, \dots, D$ and $l = 1, \dots, L$,

$$v_d = \begin{bmatrix} x_{d,1}/\sigma \\ (x_{d,2} - \exp(-\rho y_{d,1}) x_{d,1}) / \sqrt{\sigma^2(1 - \exp(-2\rho y_{d,1}))} \\ \vdots \\ (x_{d,n_d} - \exp(-\rho y_{d,n_{d-1}}) x_{d,n_{d-1}}) / \sqrt{\sigma^2(1 - \exp(-2\rho y_{d,n_{d-1}}))} \end{bmatrix}$$

and

$$W_{dl} = \begin{bmatrix} B_l(t_{d,1})/\sigma \\ (B_l(t_{d,2}) - \exp(-\rho y_{d,1}) B_l(t_{d,1})) / \sqrt{\sigma^2(1 - \exp(-2\rho y_{d,1}))} \\ \vdots \\ (B_l(t_{d,n_d}) - \exp(-\rho y_{d,n_{d-1}}) B_l(t_{d,n_d-1})) / \sqrt{\sigma^2(1 - \exp(-2\rho y_{d,n_{d-1}}))} \end{bmatrix}.$$

Drawing from $p(\beta \,|\, z, s, x, y)$

We update β using a Metropolis-Hastings step. For the purposes of drawing proposals, we use the logistic transformation $\vartheta(\beta) = (\log(\beta_1/\beta_J), \dots, \log(\beta_{J-1}/\beta_J))$. Our proposal for β^* consists of

Gaussian random walk proposal for ϑ^*, which is transformed back to β^* using the inverse transformation $\beta(\vartheta) = (1 + \sum_{j=1}^{J-1} \exp(\vartheta_j))^{-1}(\exp(\vartheta_1), \ldots, \exp(\vartheta_{J-1}), 1)$. We accept β^* with probability

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

where the second term comes form the Jacobian of the logistic transformation. The covariance of the random walk proposal is again tuned during the burn-in period following the adaptive scheme described in Vihola (2012) by targeting the optimal acceptance probability corresponding to the dimension of $\vartheta(\beta)$ given in Gelman et al. (1996).

Drawing from $p(s \mid \xi, \pi, \beta, x, y)$

Latent regime indicators are updated via a single-move, or one-at-a-time, sampler, which is considerably faster than drawing states as a block, as no filtering is required (Frühwrith-Schnatter 2006). Since there is little posterior autocorrelation (most probabilities are close to zero or one, regardless of past and future values) there is little loss of numerical precision. The relative conditional probabilities of drawing $s_{d,i} = 0$ and $s_{d,i} = 1$, given $s_{-(d,i)}$, the rest of the indicators, are given by

$$\Pr(s_{d,i} = j \mid \xi, \pi, \beta, s_{-(d,i)}, x, y) \propto \xi_{s_{d,i-1},j} p(y_{d,i} \mid \pi, \beta, s_{d,i} = j, x_{d,i}) \xi_{j,s_{d,i+1}}, \tag{3.13}$$

for $d = 1, \ldots, D$ and $i = 1, \ldots, n_d$. The first and the last conditional transition probabilities are understood to be $\xi_{j,s_{d,0}} \equiv \xi_j$ and $\xi_{j,s_{d,n_d+1}} \equiv 1$.

Drawing from $p(\xi \mid s)$

Given the latent regime indicators, we update the transition probabilities using a Metropolis-Hasting step. The prior for (ξ_{00}, ξ_{11}) (ξ_{00} and ξ_{11} are independent betas) is nearly conditionally conjugate, but since the first indicator of each day comes from the marginal distribution, not exactly so. The target density can be written as

$$p(\xi \mid s) \propto \left[\prod_{d=1}^{D} \frac{1 - \xi_{11}^{1-s_{d,1}} - \xi_{00}^{s_{d,1}}}{2 - \xi_{00} - \xi_{11}} \right] \xi_{00}^{N_{00}+\bar{a}_0-1}(1-\xi_{00})^{N_{01}+\bar{b}_0-1} \xi_{11}^{N_{11}+\bar{a}_1-1}(1-\xi_{11})^{N_{10}+\bar{b}_1-1}, \tag{3.14}$$

Table 3.1: Prior hyper-parameter values used in the Getting it right experiment (GIR) and in the empirical example using transaction data from the Toronto Stock Exchange (TSX).

Hyper-parameters	GIR	TSX
$(\bar{\mu}_\sigma, \bar{h}_\sigma)$	$(-1.0, 10^3)$	$(-0.9, 4)$
$(\bar{\mu}_\rho, \bar{h}_\rho)$	$(-2.3, 10^3)$	$(-\log 10\bar{y}_+, 4)$
$(\bar{\mu}_\delta, \bar{h}_\delta)$	$(1.5, 250)$	$(\log \bar{y}_+, 1)$
$(\bar{s}, \bar{\nu})$	$(10, 500)$	$(1, 200)$
$(\bar{M}; \bar{\beta})$	$(250; 0.4, 0.3, 0.3)$	$(5J; 1/J, \ldots, 1/J)$
$(\bar{a}_0, \bar{b}_0)$	$(100, 400)$	$(5, 2)$
$(\bar{a}_1, \bar{b}_1)$	$(400, 100)$	$(2, 5)$
$(\bar{a}_\pi, \bar{b}_\pi)$	$(200, 50)$	$(100, 3)$

Note. $\bar{y}_+$ is the sample average of nonzero durations.

where $N_{lk} = \sum_{d,i} \mathbf{1}\{s_{d,i} = l, s_{d,i+1} = k\}$ is the number of transitions from l to k over all D days. We draw a proposal using two independent beta distributions, $\xi_{00}^* \mid s \sim \text{Beta}(N_{00} + \bar{a}_0, N_{01} + \bar{b}_0)$ and $\xi_{11}^* \mid s \sim \text{Beta}(N_{11} + \bar{a}_1, N_{10} + \bar{b}_1)$. This would be an exact draw from the conditional posterior if we were conditioning on the first indicator in each day. We correct for the approximation by accepting the proposal with probability

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

Drawing from $p(\pi \mid s, y)$

Given the latent regime indicators, a Beta prior for π is conditionally conjugate and the conditional posterior distribution is $\pi \mid s, y \sim \text{Beta}(\bar{a}_\pi + N_0, \bar{b}_\pi + N_1)$, where $N_k = \sum_{d,i} \mathbf{1}\{s_{d,i} = 0, y_{d,i} = k\}$.

3.4 Results

In this section, we first report results of tests of program correctness meant to verify the conceptual validity and the implementation in code of our posterior simulator. We then provide an illustrative empirical example for the use of our censored SCD model using transaction data for three equities traded on the Toronto Stock Exchange.

3.4.1 Getting it right (GIR)

The correctness tests described here are similar to those described in Geweke (2004). Those tests take the form of tests for the equality of two stationary distributions and have power against a wide array of conceptual and programming errors. The idea is to generate an MCMC sample for $\{\Theta^{(m)}, s^{(m)}, x^{(m)}, y^{(m)}\}_{m=1}^{M}$ from the joint distribution of Θ, s, x and y using the simulation methods that will be used for posterior simulation to update (Θ, s, x) conditional on y and an additional Gibbs block to update y conditional on (Θ, s, x).[6] Under the hypothesis that the posterior simulation methods are correctly implemented and conceptually valid, the marginal distribution of the parameter vector $\{\Theta^{(m)}\}_{m=1}^{M}$ is identical to its known prior distribution.

As mentioned in Section 2.4.1 of the second chapter, some difficulties with the underlying continuous-time OU process $x(t)$ make this exercise impractical without minor adjustments. Hence, we make a similar modification to the latent state process (3.3) and replace it for the GIR simulation by a homogeneous autoregressive process, parametrized in a way that resembles to the sampled OU:

$$x_{d,i+1} \,|\, x_{d,i} \sim \mathcal{N}\left(m(t_i) + e^{-\rho}(x_{d,i} - m(t_{i-1})), \sigma^2(1 - e^{-2\rho})\right). \tag{3.15}$$

This implies that the number of state variables equals the number of observations, that is, unlike in the empirical application, we draw one state variable $x_{d,i}$ by observation even if $y_{d,i} = 0$. This make sure the number of states does not change from iteration to iteration. We also turn off the adaptive schemes implemented to tuned the random walk proposals in the Gibbs block updating (σ, ρ, x) and β and instead fix the proposal covariances to the prior covariance of the parameter vector in question.

We generate a simulation sample of size $M = 10^6$ for analysis from a model with a normalized distribution with $J = 3$ components and a diurnal pattern defined by a B-spline function with two knots, t_{open} and t_{close}, giving an expansion with $L = 4$ cubic polynomials. To avoid trades in simulations occurring after t_{close}, where the diurnal pattern is undefined, we choose the sample size n for the durations and prior distributions such that the probability that the last transaction of the day occurs after t_{close} is extremely small. For the sample of durations, we fix $D = 1$ and choose a sample size of $n = 20$ observations, and set the length of the trading session at 600 seconds. Values of the prior hyper-parameters are shown in Table 3.1. The tighter prior distribution, and

[6] Details of this additional Gibbs block is provided in Appendix C.2.

Table 3.2: Difference between prior and simulation sample first and second moments in the Getting it right experiment.

θ	$\mathrm{E}[\theta] - \bar{\theta}$	$\hat{\sigma}_{\mathrm{nse}}$	t-stat	$\mathrm{E}[\theta^2] - \bar{\theta}^2$	$\hat{\sigma}_{\mathrm{nse}}$	t-stat
$\log \sigma$	-0.000035	0.000099	-0.357	0.000070	0.000197	0.353
$\log \rho$	0.000048	0.000102	0.469	-0.000222	0.000468	-0.474
μ	-0.000308	0.000239	-1.290	-0.000870	0.000716	-1.215
τ	-0.019117	0.010328	-1.851	-1.936401	1.037820	-1.866
β_1	-0.000028	0.000316	-0.090	-0.000012	0.000252	-0.049
β_2	-0.000333	0.000290	-1.150	-0.000186	0.000174	-1.065
β_3	0.000362	0.000281	1.286	0.000221	0.000170	1.301
ξ_{00}	-0.000052	0.000055	-0.949	-0.000019	0.000022	-0.845
ξ_{11}	-0.000094	0.000065	-1.444	-0.000150	0.000104	-1.445
π	0.000012	0.000078	0.159	0.000022	0.000124	0.175

much smaller number of observations, compared with our empirical examples, ensure high numerical precision with a moderate amount of computation.

Table 3.2 shows the results for the comparison between the prior and the simulation sample first and second moment. In each case, the first column gives, for selected parameters θ, the difference between the prior moment $\mathrm{E}[\theta^k]$ and the simulation sample moment $\bar{\theta}^k$ $(k = 1, 2)$; the second, the numerical standard error (i.e. the simulation standard deviation quantifying error in finite simulation) of the sample moment; and the third, the t-statistic for the test of the hypothesis that the moment of the parameter, with respect to the stationary distribution of the Markov chain, equals the (known) prior moment. Numerical standard errors are computed using the overlapping batch means method (see. Flegal & Jones 2010). Rather than reporting results for all elements of δ, we report results for their mean $\mu \equiv u'\delta$ and the smoothing parameter τ. Sample moments are close to the true prior moments, relative to the numerical standard error. Each of the hypotheses is a necessary condition for the correctness of our simulation methods. The results fail to cast doubt on this correctness: no hypothesis (out of 20) is rejected at the 10% level.

3.4.2 An empirical example

Data

We demonstrate our censored SCD model and posterior simulation methods using transaction data from March 17 to March 21, 2014 for three equities traded on the Toronto Stock Exchange (TSX):

Table 3.3: Descriptive statistics of the cleaned sample from March 17 to 21, 2014.

	Trades	Mean	Std.	Max.	C.V.	0s %	1s %	2s %	3s %	4s %	5s %
POT	25203	4.28	14.21	434	3.32	75.5	3.5	2.0	1.6	1.3	1.1
RY	31343	3.44	9.81	236	2.85	71.3	4.6	2.6	2.2	1.9	1.6
TD	32359	3.34	10.38	266	3.11	73.5	4.3	2.5	2.1	1.7	1.5

The first column gives the number of transactions. The next four columns give the mean, standard deviation, maximum value and coefficient of variation of durations. The last six columns give the percentage of trade durations equal to $0s$, $1s$, $2s$, $3s$, $4s$ and $5s$.

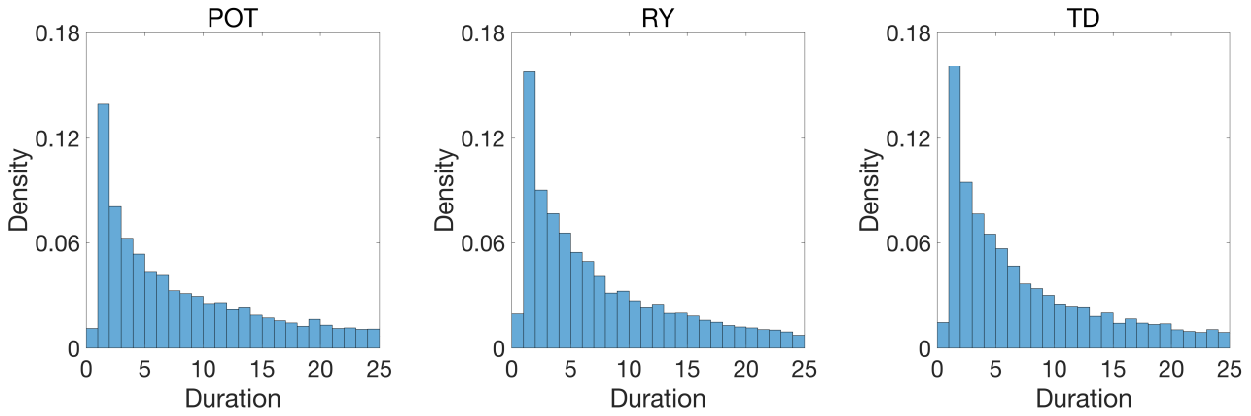

Figure 3.1: Histograms of regular durations, as classified by the GW aggregation rule, from $0s$ to $25s$. Bins are aligned with clock seconds.

the Potash Corporation (POT), the Royal Bank of Canada (RY) and the Toronto-Dominion Bank (TD).[7] The data comes from the TickData database, where it is freely available.[8] For each equity, we remove observations that result from atypical market conditions or have aberrant prices, following the procedure detailed in Brownlees & Gallo (2006). We select for analysis the transactions recorded between 10:00 am and 4:00 pm. Trading times are truncated to the second; in other data sets, transaction times can be recorded to higher levels of precision. Hence, in this case, the *observed* transaction times are equivalent to the *censored* transaction times. Descriptive statistics of the cleaned data are reported in Table 3.3. For each duration series, we report the number of observations, followed by the sample mean, standard deviation, maximum, and coefficient of variation. An exponential has a coefficient of variation of 1, so the empirical distributions are considerably overdispersed relative to the exponential. The second part of the table reports the percentage of discrete durations recorded as $0s$, $1s$, $2s$, $3s$, $4s$, and $5s$.

[7]The Potash Corporation merged with Agrium on January 1, 2018 to form Nutrien Ltd (NTR).
[8]https://www.tickdata.com/equity-data/

Model specification and prior distribution

We analyze the full samples directly using the all-duration model (A-CSCD). To compare our probabilistic approach to classifying durations as regular or cluster durations with deterministic aggregation rules, we also construct subsamples of the durations classified as regular by the GW aggregation rule. We will refer to these subsamples as GW-filtered subsamples, and analyze them using the regular-duration model (R-CSCD). Figure 3.1 shows histograms of durations that are classified as regular by this rule. We remind the reader that the same-second aggregation rule of common practice is cruder than the GW aggregation rule; we would expect distortions arising from the same-second aggregation rule to be even more serious than those reported here. For both specifications (all-duration and regular-duration) we report results for five models, each with a fixed value of J, the number of terms in the normalized conditional density; those values are $J = 1, 2, 3, 4$ and 5. In all models, the diurnal pattern is specified as a B-spline function defined on knots set on each half-hour, giving an expansion with $L = 15$ piecewise polynomials.

Values of the prior hyper-parameters are shown in the third column of Table 3.1, where $\bar{y}_+$ denotes the sample average of the discrete nonzero durations. Overall, we select fairly diffuse prior distributions for the parameters of the latent intensity state $x_{d,i}$, the mean of the coefficient of the B-spline function $m(t)$ describing diurnal patterns and the transition probabilities of the latent indicator $s_{d,i}$. The hyperparameter $\bar{\mu}_\rho$ for the mean reversion parameter ρ gives for each series an autocorrelation of approximately 0.9 at the prior mean for $\bar{y}_+$. The values of $(\bar{s}, \bar{\nu})$ give a more diffuse prior distribution for the smoothing parameter τ than the one suggested by Lang & Brezger (2004). For reasons given in the introduction, we center the prior distribution for the normalized conditional duration density $p_\epsilon(\cdot)$ around the exponential distribution and set the concentration parameter for moderate shrinkage toward it: most of posterior variances of the β_j coefficients are less than a quarter of the prior variances. Finally, we select an informative prior distribution for π to strongly favour cluster durations of $0s$ over $1s$.

Parameter estimates

Tables 3.4 , 3.5 and 3.6 show the results for the POT, RY and TD series, respectively. For each selected parameter, we report the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the posterior mean. Defined in Geweke (1989), the relative numerical efficiency is a variance ratio that quantifies the numerical precision of the sample mean of a ergodic process,

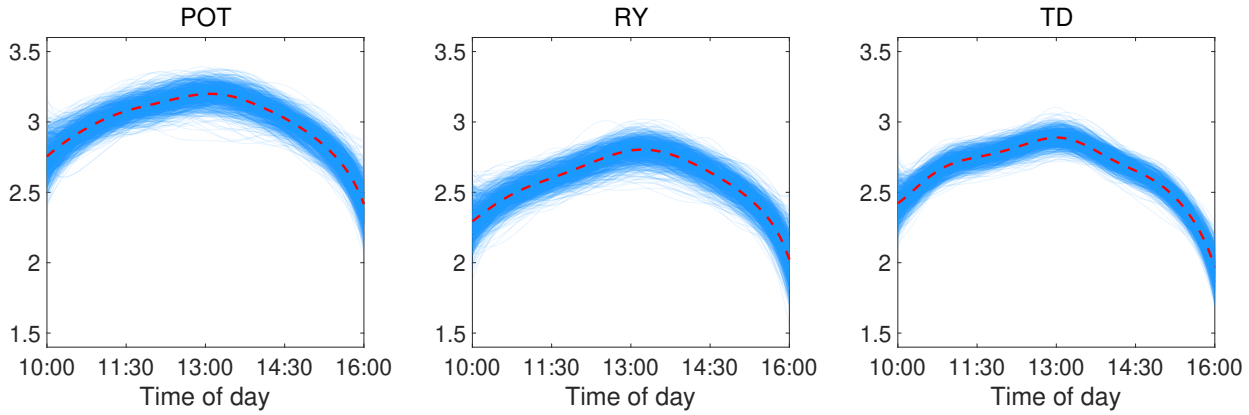

Figure 3.2: Diurnal pattern at the posterior mean (dashed line) and for 1000 posteriors draws (solid lines) obtained for the full samples and the all-duration model with $J = 1$. The figure on the left is for the POT series; the one on centre, for the RY series ; the one on the left, the TD series.

relative to that of a (hypothetical) iid sample. RNE times sample size gives the size of an iid sample with the same numerical standard error. Numerical standard errors are computed using the overlapping batch mean method (Flegal & Jones 2010). The posterior samples consist of 25,000 retained draws recorded after a burn-in period of 10,000 draws.

Figure 3.2 shows the posterior mean (dashed line) and 1000 posterior draws of the diurnal pattern $m(t)$, for the full samples and the all-duration model with $J = 1$ specification. In each case, we obtain the usual inverted U-shaped diurnal pattern found in most studies, with more trading intensity near the opening and closing times. Although different in level, the diurnal patterns are similar in shape. The posterior variation in the diurnal patterns is fairly small compared to the variation in average intensity through the day. The posterior distribution for δ is not very sensitive to J or to the choice between deterministic and probabilistic classification. For this reason, we do not show illustrations similar to Figure 3.2 for other specifications.

Model comparison

In this subsection, we first compare the normalized densities for the all-duration and the regular-duration models. We then compare parameters of the latent intensity state between models and specifications. Finally, we propose a criteria to select the most appropriate number of terms in the normalized density based on in-sample predictive performance.

Figure 3.3 shows normalized density functions $p_\epsilon(\cdot)$ at the posterior mean of the β coefficients. Upper panels are for the all-duration models and the full samples; lower panels for the regular-

Table 3.4: Posterior mean and standard deviation of various parameters, for the GW-filtered POT subsample and the regular-duration model; and for the full POT sample and the all-duration model.

Panel A: All-duration models

	A-CSCD(1)			A-CSCD(2)			A-CSCD(3)			A-CSCD(4)			A-CSCD(5)		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
$\rho^\dagger$	0.253	7.04e-02	0.077	0.128	3.15e-02	0.068	0.108	2.69e-02	0.051	0.106	2.71e-02	0.037	0.106	2.61e-02	0.033
σ	0.539	2.79e-02	0.107	0.527	3.51e-02	0.077	0.523	3.68e-02	0.073	0.522	3.79e-02	0.081	0.522	3.70e-02	0.068
β_1				0.732	3.23e-02	0.166	0.529	2.32e-02	0.078	0.407	3.08e-02	0.041	0.328	2.84e-02	0.030
β_2				0.268	3.23e-02	0.166	0.216	2.54e-02	0.060	0.223	3.42e-02	0.046	0.202	3.49e-02	0.039
β_3							0.256	2.82e-02	0.061	0.183	2.91e-02	0.060	0.177	3.34e-02	0.054
β_4										0.186	2.98e-02	0.054	0.138	2.44e-02	0.056
β_5													0.155	2.62e-02	0.045
ξ_{00}	0.796	3.19e-03	0.542	0.794	3.25e-03	0.538	0.791	3.31e-03	0.354	0.791	3.35e-03	0.254	0.791	3.44e-03	0.139
ξ_{11}	0.364	6.80e-03	0.618	0.367	6.78e-03	0.683	0.370	6.85e-03	0.510	0.371	6.88e-03	0.392	0.371	6.95e-03	0.271
π	0.983	1.47e-03	0.237	0.984	1.45e-03	0.179	0.987	1.55e-03	0.105	0.987	1.62e-03	0.072	0.987	1.72e-03	0.047

Panel B: Regular-duration models

	R-CSCD(1)			R-CSCD(2)			R-CSCD(3)			R-CSCD(4)			R-CSCD(5)		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
$\rho^\dagger$	0.329	8.81e-02	0.085	0.142	3.35e-02	0.067	0.109	2.70e-02	0.052	0.109	2.69e-02	0.044	0.111	2.71e-02	0.039
σ	0.532	2.51e-02	0.107	0.508	3.28e-02	0.081	0.492	3.59e-02	0.078	0.493	3.47e-02	0.070	0.493	3.47e-02	0.061
β_1				0.740	2.95e-02	0.149	0.538	2.28e-02	0.081	0.397	2.90e-02	0.056	0.307	2.41e-02	0.042
β_2				0.260	2.95e-02	0.149	0.223	2.14e-02	0.081	0.251	3.02e-02	0.064	0.238	2.99e-02	0.051
β_3							0.238	2.60e-02	0.077	0.164	2.83e-02	0.055	0.158	3.24e-02	0.060
β_4										0.188	2.94e-02	0.052	0.136	2.49e-02	0.053
β_5													0.160	2.64e-02	0.037

†: Rescale $\times 60$ (mean reversion for duration in minutes).

Note. The table gives the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the posterior mean, based on 25,000 posterior draws recorded after a burn-in period of 10,000 draws. The five groups of these three columns are for one to five component normalized duration densities.

Table 3.5: Posterior mean and standard deviation of various parameters, for the GW-filtered RY subsample and the regular-duration model; and for the full RY sample and the all-duration model.

Panel A: All-duration models

	A-CSCD(1)			A-CSCD(2)			A-CSCD(3)			A-CSCD(4)			A-CSCD(5)		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
$\rho^{\dagger}$	0.191	5.25e-02	0.062	0.110	2.72e-02	0.054	0.103	2.63e-02	0.037	0.104	2.73e-02	0.028	0.105	2.71e-02	0.024
σ	0.503	2.82e-02	0.083	0.505	3.62e-02	0.074	0.500	3.61e-02	0.073	0.502	3.56e-02	0.076	0.501	3.61e-02	0.068
β_1				0.673	3.01e-02	0.111	0.469	1.82e-02	0.060	0.357	2.79e-02	0.028	0.291	2.39e-02	0.026
β_2				0.327	3.01e-02	0.111	0.244	2.59e-02	0.045	0.229	3.46e-02	0.027	0.186	3.46e-02	0.028
β_3							0.287	2.59e-02	0.057	0.199	2.98e-02	0.047	0.197	3.32e-02	0.044
β_4										0.215	2.82e-02	0.058	0.144	2.59e-02	0.039
β_5													0.182	2.59e-02	0.037
ξ_{00}	0.757	3.20e-03	0.392	0.754	3.29e-03	0.395	0.751	3.49e-03	0.165	0.750	3.64e-03	0.076	0.750	3.69e-03	0.069
ξ_{11}	0.410	5.89e-03	0.514	0.414	5.84e-03	0.462	0.418	6.01e-03	0.300	0.419	6.14e-03	0.126	0.420	6.15e-03	0.116
π	0.983	1.66e-03	0.119	0.986	1.66e-03	0.110	0.989	1.94e-03	0.054	0.989	2.12e-03	0.026	0.990	2.18e-03	0.027

Panel B: Regular-duration models

	R-CSCD(1)			R-CSCD(2)			R-CSCD(3)			R-CSCD(4)			R-CSCD(5)		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
$\rho^{\dagger}$	0.276	7.88e-02	0.075	0.129	3.40e-02	0.056	0.126	3.35e-02	0.037	0.129	3.31e-02	0.031	0.133	3.73e-02	0.024
σ	0.485	2.39e-02	0.098	0.471	3.13e-02	0.071	0.471	3.09e-02	0.089	0.465	3.07e-02	0.065	0.465	3.03e-02	0.076
β_1				0.656	3.17e-02	0.112	0.439	2.09e-02	0.073	0.267	1.96e-02	0.053	0.196	1.84e-02	0.037
β_2				0.344	3.17e-02	0.112	0.321	2.01e-02	0.082	0.349	2.53e-02	0.062	0.310	2.77e-02	0.053
β_3							0.241	2.72e-02	0.081	0.132	2.64e-02	0.048	0.140	3.07e-02	0.050
β_4										0.252	2.51e-02	0.047	0.152	2.55e-02	0.046
β_5													0.201	2.56e-02	0.029

†: Rescale $\times 60$ (mean reversion for duration in minutes).

Note. The table gives the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the posterior mean, based on 25,000 posterior draws recorded after a burn-in period of 10,000 draws. The five groups of these three columns are for one to five component normalized duration densities.

Table 3.6: Posterior mean and standard deviation of various parameters, for the GW-filtered TD subsample and the regular-duration model; and for the full TD sample and the all-duration model.

Panel A: All-duration models

	A-CSCD(1)			A-CSCD(2)			A-CSCD(3)			A-CSCD(4)			A-CSCD(5)		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
$\rho^{\dagger}$	0.729	1.34e-01	0.087	0.360	8.14e-02	0.038	0.262	6.62e-02	0.023	0.271	6.68e-02	0.017	0.262	6.33e-02	0.018
σ	0.595	2.01e-02	0.085	0.571	2.48e-02	0.067	0.561	2.74e-02	0.068	0.562	2.57e-02	0.076	0.564	2.76e-02	0.070
β_1				0.700	3.28e-02	0.089	0.509	1.99e-02	0.033	0.396	2.77e-02	0.032	0.324	2.49e-02	0.029
β_2				0.300	3.28e-02	0.089	0.184	2.52e-02	0.037	0.196	3.37e-02	0.037	0.175	3.16e-02	0.029
β_3							0.307	2.66e-02	0.035	0.186	3.26e-02	0.039	0.176	3.22e-02	0.035
β_4										0.222	3.11e-02	0.033	0.142	2.74e-02	0.035
β_5													0.183	2.90e-02	0.037
ξ_{00}	0.804	2.90e-03	0.412	0.803	2.95e-03	0.358	0.800	3.05e-03	0.204	0.800	3.06e-03	0.190	0.799	3.15e-03	0.123
ξ_{11}	0.482	6.12e-03	0.548	0.485	6.11e-03	0.528	0.491	6.20e-03	0.347	0.492	6.25e-03	0.253	0.492	6.23e-03	0.228
π	0.987	1.41e-03	0.118	0.989	1.43e-03	0.088	0.992	1.47e-03	0.057	0.992	1.52e-03	0.048	0.993	1.60e-03	0.036

Panel B: Regular-duration models

	R-CSCD(1)			R-CSCD(2)			R-CSCD(3)			R-CSCD(4)			R-CSCD(5)		
	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE	Mean	Std.	RNE
$\rho^{\dagger}$	0.606	1.12e-01	0.123	0.275	6.39e-02	0.037	0.252	6.52e-02	0.025	0.268	6.65e-02	0.020	0.283	7.10e-02	0.016
σ	0.563	1.98e-02	0.114	0.535	2.53e-02	0.097	0.531	2.66e-02	0.087	0.527	2.53e-02	0.072	0.524	2.51e-02	0.050
β_1				0.703	3.86e-02	0.085	0.476	2.75e-02	0.044	0.264	2.18e-02	0.035	0.173	1.97e-02	0.032
β_2				0.297	3.86e-02	0.085	0.309	2.13e-02	0.042	0.384	2.47e-02	0.068	0.366	2.67e-02	0.037
β_3							0.215	2.90e-02	0.072	0.094	2.37e-02	0.060	0.112	2.75e-02	0.034
β_4										0.258	2.46e-02	0.040	0.126	2.39e-02	0.045
β_5													0.223	2.53e-02	0.032

†: Rescale ×60 (mean reversion for duration in minutes).

Note. The table gives the posterior mean and standard deviation, and the relative numerical efficiency (RNE) for the posterior mean, based on 25,000 posterior draws recorded after a burn-in period of 10,000 draws. The five groups of these three columns are for one to five component normalized duration densities.

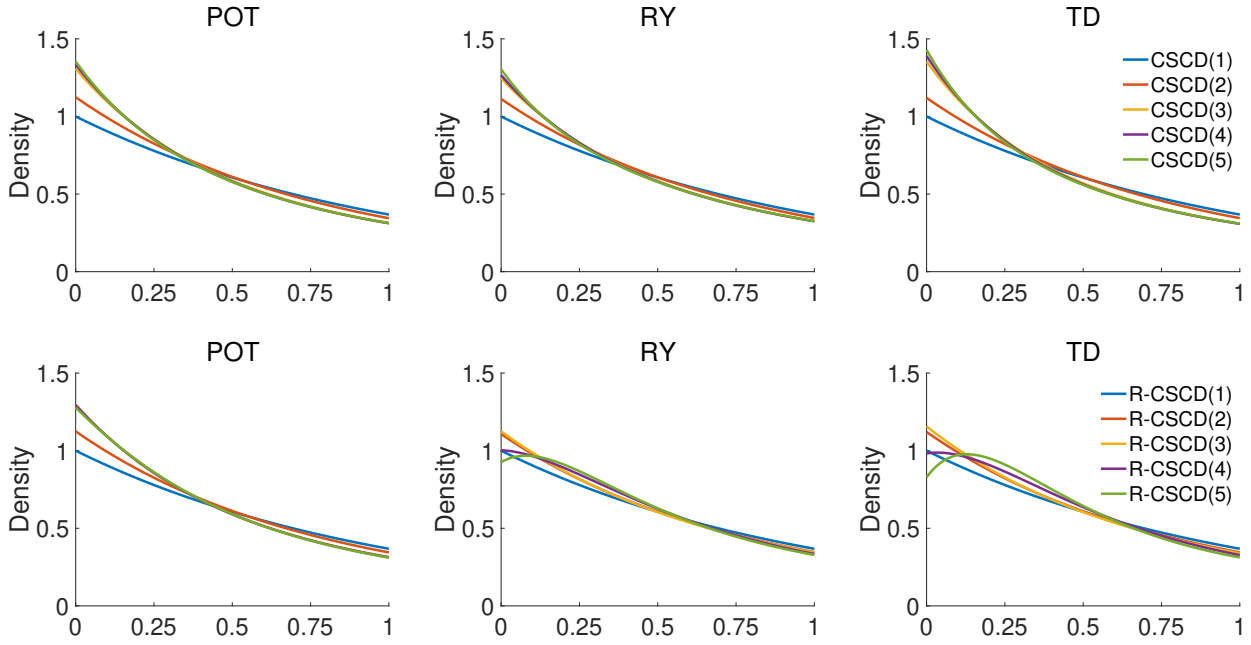

Figure 3.3: Normalized density functions at the posterior mean of β. Upper panels are for the all-duration models and the full samples; lower panels, for the regular-duration models and the GW-filtered subsamples. Panels on the right are for the POT series; panels in the middle, for the RY series; panels on the left, for TD series.

duration models and the GW-filtered subsamples; POT on the left, RY in the middle and TD on the right. We see that the conditional density at the posterior mean of β varies much more with J at shorter durations than it does at longer durations. With the exception of the POT series, there is a non-negligible difference between the densities for the all-duration and regular-duration models obtained with 4 and 5 terms. For the regular-duration models (lower panels), the density at zero decreases with J beyond three terms for the RY and the TD series. With extra flexibility, the density increasingly fits the spurious scarcity of $0s$ durations in the GW-filtered sample. For the all-duration models (upper panels), the density at zero increases with J, and this, for the three series. With the latent classification approach, there is more variation of the density near zero, where the uncertainty about the number of regular $0s$ and $1s$ durations comes into play. For each series, the mean densities obtained using 3, 4 and 5 terms are barely distinguishable. These results suggest that, for the all-duration models, a specification using $J = 3$ term is able to capture well the normalized density.

Table 3.7 reports posterior quantiles and moments of the half-life $t_{1/2}$, measured in seconds, of the OU process $x_d(t)$: the length of time it takes for $x_d(t)$ and $x_d(t + t_{1/2})$ to have a correlation of $1/2$ between them. This quantity is more easily interpretable than the mean reversion parameter

Table 3.7: Posterior quantiles and moments of the half-life $t_{1/2}$, measured in seconds, of the latent intensity state OU process $x_d(t)$.

		All-duration models					Regular-duration models				
		Mean	Std.	$\mathcal{Q}_{0.01}$	$\mathcal{Q}_{0.50}$	$\mathcal{Q}_{0.99}$	Mean	Std.	$\mathcal{Q}_{0.01}$	$\mathcal{Q}_{0.50}$	$\mathcal{Q}_{0.99}$
POT	$J = 1$	178	50	93	171	324	136	37	72	131	249
	$J = 5$	416	106	232	401	727	399	100	220	386	695
RY	$J = 1$	234	64	120	224	422	163	48	84	156	307
	$J = 5$	422	108	224	409	732	337	94	171	324	610
TD	$J = 1$	59	11	38	58	89	71	13	46	70	110
	$J = 5$	168	41	96	163	293	156	39	84	152	269

ρ. The left columns are for the full samples and the all-duration models and the right columns for the GW-filtered subsamples and the regular-duration models. For each series and model, we report the results for the $J = 1$ and the $J = 5$ specifications. Persistence of the latent intensity process is fairly high, but much less so for the TD series. 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 to the number of terms, J, in the normalized duration density. The marginal standard deviation σ is estimated more precisely (see Tables 3.4 , 3.5 and 3.6). Its distribution is less affected by how the regular durations are classified and depends very little on J.

Rather than estimate J, which is difficult in models with two different state variables, we compare specifications in term of their in-sample predictive performance using log-predictive scores (LPS):

$$\text{LPS} \equiv -\frac{1}{n}\sum_{d=1}^{D}\sum_{i=1}^{n_d}\log p(y_{d,i}\,|\,y) \tag{3.16}$$

where $p(y_{d,i}\,|\,y)$ is the pointwise predictive density defined as

$$p(y_{d,i}\,|\,y) = \int p(y_{d,i}\,|\,\Theta, s, x)p(\Theta, s, x\,|\,y)d\Theta ds dx.$$

To compute this density, we evaluate the expectation through simulation using draws from the posterior distribution:

$$p(y_{d,i}\,|\,y) \approx \frac{1}{M}\sum_{m=1}^{M} p(y_{d,i}\,|\,s_{d,i}^{(m)}, x_{d,i}^{(m)}, \pi^{(m)}, \beta^{(m)}), \tag{3.17}$$

where $s_{d,i}^{(m)}$, $x_{d,i}^{(m)}$, $\pi^{(m)}$ and $\beta^{(m)}$ are the m'th posterior draws generated from our posterior simulation

Table 3.8: In sample log-predictive score

		A-CSCD(1)	A-CSCD(2)	A-CSCD(3)	A-CSCD(4)	A-CSCD(5)
	LPS	1.0883	1.0948	1.0965	1.0967	1.0967
POT	LPS 0-1	0.1712	0.1691	0.1648	0.1644	0.1642
	LPS 2+	4.5227	4.5611	4.5855	4.5875	4.5884
	LPS	1.1807	1.1854	1.1857	1.1856	1.1856
RY	LPS 0-1	0.2180	0.2134	0.2080	0.2074	0.2068
	LPS 2+	4.2212	4.2551	4.2737	4.2751	4.2771
	LPS	1.0857	1.0956	1.0999	1.0998	1.1001
TD	LPS 0-1	0.1940	0.1922	0.1863	0.1856	0.1851
	LPS 2+	4.2161	4.2671	4.3073	4.3093	4.3120
		R-CSCD(1)	R-CSCD(2)	R-CSCD(3)	R-CSCD(4)	R-CSCD(5)
	LPS	4.2965	4.3294	4.3408	4.3405	4.3400
POT	LPS 0-1	3.0519	3.0219	2.9615	2.9624	2.9659
	LPS 2+	4.5170	4.5610	4.5852	4.5847	4.5834
	LPS	3.9703	3.9955	3.9963	3.9953	3.9941
RY	LPS 0-1	2.8355	2.8129	2.8074	2.8302	2.8365
	LPS 2+	4.2149	4.2504	4.2525	4.2463	4.2435
	LPS	3.9643	4.0033	4.0076	4.0040	4.0012
TD	LPS 0-1	2.7520	2.7472	2.7391	2.7676	2.7771
	LPS 2+	4.2226	4.2709	4.2778	4.2674	4.2620

methods and M is the total number of posterior draws. A lower LPS is an indication of a better in-sample predictive performance. In addition to LPS, we also compute (in-sample) log-predictive scores restricted to zero and one second durations (LPS 01), and durations of two seconds and more (LPS 2+), each obtained by averaging the pointwise densities of the corresponding observations. Table 3.8 shows the results; the upper panel are for full samples and all-duration models; the lower panel for GW-filtered subsamples and regular-duration models. For the all-duration models, the $J = 1$ specification provides the best in-sample predictive performance for each series; additional flexibility allows a better fit for the $0s$ and $1s$ durations, but at the expense of a worse fit for longer durations. Notice that we obtain similar results for the regular-duration models. Based on in-sample LPS, an all-duration model with an exponential distribution as normalized conditional density is a sensible choice for modelling censored trade durations with zero inflation.

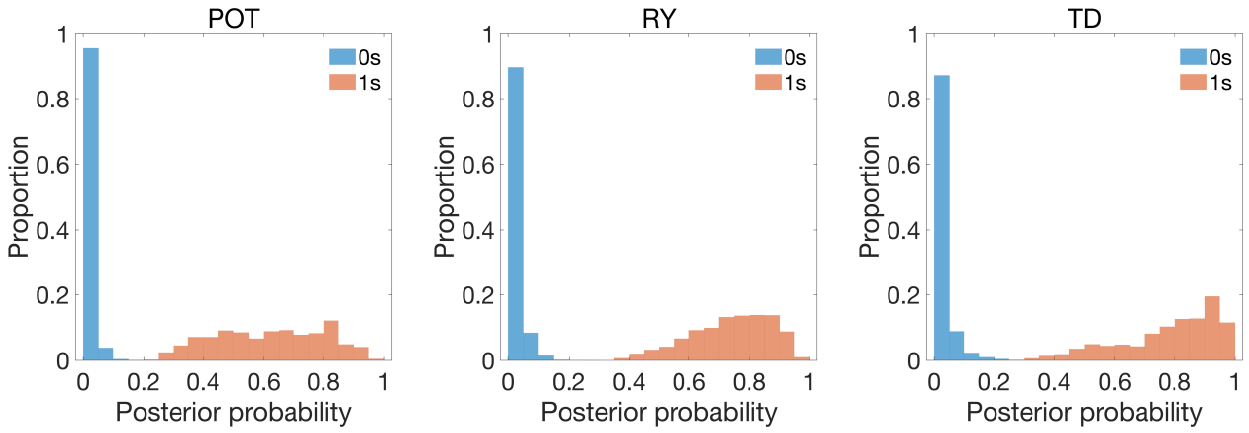

Figure 3.4: Histogram of posterior probabilities of being regular, for $0s$ or $1s$ durations, for the full sample and the all-duration model with $J = 1$. The histogram on the left is for the POT series; the one in the middle, the RY series; the one on the right, the TD series. This histograms illustrate variation over observations, not posterior uncertainty.

Table 3.9: Posterior quantiles and moments for the number of $0s$ and $1s$ discrete durations classified as regular.

	$0s$	$1s$	Mean	Std.	$q_{0.01}$	$q_{0.99}$	Mean	Std.	$q_{0.01}$	$q_{0.99}$
POT	19017	870	275	19	234	320	540	22	489	591
RY	22359	1443	554	27	493	617	1063	32	989	1136
TD	23782	1399	615	30	548	685	1097	29	1029	1163

Classification

For all-duration models, we record, at each posterior draw, whether each duration recorded as $0s$ or $1s$ is a regular or a cluster duration; classifications vary from draw to draw. We summarize this in two ways: first, we compute, for each of these durations, the probability that it is regular and illustrate the variation of this probability over durations; second, we describe the posterior distribution of the number of these durations that are regular. Figure 3.4 shows histograms of classification probabilities for durations of $0s$ and $1s$ under the all-duration model with $J = 1$, for each series. The horizontal axes give the posterior probability that a duration is regular, and the vertical height of the bar at a given histogram bin gives the proportion of $0s$ or $1s$ durations whose posterior probability of being regular is within that bin. Almost every $0s$ duration has a posterior probability less than 0.2 of being a regular duration; for a large majority, it is less than 0.05. In contrast, most $1s$ durations have a posterior probability of more than 0.5 of being regular.

While $0s$ durations are each quite unlikely to be regular, they are very numerous, and so the

probability that many of them are regular is nonetheless very high. Table 3.9 shows posterior quantiles and moments of the number of $0s$ and $1s$ durations classified as regular. The first two columns report the total number of $0s$ and $1s$ durations. Results for $0s$ durations are reported in the next four columns; those for $1s$ durations, in the last four columns. For comparison, the GW rule classifies 71 durations of $0s$ as regular for the POT series, 181 for RY series, and 129 for TD series. The GW rule does not apply to durations of $1s$ and so all them are considered regular (third column). Our probabilistic approach classifies as regular many more $0s$ durations and slightly fewer $1s$ durations.

To summarize: for each $0s$ duration, classifying it as a cluster duration is the right choice, under symmetric loss. Collectively, these zero/one decisions lead one to severely underestimate the number of these durations that are regular. With our probabilistic approach, we get a very good idea about how many of these durations are regular, although of course we do not magically discover which ones. This approach is exactly how we avoid artifacts arising from the spurious aggregation of unrelated, but nearly simultaneous, trades.

3.5 Concluding remarks

Models in the literature are designed to capture regular durations, those between unrelated trades. They are not intended, nor well suited, to capture the observed clustering of related trades. Common practice is to aggregate seemingly related trades into clusters and 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 aggregation rule, the same-second rule, amounts to calling all $0s$ durations cluster durations, and all others, regular durations. It is clear, however, that many of the $0s$ durations must be regular by happenstance—we just don't know which ones—and they are erroneously classified as cluster durations by the same-second 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 durations between $0s$ and $1s$ makes it difficult to fit a conditional duration distribution. The rule suggested by Grammig & Wellner (2002) clearly mitigates the problem, but it does not eliminate it.

The solution we proposed is (1) to work with truncated transaction times and to model the resulting discrete durations as a censored continuous-time process and (2) to make our model a mixture model, with a binary state variable indicating which durations are cluster durations and which are regular. Identification of the two states comes from the very tight distribution of cluster durations and, more subtly, the shrinkage of the underlying conditional duration density towards an exponential density, which varies slowly near zero. This probabilistic, rather than deterministic, classification, allows us to learn that any given pair of trades recorded in the same second are very probably related, at the same time as we learn that many such pairs are not in fact pairs of related trades.

In an empirical example, we found that the underlying conditional duration density is well captured by an exponential distribution, whether or not the cluster durations are classified using our probabilistic approach or the deterministic rule suggested by Grammig & Wellner (2002). However, the measurement of trading intensity differs considerably between the classification rules. We claim that this is an artifact of the misclassification of many unrelated but nearly simultaneous trades as being related.

Despite not learning which discrete $0s$ durations are regular, we were able to estimate quite precisely the underlying conditional duration density. 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, 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.

Bibliographie

Bibliography

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.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* **12**(3), C1–C32.

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**(2), 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. (2020), Zero-inflated autoregressive conditional duration model for discrete trade durations with excessive zeros, Technical report, ArXiv.
URL: *https://arxiv.org/pdf/1812.07318.pdf*

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

Chan, J. C. & Jeliazkov, I. (2009), Efficient simulation and integrated likelihood estimation in state-space models, Working paper.

Chen, Y., Hanson, T. & Zhang, J. (2014), 'Accelerated hazards model based on parametric families generalized with Berntein polynomials', *Biometrics* **70**(1), 192–201.

Chib, S. (1995), 'Marginal likelihood from the gibbs output', *Journal of the American Statistical Association* **90**(432), 1313–1321.

Chib, S., Nardari, F. & Shephard, N. (2002), 'Markov chain monte carlo methods for stochastic volatility models', *Journal of Econometrics* **108**(2), 281–316.

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 & 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.

Djegnene, B. & McCausland, W. J. (2015), 'The HESSIAN method with conditional dependence', *Journal of Financial Econometrics* **13**, 722–755.

Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987), 'Hybrid Monte Carlo', *Physics Letters B* **195**(2), 216–222.

Durbin, J. & Koopman, S. J. (1997), 'Monte Carlo maximum likelihood estimation for non-Gaussian state space models', *Biometrika* **84**(3), 669–684.

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 frontiers 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**(5), 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.

Frühwirth-Schnatter, S. & Frühwirth, R. (2007), 'Auxiliary mixture sampling with applications to logistic models', *Computational Statisitcs & Data Analysis* **51**, 3509–3528.

Frühwirth-Schnatter, S. & Wagner, H. (2006), 'Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling', *Biometrika* **93**, 827–841.

Frühwrith-Schnatter, S. (2006), *Finite Mixture and Markov Swithching Models*, Springer Series in Statistics, Springer-Verlag New York.

Gallant, A. R. (1981), 'On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form', *Journal of Econometrics* **15**, 211–245.

Gelman, A., Roberts, G. O. & Gilks, W. R. (1996), 'Efficient metropolis jumping rules', *Bayesian Statistics* **5**, 599–607.

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**(467), 799–804.

Gourieroux, C., Jasiak, J. & Fol, G. L. (1999), 'Intra-day market activity', *Journal of Financial Market* **2**(3), 193–226.

Grammig, J. & Maurer, K.-O. (2000), 'Non-monotonic hazard functions and the autoregressive conditional duration model', *The Econometrics Journal* **3**(1), 16–38.

Grammig, J. & Wellner, M. (2002), 'Modeling the interdependence of volatility and inter-transaction duration processes', *Journal of Econometrics* **106**(2), 369–400.

Haario, H., Saksman, E. & Tamminen, J. (2001), 'An adaptive metropolis algorithm', *Bernoulli* **7**(2), 223–242.

Harvey, A. & Ito, R. (2020), 'Modeling time series when some observations are zero', *Journal of Econometrics* **214**(1), 33–45.

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

Hautsch, N., Malec, P. & Schienle, M. (2014), 'Capturing the zero: A new class of zero-augmented distribution and multiplicative error process', *Journal of Financial Econometrics* **12**(1), 89–121.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business and Economic Statistics* **20**(1), 69–87.

Jung, R. C., Kukuk, M. & Liesenfeld, R. (2006), 'Time series of count data: modeling, estimation and diagnostics', *Computational Statisitcs & Data Analysis* **51**, 2350–2364.

Kastner, G. & Frühwirth-Schnatter, S. (2014), 'Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models', *Computational Statisitcs & Data Analysis* **76**, 408–423.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *The Review of Economic Studies* **65**(3), 361–393.

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

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

Liesenfeld, R., Nolte, I. & Pohlmeier, W. (2006), 'Modelling financial transaction price movements: a dynamic integer count data model', *Empirical Economics* **30**, 795–825.

Liu, J. S. (2004), *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag New York.

Lorentz, G. G. (1953), *Bernstein polynomials*, University of Toronto Press.

Lunde, A. (1999), A generalized gamma autoregressive conditional duration model, Dicussion paper, Aalborg University.

McCausland, W. J. (2012), 'The HESSIAN method: Highly efficient simulation smoothing, in a nutshell', *Journal of Econometrics* **168**(2), 189–206.

McCausland, W. J., Miller, S. & Pelletier, D. (2011), 'Simulation smoothing for state-space models: A computational efficiency analysis', *Computational Statisitcs & Data Analysis* **55**(1), 199–212.

Men, Z., Kolkeiwicz, A. W. & Wirjanto, T. S. (2015), 'Bayesian analysis of asymmetric stochastic conditional duration model', *Journal of Forcasting* **34**(1), 36–56.

Omori, Y., Chib, S., Shephard, N. & Nakajima, J. (2007), 'Stochastic volatility with leverage: fast and efficient likelihood inference', *Journal of Econometrics* **140**, 425–449.

Omori, Y. & Watanabe, T. (2008), 'Block sampler and posterior mode estimation for asymmetric stochastic volatility models', *Computational Statisitcs & Data Analysis* **52**, 2892–2910.

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', *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 (Statistical Methodology)* **64**(1), 79–100.

Quintana, F. A., Steel, M. F. J. & Ferreira, J. T. A. S. (2009), 'Flexible univariate continuous distribution', *Bayesian Analysis* **4**(4), 497–522.

Richard, J.-F. & Zhang, W. (2007), 'Efficient high-dimensional importance sampling', *Journal of Econometrics* **141**(2), 1385–1411.

Roberts, G. O. & Rosenthal, J. S. (1998), 'Optimal scaling of discrete approximations to langevin diffusions', *Journal of the Royal Statistical Society Series B* **60**(1), 255–268.

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

Roberts, G. & Stramer, O. (2003), 'Langevin diffusions and metropolis-hastings algorithms', *Methodology and Computing in Applied Probability* **4**, 337–358.

Russell, J. R. & Engle, R. F. (2005), 'A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model', *Journal of Business & Economic Statistics* **23**(2), 166–180.

Rydberg, T. & Shephard, N. (2003), 'Dynamics of trade-by-trade price movements: decompositions and models', *Journal of Financial Econometrics* **1**, 2–25.

Shephard, N. & Pitt, M. K. (1997), 'Likelihood analysis of non-Gaussian measurement time series', *Biometrika* **84**(3), 653–667.

Strickland, C. M., Forbes, C. S. & Martin, G. M. (2006), 'Bayesian analysis of the stochastic conditional duration model', *Computational Statisitcs & Data Analysis* **50**(9), 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.

Stroud, J. R., Müller, P. & Polson, N. G. (2003), 'Nonlinear state-space models with state-dependent variances', *Journal of the American Statistical Association* **98**, 377–386.

Taylor, S. J. (1982), Financial returns modelled by the product of two stochastic processes–a study of daily sugar prices 1691–1679., *in* O. D. Anderson, ed., 'Financial returns modelled by the product of two stochastic processes–a study of daily sugar prices 1691–1679.', North–Holland, pp. 203–226.

Tsay, R. (2002), *Analysis of Financial Time Series*, Wiley.

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).

Vihola, M. (2012), 'Robust adaptive Metropolis algorithm with coerced acceptance rate', *Statistics and Computing* **22**, 997–1008.

Watanabe, T. & Omori, Y. (2004), 'A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard and Pitt (1997)', *Biometrika* **91**, 246–248.

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

Xu, D., Knight, J. & Wirjanto, T. S. (2011), 'Asymmetric stochastic conditional duration model-a mixture-of-normal approach', *Journal of Financial Econometrics* **9**(3), 469–488.

Yatigammana, R. P., Chan, J. S. K. & Gerlach, R. H. (2019), 'Forecasting trade durations via ACD models with mixture distributions', *Quantitative Finance* **19**(12), 2051–2067.

Annexes

Appendices

A Appendices for Chapter 1

A.1 Computing the approximate gradient and Hessian of $\ln p(y|\theta)$

Here we compute approximations of $g_{y|\theta}(\theta)$ and $H_{y|\theta}(\theta)$, the gradient and Hessian of $\ln p(y|\theta)$ with respect to θ. For the special case where θ is the complete parameter vector, $g_{y|\theta}(\theta)$ and $H_{y|\theta}(\theta)$ are the gradient and Hessian of the log likelihood function.

This appendix is organized as follows. First, we show that $g_{y|\theta}(\theta) = E_{x|\theta,y}[g_{x|\theta}(\theta)]$ and $H_{y|\theta}(\theta) = E_{x|\theta,y}[H_{x|\theta}(\theta)] + \mathrm{Var}_{x|\theta,y}[g_{x|\theta}(\theta)]$, where $g_{x|\theta}(\theta)$ and $H_{x|\theta}(\theta)$ are the gradient and Hessian of $\ln p(x|\theta)$ with respect to θ. We then compute exact expressions for $g_{x|\theta}(\theta)$ and $H_{x|\theta}(\theta)$ followed by a description of exact, but infeasible, sequential computation for their moments $E_{x|\theta,y}[g_{x|\theta}(\theta)]$, $E_{x|\theta,y}[H_{x|\theta}(\theta)]$, and $\mathrm{Var}_{x|\theta,y}[g_{x|\theta}(\theta)]$. Next, we modify these computations so that they are feasible, but approximate, drawing on the approximate distribution for $x|\theta, y$ constructed using the HESSIAN method. Finally, we provide a step-by-step overview of these computations.

For the rest of this appendix, all expectations, including variances and covariances, are conditional on θ and y. To simplify the exposition, we suppress the notation for this conditioning, and speak of functions of θ and y as constants.

Derivatives of $\ln p(y|\theta)$ as expectations with respect to $x|\theta, y$

We can write the conditional posterior density of x as

$$p(x|\theta, y) = \frac{p(x|\theta)}{p(y|\theta)} p(y|x) = e^{\psi(x,\theta)} p(y|x),$$

where $\psi(x,\theta) = \log p(x|\theta) - \log p(y|\theta)$. Integrating both sides of this equation yields the identity

$$\int_{\mathbb{R}^n} e^{\psi(x,\theta)} p(y|x)\, dx = 1.$$

We now proceed to take partial derivatives with respect to elements of θ, denoting partial derivatives of $\psi(x,\theta)$ using subscripts. The partial derivative with respect to θ_i is

$$\int_{\mathbb{R}^n} \psi_i(x,\theta) e^{\psi(x,\theta)} p(y|x)\, dx = E_{x|\theta}[\psi_i(x,\theta)] = 0,$$

and organizing the three first order partial derivatives in vectorial form gives

$$g_{y|\theta}(\theta) = E_{x|\theta,y}[g_{x|\theta}(\theta)]. \tag{A.1}$$

Taking a second partial derivative, with respect to θ_j, gives

$$\int_{\mathbb{R}^n} [\psi_{ij}(x,\theta) + \psi_i(x,\theta)\psi_j(x,\theta)]\, e^{\psi(x,\theta)} p(y|x)\, dx = E_{x|\theta,y}[\psi_{ij}(x,\theta) + \psi_i(x,\theta)\psi_j(x,\theta)] = 0,$$

and organizing the second order partial derivatives in matrix notation gives $E_{x|\theta,y}[H_{y|\theta}(\theta) - H_{x|\theta}(\theta)] + \mathrm{Var}_{x|\theta,y}[g_{x|\theta}(\theta)] = 0$ or, equivalently,

$$H_{y|\theta}(\theta) = E_{x|\theta,y}[H_{x|\theta}(\theta)] + \mathrm{Var}_{x|\theta,y}[g_{x|\theta}(\theta)]. \tag{A.2}$$

Computing $g_{x|\theta}(\theta)$ and $H_{x|\theta}(\theta)$

We now compute exact expressions for $g_{x|\theta}(\theta)$ and $H_{x|\theta}(\theta)$. We first write

$$\ln p(x|\theta) = \frac{n}{2}(\ln\omega - \ln 2\pi) + \frac{1}{2}\ln(1-\phi^2) - \frac{\omega}{2} e^\top Q e = \frac{n}{2}(\theta_1 - \ln 2\pi) + \frac{1}{2}\ln(1-\tanh^2\theta_2) - \frac{e^{\theta_1}}{2} e^\top Q e,$$

where

$$Q = \begin{bmatrix} 1 & -\phi & 0 & \cdots & 0 \\ -\phi & 1+\phi^2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\phi^2 & -\phi \\ 0 & \cdots & 0 & -\phi & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\tanh\theta_2 & 0 & \cdots & 0 \\ -\tanh\theta_2 & 1+\tanh^2\theta_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1+\tanh^2\theta_2 & -\tanh\theta_2 \\ 0 & \cdots & 0 & -\tanh\theta_2 & 1 \end{bmatrix}.$$

The gradient and Hessian of $\ln p(x|\theta)$ with respect to θ are

$$g_{x|\theta}(\theta) = \begin{bmatrix} \frac{n}{2} - \frac{\omega}{2} e^\top Q e \\ -\phi - \frac{\omega}{2} e^\top Q' e \\ \omega q e \end{bmatrix}, \quad H_{x|\theta}(\theta) = \begin{bmatrix} -\frac{\omega}{2} e^\top Q e & -\frac{\omega}{2} e^\top Q' e & \omega q e \\ -\frac{\omega}{2} e^\top Q' e & -(1-\phi^2) - \frac{\omega}{2} e^\top Q'' e & \omega q' e \\ \omega q e & \omega q' e & -\omega q \iota \end{bmatrix}, \tag{A.3}$$

where Q' and Q'' are the first and second derivatives of the matrix Q with respect to the scalar θ_2, $q \equiv \iota^\top Q = (1-\phi, (1-\phi)^2, \ldots, (1-\phi)^2, (1-\phi))$ and $q' \equiv \iota^\top Q' = -(1-\phi^2)(1, 2(1-\phi), \ldots, 2(1-\phi), 1)$.

We compute Q' and Q'' as follows. There are two unique non-constant elements of Q, $-\phi = -\tanh\theta_2$ and $1+\phi^2 = 1+\tanh^2\theta_2$. Using $d\tanh\theta/d\theta = 1-\tanh^2\theta$, we obtain the first two derivatives of these two expressions:

$$\frac{d}{d\theta_2}[-\tanh\theta_2] = -(1-\tanh^2\theta), \quad \frac{d^2}{d\theta_2^2}[-\tanh\theta_2] = 2\tanh\theta_2(1-\tanh^2\theta_2),$$

$$\frac{d}{d\theta_2}[1+\tanh^2\theta_2] = 2\tanh\theta_2(1-\tanh^2\theta_2), \quad \frac{d^2}{d\theta_2^2}[1+\tanh^2\theta_2] = (2-6\tanh^2\theta_2)(1-\tanh^2\theta).$$

Then in terms of $\phi = \tanh\theta_2$,

$$Q' = (1-\phi^2)\begin{bmatrix} 0 & -1 & 0 & \cdots & 0 \\ -1 & 2\phi & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2\phi & -1 \\ 0 & \cdots & 0 & -1 & 0 \end{bmatrix}, \quad Q'' = 2(1-\phi^2)\begin{bmatrix} 0 & \phi & 0 & \cdots & 0 \\ \phi & 1-3\phi^2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1-3\phi^2 & \phi \\ 0 & \cdots & 0 & \phi & 0 \end{bmatrix}.$$

In the expressions for $g_{x|\theta}(\theta)$ and $H_{x|\theta}(\theta)$ in equation (A.3), we find the quadratic forms $e^\top Q e$, $e^\top Q' e$ and $e^\top Q'' e$ and the linear forms qe and $q'e$. In order to take means, variances and covariances of these, it is convenient to re-center them around the posterior mode $x^\circ = (x_1^\circ, \ldots, x_n^\circ)$, which is available as a computational byproduct of the HESSIAN method. For each t, define d_t and ϵ_t through the decomposition $e_t = x_t - \mu_t = (x_t^\circ - \mu_t) + (x_t - x_t^\circ) \equiv d_t + \epsilon_t$. Also let $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ and $d = (d_1, \ldots, d_n)$. Then $e^\top Q e = d^\top Q d + 2d^\top Q\epsilon + \epsilon^\top Q\epsilon$, and similarly for $e^\top Q' e$ and $e^\top Q'' e$. Also, $qe = qd + q\epsilon$ and $q'e = q'd + q'\epsilon$. Since d is constant,

$$E[e^\top Q e] = d^\top Q d + E[(2d+\epsilon)^\top Q\epsilon],$$

$$\mathrm{Var}[e^\top Qe] = \mathrm{Var}[(2d+\epsilon)^\top Q\epsilon]$$

$$\mathrm{Cov}[e^\top Qe, e^\top Q'e] = \mathrm{Cov}[(2d+\epsilon)^\top Q\epsilon, (2d+\epsilon)^\top Q'\epsilon],$$

$$\mathrm{Cov}[e^\top Qe, qe] = \mathrm{Cov}[(2d+\epsilon)^\top Q\epsilon, q\epsilon],$$

$$E[qe] = qd + E[q\epsilon], \quad \text{and} \quad \mathrm{Var}[qe] = \mathrm{Var}[q\epsilon].$$

Each of the expressions $E[e^\top Q'e]$, $E[e^\top Q''e]$, $E[q'e]$, $\mathrm{Var}[e^\top Q'e]$ and $\mathrm{Cov}[e^\top Qe, q'e]$ is similar to one of the above.

Sequential computation of $E_{x|\theta,y}[g_{x|\theta,y}]$, $E_{x|\theta,y}[H_{x|\theta,y}]$ and $\mathrm{Var}_{x|\theta,y}[g_{x|\theta,y}]$

The five quantities $(2d+\epsilon)^\top Q\epsilon$, $(2d+\epsilon)^\top Q'\epsilon$, $(2d+\epsilon)^\top Q''\epsilon$, $q\epsilon$ and $q'\epsilon$, can be expressed as the five sums

$$w_n^{(i)}(\epsilon_n) + \sum_{\tau=1}^{n-1} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}), \quad i = 1, \ldots, 5,$$

each for an appropriate choice of functions w_t. For the first quantity, $(2d+\epsilon)^\top Q\epsilon$, we have

$$w_1^{(1)}(\epsilon_1, \epsilon_2) = 2(d_1 - \phi d_2)\epsilon_1 + \epsilon_1^2 - 2\phi\epsilon_1\epsilon_2, \quad w_n^{(1)}(\epsilon_n) = 2(d_n - \phi d_{n-1})\epsilon_n + \epsilon_n^2,$$

and for $t = 2, 3, \ldots, n-1$,

$$w_t^{(1)}(\epsilon_t, \epsilon_{t+1}) = 2((1+\phi^2)d_t - \phi(d_{t-1} + d_{t+1}))\epsilon_t + (1+\phi^2)\epsilon_t^2 - 2\phi\epsilon_t\epsilon_{t+1}.$$

The expressions for $w_t^{(2)}(\epsilon_t, \epsilon_{t+1})$ and $w_t^{(3)}(\epsilon_t, \epsilon_{t+1})$, giving $(2d+\epsilon)^\top Q'\epsilon$ and $(2d+\epsilon)^\top Q''\epsilon$ are similar. For the fourth quantity, $q\epsilon$, we have

$$w_1^{(4)}(\epsilon_1, \epsilon_2) = (1-\phi)\epsilon_1, \quad w_n^{(4)}(\epsilon_n) = (1-\phi)\epsilon_n, \quad w_t^{(4)}(\epsilon_t, \epsilon_{t+1}) = (1-\phi)^2\epsilon_t,\ t = 2, \ldots, n-1.$$

The expression for $w_t^{(5)}(\epsilon_t, \epsilon_{t+1})$, giving $q'\epsilon$, is similar to this. The special structure of the quadratic forms comes from the fact that Q, Q' and Q'' are tridiagonal.

The computation of $E[g_{x|\theta}(\theta)]$, $E[H_{x|\theta}(\theta)]$ and $\mathrm{Var}[g_{x|\theta}(\theta)]$, requires computing the expectations

$$m_n^{(i)} = E\left[w_n^{(i)}(\epsilon_n) + \sum_{\tau=1}^{n-1} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1})\right], \tag{A.4}$$

for $i = 1, \ldots, 5$ and the covariances (including variances)

$$c_n^{(i,j)} = \mathrm{Cov}\left[w_n^{(i)}(\epsilon_n) + \sum_{\tau=1}^{n-1} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}),\ w_n^{(j)}(\epsilon_n) + \sum_{\tau=1}^{n-1} w_\tau^{(j)}(\epsilon_\tau, \epsilon_{\tau+1}) \right], \tag{A.5}$$

for $(i,j) \in I \equiv \{(1,1),(2,2),(4,4),(1,2),(1,4),(2,4)\}$. The $m_n^{(i)}$ can be obtained by repeated application of the law of iterated expectations: define, for $t = 0, \ldots, n-1$,

$$m_t^{(i)}(\epsilon_{t+1}) \equiv E\left[\sum_{\tau=1}^{t} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_{t+1} \right].$$

Then $m_0^{(i)}(\epsilon_1) = 0$ trivially and for $t = 1, \ldots, n-1$,

$$\begin{aligned} m_t^{(i)}(\epsilon_{t+1}) &= E\left[E\left[\sum_{\tau=1}^{t} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_t, \epsilon_{t+1} \right] \,|\, \epsilon_{t+1} \right] \\ &= E\left[w_t^{(i)}(\epsilon_t, \epsilon_{t+1}) + E\left[\sum_{\tau=1}^{t-1} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_t \right] \,|\, \epsilon_{t+1} \right] \\ &= E[w_t^{(i)}(\epsilon_t, \epsilon_{t+1}) + m_{t-1}^{(i)}(\epsilon_t) \,|\, \epsilon_{t+1}]. \end{aligned} \tag{A.6}$$

It is the posterior Markov property of $\{\epsilon_t\}$ that allows us to drop the ϵ_{t+1} from the inner expectation in the second line. Using a similar development, we obtain the final value $m_n^{(i)} = E[w_n^{(i)}(\epsilon_n) + m_{n-1}^{(i)}(\epsilon_n)]$.

Likewise, the $c_n^{(i,j)}$ can be obtained by repeated application of the law of total covariance: define, for $t = 0, \ldots, n-1$,

$$c_t^{(i,j)}(\epsilon_{t+1}) \equiv \mathrm{Cov}\left[\sum_{\tau=1}^{t} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}),\ \sum_{\tau=1}^{t} w_\tau^{(j)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_{t+1} \right].$$

Then $c_0^{(i,j)}(\epsilon_1) = 0$ and for $t = 1, \ldots, n-1$,

$$\begin{aligned} c_t^{(i,j)}(\epsilon_{t+1}) &= E\left[\mathrm{Cov}\left[\sum_{\tau=1}^{t} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}),\ \sum_{\tau=1}^{t} w_\tau^{(j)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_t, \epsilon_{t+1} \right] \,|\, \epsilon_{t+1} \right] \\ &\quad + \mathrm{Cov}\left[E\left[\sum_{\tau=1}^{t} w_\tau^{(i)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_t, \epsilon_{t+1} \right],\ E\left[\sum_{\tau=1}^{t} w_\tau^{(j)}(\epsilon_\tau, \epsilon_{\tau+1}) \,|\, \epsilon_t, \epsilon_{t+1} \right] \,|\, \epsilon_{t+1} \right] \\ &= E[c_{t-1}^{(i,j)}(\epsilon_t) \,|\, \epsilon_{t+1}] + \mathrm{Cov}[w_t^{(i)}(\epsilon_t, \epsilon_{t+1}) + m_{t-1}^{(i)}(\epsilon_t),\ w_t^{(j)}(\epsilon_t, \epsilon_{t+1}) + m_{t-1}^{(j)}(\epsilon_t) \,|\, \epsilon_{t+1}]. \end{aligned} \tag{A.7}$$

Finally, we obtain the desired value

$$c_n^{(i,j)} = E[c_{n-1}^{(i,j)}(\epsilon_n)] + \mathrm{Cov}[w_n^{(i)}(\epsilon_n) + m_{n-1}^{(i)}(\epsilon_n),\ w_n^{(j)}(\epsilon_n) + m_{n-1}^{(j)}(\epsilon_n)].$$

Now we can write, using (A.3), (A.4) and (A.5),

$$g_{y|\theta}(\theta) = E[g_{x|\theta}(\theta)] = \begin{bmatrix} \frac{n}{2} \\ -\phi \\ 0 \end{bmatrix} + \frac{\omega}{2} \begin{bmatrix} -d^\top Q d - m_n^{(1)} \\ -d^\top Q' d - m_n^{(2)} \\ 2(qd + m_n^{(4)}) \end{bmatrix} \tag{A.8}$$

and

$$\begin{aligned} H_{y|\theta}(\theta) = \mathrm{Var}[g_{x|\theta}(\theta)] + E[H_{x|\theta}(\theta)] &= \frac{\omega^2}{4} \begin{bmatrix} c_n^{(1,1)} & c_n^{(1,2)} & -2c_n^{(1,4)} \\ c_n^{(1,2)} & c_n^{(2,2)} & -2c_n^{(2,4)} \\ -2c_n^{(1,4)} & -2c_n^{(2,4)} & 4c_n^{(4,4)} \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 & 0 \\ 0 & -(1-\phi^2) & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{\omega}{2} \begin{bmatrix} -d^\top Q d - m_n^{(1)} & -d^\top Q' d - m_n^{(2)} & 2(qd + m_n^{(4)}) \\ -d^\top Q' d - m_n^{(2)} & -d^\top Q'' d - m_n^{(3)} & 2(q'd + m_n^{(5)}) \\ 2(qd + m_n^{(4)}) & 2(q'd + m_n^{(5)}) & -2q\iota \end{bmatrix}. \end{aligned} \tag{A.9}$$

The matrices Q, Q' and Q'' are tridiagonal and each has only three unique elements. Thus the quadratic form $d^\top Q d$ is efficiently computed with $O(n)$ operations as

$$\begin{aligned} d^\top Q d &= Q_{11} d_1^2 + Q_{nn} d_n^2 + \sum_{t=2}^{n-1} Q_{tt} d_t^2 + \sum_{t=1}^{n-1} Q_{t,t+1} d_t d_{t+1} \\ &= Q_{11}(d_1^2 + d_n^2) + Q_{22} \sum_{t=2}^{n-1} d_t^2 + Q_{12} \sum_{t=1}^{n-1} d_t d_{t+1}, \end{aligned}$$

and the quadratic forms $d^\top Q' d$ and $d^\top Q'' d$ with $O(1)$ additional operations by replacing Q_{11}, Q_{22} and Q_{12} in the second line with Q'_{11}, Q'_{22} and Q'_{12} and then Q''_{11}, Q''_{22} and Q''_{12}.

Feasible approximations of $E_{x|\theta,y}[g_{x|\theta,y}]$, $E_{x|\theta,y}[H_{x|\theta,y}]$ and $\mathrm{Var}_{x|\theta,y}[g_{x|\theta,y}]$

Unfortunately, we cannot evaluate the expectations and covariances in (A.6) and (A.7). Here, we provide feasible but approximate computations, similar to the exact but infeasible computations in A.1, but using the HESSIAN method proposal distribution $q(x|\theta, y)$ to approximate those ex-

pectations and covariances. Our strategy involves approximating various conditional moments by cubic polynomials. At each step $t = 1, \ldots, n-1$, we first compute approximations of $E[\epsilon_t|\epsilon_{t+1}]$ and several other conditional moments of ϵ_t given ϵ_{t+1}, as cubic polynomials in ϵ_{t+1}. We then compute approximations of the $m_t^{(i)}(\epsilon_{t+1})$ and $c_t^{(i,j)}(\epsilon_{t+1})$, which are conditional moments of functions of $(\epsilon_1, \ldots, \epsilon_t)$ given ϵ_{t+1}, as cubic polynomials in ϵ_{t+1}. At the last step $t = n$, we compute approximations of the final values $m_n^{(i)}$, $i = 1, \ldots, 5$ and $c_n^{(i,j)}$, $(i,j) \in I$, which are unconditional expectations of functions of ϵ_n. The following description is at a relatively high level, to avoid unnecessary detail. For example, we will express the product of two polynomials directly, rather than give expressions for individual coefficients. Our implementation in code mirrors this modularity; to continue the example, we use function calls for polynomial multiplication.

The HESSIAN method of McCausland (2012) generates many computational byproducts including, for $t = 1, \ldots, n-1$, a polynomial function $b_t(\epsilon_{t+1})$ approximating the conditional mode of ϵ_t given ϵ_{t+1}, and the polynomial functions $\delta_t(\epsilon_{t+1})$ and $s_t(\epsilon_{t+1})$ approximating $E[\epsilon_t - b_t(\epsilon_{t+1})|\epsilon_{t+1}]$ and $\ln E[(\epsilon_t - b_t(\epsilon_{t+1}))^2|\epsilon_{t+1}]$. It also generates analogous constants b_n, approximating the unconditional mode of ϵ_n; δ_n, approximating $E[\epsilon_n] - b_n$; and s_n, approximating $\ln E[(\epsilon_n - b_n)^2]$. We use $s_t(\epsilon_{t+1})$ to construct a polynomial $\Sigma_t(\epsilon_{t+1})$ approximating $E[(\epsilon_t - b_t(\epsilon_{t+1}))^2|\epsilon_{t+1}]$, based on a first order Taylor expansion of the exponential function. Reasonable approximations of $E[(\epsilon_t - b_t(\epsilon_{t+1}))^3|\epsilon_{t+1}]$ and $E[(\epsilon_t - b_t(\epsilon_{t+1}))^4|\epsilon_{t+1}]$ are $5\delta_t(\epsilon)\Sigma_t(\epsilon_{t+1})$ and $3\Sigma_t^2(\epsilon_{t+1})$, respectively. Although the HESSIAN method provides fifth order polynomials, we truncate $b_t(\epsilon_{t+1})$, $\delta_t(\epsilon_{t+1})$ and $\Sigma_t(\epsilon_{t+1})$ to cubic polynomials.

Using these, we define the following approximate conditional moments of ϵ_t given ϵ_{t+1}:

$$E[\epsilon_t|\epsilon_{t+1}] \approx \tilde{E}_{t,1}(\epsilon_{t+1}) \equiv b_t(\epsilon_{t+1}) + \delta_t(\epsilon_{t+1}), \tag{A.10}$$

$$\mathrm{Var}[\epsilon_t|\epsilon_{t+1}] \approx \tilde{V}_{t,1}(\epsilon_{t+1}) \equiv \Sigma_t(\epsilon_{t+1}) - \delta_t^2(\epsilon_{t+1}), \tag{A.11}$$

$$E[\epsilon_t^2|\epsilon_{t+1}] \approx \tilde{E}_{t,2}(\epsilon_{t+1}) \equiv \tilde{E}_{t,1}^2(\epsilon_{t+1}) + \tilde{V}_{t,1}(\epsilon_{t+1}), \tag{A.12}$$

$$\mathrm{Cov}[\epsilon_t, \epsilon_t^2|\epsilon_{t+1}] \approx \tilde{C}_{t,12}(\epsilon_{t+1}) \equiv 2b_t(\epsilon_{t+1})\tilde{V}_{t,1}(\epsilon_{t+1}) + 4\delta_t(\epsilon_{t+1})\Sigma_t(\epsilon_{t+1}), \tag{A.13}$$

$$E[\epsilon_t^3|\epsilon_{t+1}] \approx \tilde{E}_{t,3}(\epsilon_{t+1}) \equiv \tilde{E}_{t,1}(\epsilon_{t+1})\tilde{E}_{t,2}(\epsilon_{t+1}) + \tilde{C}_{t,12}(\epsilon_{t+1}), \tag{A.14}$$

$$\mathrm{Var}[\epsilon_t^2|\epsilon_{t+1}] \approx \tilde{V}_{t,2}(\epsilon_{t+1}) \equiv 2b_t^2(\epsilon_{t+1})\tilde{C}_{t,12}(\epsilon_{t+1}) + 2\Sigma_t^2(\epsilon_{t+1}), \tag{A.15}$$

where polynomial multiplication is taken to be multiplication followed by truncation of the result

to a cubic polynomial. The analogous time $t = n$ values are constant and defined as

$$\tilde{E}_{n,1} \equiv b_n + \delta_n, \quad \tilde{V}_{n,1} \equiv \Sigma_n - \delta_n^2, \quad \tilde{E}_{n,2} \equiv \tilde{E}_{t,1}^2 + \tilde{V}_{n,1} \tag{A.16}$$

$$\tilde{C}_{n,12} \equiv 2b_n \tilde{V}_{n,1} + 4\delta_n \Sigma_n, \quad \tilde{V}_{n,1} \equiv 2b_n^2 \tilde{C}_{n,12} + 2\Sigma_n^2. \tag{A.17}$$

We now define two operators that approximate conditional expectations and conditional covariances, respectively, for polynomials that take the form of a cubic in ϵ_t plus a term consisting of a coefficient times $\epsilon_t \epsilon_{t+1}$. Let $v^{(1)}(\epsilon_t, \epsilon_{t+1})$ and $v^{(2)}(\epsilon_t, \epsilon_{t+1})$ be two polynomials of this form:

$$v^{(i)}(\epsilon_t, \epsilon_{t+1}) = v_0^{(i)} + v_1^{(i)}\epsilon_t + v_2^{(i)}\epsilon_t^2 + v_3^{(i)}\epsilon_t^3 + v_{11}^{(i)}\epsilon_t\epsilon_{t+1}, \quad i = 1, 2.$$

We define the following operator approximating the conditional mean operator:

$$\bar{\mathrm{E}}_t[v^{(1)}(\epsilon_t, \epsilon_{t+1})] \equiv v_0^{(1)} + \sum_{k=1}^{3} v_k^{(1)} \tilde{E}_{t,k}(\epsilon_{t+1}) + v_{11}^{(1)} \epsilon_{t+1} \tilde{E}_{t,1}(\epsilon_{t+1}).$$

We define the following operator approximating the conditional covariance operator:

$$\begin{aligned}\bar{\mathrm{C}}_t[v^{(1)}(\epsilon_t, \epsilon_{t+1}), v^{(2)}(\epsilon_t, \epsilon_{t+1})] \equiv{}& (v_1^{(1)} + v_{11}^{(1)}\epsilon_{t+1})(v_1^{(2)} + v_{11}^{(2)}\epsilon_{t+1}) \cdot \tilde{V}_{t,1}(\epsilon_{t+1}) \\ &+ [(v_1^{(1)} + v_{11}^{(1)}\epsilon_{t+1})v_2^{(2)} + v_2^{(1)}(v_1^{(2)} + v_{11}^{(2)}\epsilon_{t+1})] \cdot \tilde{C}_{t,12}(\epsilon_{t+1}) \\ &+ v_2^{(1)} v_2^{(2)} \cdot \tilde{V}_{t,2}(\epsilon_{t+1}).\end{aligned}$$

Note that this uses only terms up to second order in the polynomials $v^{(1)}(\epsilon_t, \epsilon_{t+1})$ and $v^{(2)}(\epsilon_t, \epsilon_{t+1})$.

We approximate the $m_t^{(i)}(\epsilon_{t+1})$, $i = 1, \ldots, 5$, analogously with (A.6), as

$$\tilde{m}_t^{(i)}(\epsilon_{t+1}) \equiv \bar{\mathrm{E}}_t[w_t^{(i)}(\epsilon_t, \epsilon_{t+1}) + \tilde{m}_{t-1}^{(i)}(\epsilon_t)] \tag{A.18}$$

and the $c_t^{(i,j)}(\epsilon_{t+1})$, $(i, j) \in I$, analogously with (A.7), as

$$\tilde{c}_t^{(i,j)}(\epsilon_{t+1}) \equiv \bar{\mathrm{E}}_t[\tilde{c}_{t-1}^{(i,j)}(\epsilon_t)] + \bar{\mathrm{C}}_t[w_t^{(i)}(\epsilon_t, \epsilon_{t+1}) + \tilde{m}_{t-1}^{(i)}(\epsilon_t),\ w_t^{(j)}(\epsilon_t, \epsilon_{t+1}) + \tilde{m}_{t-1}^{(j)}(\epsilon_t)]. \tag{A.19}$$

Summary of computations

We summarize the computational routine as follows:

1. For $t = 1, \ldots, n-1$,

   (a) compute $\tilde{E}_{t,1}(\epsilon_{t+1})$, $\tilde{E}_{t,2}(\epsilon_{t+1})$, $\tilde{V}_{t,1}(\epsilon_{t+1})$, $\tilde{V}_{t,1}(\epsilon_{t+1})$, and $\tilde{C}_{t,12}(\epsilon_{t+1})$ using (A.10) through (A.15).

   (b) for $i = 1, \ldots, 5$, compute $\tilde{m}_t^{(i)}(\epsilon_{t+1})$ using (A.18).

   (c) for $(i,j) \in I$, compute $\tilde{c}_t^{(i,j)}(\epsilon_{t+1})$ using (A.18) and (A.19).

2. Compute $\tilde{E}_{n,1}$, $\tilde{E}_{n,2}$, $\tilde{V}_{n,1}$, $\tilde{V}_{n,1}$, and $\tilde{C}_{n,12}$ using (A.16) and (A.17).

3. For $i = 1, \ldots, 5$, compute $\tilde{m}_n^{(i)} \equiv \bar{\mathrm{E}}_n[w_n^{(i)} + \tilde{m}_{n-1}^{(i)}]$ using (A.18)

4. For $(i,j) \in I$ compute $\tilde{c}_n = \bar{\mathrm{E}}_n[\tilde{c}_{n-1}^{(i,j)}(\epsilon_n)] + \bar{\mathrm{C}}_n[\tilde{m}_{n-1}^{(i)}(\epsilon_n),\ \tilde{m}_{n-1}^{(j)}(\epsilon_n)]$ using (A.18) and (A.19)

5. Compute the approximation $\tilde{g}_{y|\theta}(\theta)$ of $g_{y|\theta}(\theta)$ defined by replacing $m_n^{(i)}$ with $\tilde{m}_n^{(i)}$, $i = 1, 2, 4$, in (A.8).

6. Compute the approximation $\tilde{H}_{y|\theta}(\theta)$ of $H_{y|\theta}(\theta)$ defined by replacing $m_n^{(i)}$ with $\tilde{m}_n^{(i)}$, $i = 1, \ldots, 5$, and $c_n^{(i,j)}$ with $\tilde{c}_n^{(i,j)}$, $(i,j) \in I$, in (A.9).

A.2 Using local shape information to approximate the maximum likelihood value

For μ fixed, we consider here the problem of constructing an approximation $\tilde{\vartheta}$ of $\hat{\vartheta} = \arg\max_\vartheta \ln p(y|\vartheta, \mu)$ and an approximation $\tilde{H}$ of the Hessian of $\ln p(y|\vartheta, \mu)$ with respect to ϑ at $\hat{\vartheta}$. We are given an initial value ϑ_0, and the values g and H of the gradient and Hessian of $\ln p(y|\vartheta, \mu)$ with respect to ϑ, evaluated at $\vartheta = \vartheta_0$.

An obvious choice of $\tilde{\vartheta}$ is $\tilde{\vartheta} = \vartheta_0 - H^{-1}g$, the result of a single Newton-Raphson step, and in fact that is usually what we do. If the log-likelihood were quadratic and concave, the result would be the exact maximum likelihood value $\hat{\vartheta}$, H would be the Hessian of the log likelihood there and $-\frac{1}{2}g^\top H^{-1}g$ would measure the increase of the log likelihood between ϑ_0 and $\tilde{\vartheta}$.

In practice, however, H is not constant and sometimes fails to be negative definite. We will take positive and very low values of the eigenvalues of H, as well as large values of $-g^\top H^{-1}g$, as evidence of large changes of the Hessian between ϑ_0 and $\hat{\vartheta}$.

We have found that there are regions of the parameter space where the Hessian matrix under the alternative parameterization (ϑ_1, ϕ), where $\phi \equiv \tanh \vartheta_2$, tends to be more stable. Let g_ϕ and H_ϕ

be the gradient and Hessian of $\ln p(y|\vartheta_1, \phi)$ with respect to the vector (θ_1, ϕ). Using the chain rule, and the fact that the first two derivatives of $\tanh \vartheta_2$ with respect to ϕ are $1 - \phi^2$ and $-2\phi(1 - \phi^2)$, we can write

$$g = Dg_\phi, \quad H = DH_\phi D + \begin{bmatrix} 0 & 0 \\ 0 & -2\phi g_2 \end{bmatrix}, \tag{A.20}$$

where

$$D \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 - \phi^2 \end{bmatrix}.$$

Now let

$$\Omega \equiv H + \begin{bmatrix} 0 & 0 \\ 0 & 2\phi g_2 \end{bmatrix}.$$

Under the parameterization (θ_1, ϕ), the Newton-Raphson step is

$$-H_\phi^{-1} g_\phi = -D\Omega^{-1} D \cdot D^{-1} g = -D\Omega^{-1} g,$$

and a quadratic approximation of $\ln p(y|\vartheta_1, \phi)$ gives the following approximation for the increase of the log likelihood $\ln p(y|\vartheta_1, \phi)$:

$$-\tfrac{1}{2} g_\phi^\top H_\phi^{-1} g_\phi = -\tfrac{1}{2} g^\top D^{-1} (D\Omega^{-1} D) D^{-1} g = -\tfrac{1}{2} g^\top \Omega^{-1} g.$$

When Ω is negative definite and H is not, or if Ω is negative definite and $-g^\top H^{-1} g$ is much larger than $-g^\top \Omega^{-1} g$, we take this as evidence that Ω has a more stable value in the region of the parameter space where ϑ_0 lies.

In some cases, we will perform the stabilized step $\tilde{\vartheta} = \vartheta_0 - (H + \alpha g g^\top)^{-1} g$, where $\alpha > 0$, or an analogous step using Ω. By the Woodbury matrix identity,

$$(H - \alpha g g^\top)^{-1} = H^{-1} - H^{-1} g (-\alpha^{-1} + g^\top H^{-1} g)^{-1} g^\top H^{-1} = H^{-1} + \frac{1}{\alpha^{-1} - g^\top H^{-1} g} H^{-1} g g^\top H^{-1},$$

and the stabilized step is

$$-(H - \alpha g g^\top)^{-1} g = -H^{-1} g - \frac{g^\top H^{-1} g}{\alpha^{-1} - g^\top H^{-1} g} H^{-1} g = -\left(\frac{1}{1 - \alpha g^\top H^{-1} g} \right) H^{-1} g,$$

which is the same direction as the Newton step $-H^{-1} g$ but attenuated by the factor in parentheses.

Then the increase of the likelihood under a quadratic log likelihood would be

$$-g^\top (H + \alpha g g^\top)^{-1} g = -\left(\frac{1}{1 - \alpha g^\top H^{-1} g}\right) g^\top H^{-1} g.$$

We compute $\tilde{\vartheta}$ and $\tilde{H}$ according to one of the following cases. Case 1 pertains if Ω is negative definite and either H is not negative definite or $-g^\top \Omega^{-1} g < -g^\top H^{-1} g + 1$. In this case, we compute

$$\begin{bmatrix} \tilde{\vartheta}_1 \\ \tilde{\phi} \end{bmatrix} = \begin{bmatrix} \vartheta_1 \\ \phi \end{bmatrix} - D(\Omega - \alpha g g^\top)^{-1} g,$$

where $\alpha = 0$ if $-g^\top \Omega^{-1} g < 6$ and 0.2 otherwise, and set $\tilde{\vartheta}_2 = \tanh^{-1} \tilde{\phi}$. We also set $\tilde{H} = H$.

Case 2 pertains if the condition for Case 1 fails and H is negative definite, in which case we compute $\tilde{\vartheta} = -(H - \alpha g g^\top)^{-1} g$, where $\alpha = 0$ if $-g^\top H^{-1} g < 6$ and 0.2 otherwise. We then set $\tilde{\phi} = \tanh \vartheta_2$. We set H according to equation (A.20), holding H_ϕ constant and re-evaluating with $\phi = \tilde{\phi}$.

Case 3 covers the case where both Ω and H fail to be negative definite. In this very rare case, we compute

$$\tilde{\vartheta} = -\frac{1}{\max(\frac{1}{2}, -g^\top \bar{H}^{-1} g/6)} \bar{H}^{-1} g,$$

where

$$\bar{H} = -\begin{bmatrix} 1 & \rho\sqrt{h} \\ \rho\sqrt{h} & h^2 \end{bmatrix},$$

$\rho = 0.75$ and

$$h = (\tfrac{1}{2} + \tfrac{1}{2}\phi) \max(6\phi^2 - 2, 0) + (\tfrac{1}{2} - \tfrac{1}{2}\phi) \max(2(1 + \phi)(1 + 3\phi), 0).$$

B Appendices for Chapter 2

B.1 Derivatives

We show here how to evaluate the first five derivatives of $\psi(x_{d,i}) \equiv \log p(y_{d,i} \mid z_{d,i}, x_{d,i})$ with respect to $x_{d,i}$, at an arbitrary value of $x_{d,i}$. Routines for these derivatives are required by the HESSIAN method. To avoid tedium and error, we do not provide analytic expressions for the derivatives. Instead, we give derivatives of primitive functions and show how to combine them using Fáa di Bruno's formula, a generalization of the chain rule to higher derivatives, to compute exact derivatives of $\psi(x_{d,i})$. It gives derivatives as

$$\frac{d^n}{dx^n} f(g(x)) = \sum \frac{n!}{m_1! m_2! \cdots m_n!} f^{(m_1+\cdots+m_n)}(g(x)) \prod_{j=1}^{n} \left(\frac{g^{(j)}(x)}{j!} \right)^{m_j}.$$

Substituting in equation (2.6) the exponential distribution function $F(\epsilon) = 1 - e^{-\lambda\epsilon}$ combined with the expression (2.8), conditional on $z_{d,i}$, gives

$$\begin{aligned}\psi(x_{d,i}) &= \log \left[\lambda e^{-x_{d,i}} e^{-\lambda e^{-x_{d,i}} y_{d,i}} \operatorname{Beta}\left(1 - \exp(-\lambda e^{-x_{d,i}} y_{d,i}) \mid z_{d,i}, J - z_{d,i} + 1\right) \right] \\ &= \text{Cte} - x_{d,i} + (J - z_{d,i} + 1) h(x_{d,i}) + (z_{d,i} - 1) f(x_{d,i})\end{aligned}$$

where $f(x_{d,i}) = \log(1 - g(x_{d,i}))$, $g(x_{d,i}) = \exp(h(x_{d,i}))$ and $h(x_{d,i}) = -\lambda e^{x_{d,i}} y_{d,i}$. Recall that the hazard rate λ is a linear function of β through the unit-mean normalization of the density $p_\epsilon(\cdot)$. We compute the derivatives of bottom up. The steps are

1. Compute $h(x_{d,i})$ and its first five derivatives with respect to $x_{d,i}$, $j = 1, \ldots, J$:

$$\begin{aligned}h'(x_{d,i}) = h'''(x_{d,i}) = h^{(5)}(x_{d,i}) = -h(x_{d,i}) \\ h''(x_{d,i}) = h^{(4)}(x_{d,i}) = h(x_{d,i}).\end{aligned}$$

2. Compute $g(x_{d,i})$ and first five derivatives with respect to $x_{d,i}$, using Fáa di Bruno's rule, for the exponential function composed with $h(x_{d,i})$.

3. Compute $f(x_{d,i})$ and first five derivatives with respect to $x_{d,i}$, using Fáa di Bruno's rule, for the logarithmic function composed with $1 - g(x_{d,i})$. The first five derivatives of $q(z) = \log z$

are

$$q'(z) = z^{-1}, \quad q''(z) = -z^{-2}, \quad q'''(z) = 2z^{-3}, \quad q^{(4)}(z) = -6z^{-4}, \quad q^{(5)} = 24z^{-5}(z).$$

B.2 Drawing artificial observations

Here we show how to draw observations from its conditional distribution $p(y \mid \beta, x)$. For latent state process used for the Getting it right experiment (see equation (2.14)), there is no relationship between the durations and the dynamic of the latent state. The prior distribution also implies a fairly flat diurnal pattern, thus, very little time variation in the unconditional mean of the latent state process. In this case, updating the vector of observations can be done efficiently with the following Metropolis-Hastings algorithm:

- For $d = 1, \ldots, D$, $i = 1, \ldots, n_d$, draw a proposal $y^*_{d,i} \sim p(y_{d,i} \mid \beta, x_i)$.
- For $d = 1, \ldots, D$, set $y^*_{d,0} = t_{\text{open}}$ and construct the corresponding transaction times $t^*_{d,i} = \sum_{k=0}^{i-1} y^*_{d,k}$ to evaluate the B-spline function $m(t)$.
- Accept the proposal y^* with probability

$$\min \left\{ 1, \prod_{d=1}^{D} \prod_{i=1}^{n_d - 1} \frac{p(x_{d,i+1} \mid x_{d,i}, t_{d,i}, t_{d,i-1}, \sigma, \rho, \delta, \psi_d)}{p(x_{d,i+1} \mid x_{d,i}, t^*_{d,i}, t^*_{d,i-1} \sigma, \rho, \delta, \psi_d)} \right\}.$$

C Appendices for Chapter 3

C.1 Derivatives

We show here how to evaluate the first five derivatives of $\psi(x_{d,i}) \equiv \log p(y_{d,i} \,|\, s_{d,i}, x_{d,i}, z_{d,i})$ with respect to $x_{d,i}$, at an arbitrary value of $x_{d,i}$. Routines for these derivatives are required by the HESSIAN method. To avoid tedium and error, we do not provide analytic expressions for the derivatives. Instead, we give derivatives of primitive functions and show how to combine them using Fáa di Bruno's formula, a generalization of the chain rule to higher derivatives, to compute exact derivatives of $\psi(x_{d,i})$. It gives derivatives as

$$\frac{d^n}{dx^n} f(g(x)) = \sum \frac{n!}{m_1! m_2! \cdots m_n!} f^{(m_1+\cdots+m_n)}(g(x)) \prod_{j=1}^{n} \left(\frac{g^{(j)}(x)}{j!} \right)^{m_j}.$$

For $s_{d,i} = 0$, all derivatives of $\psi(x_{d,i})$ are equal to zero. For $s_{d,i} = 1$, we use the closed from representation of $p_1(y_{d,i} \,|\, x_{d,i})$ given in equation (3.9) to the compute $\psi(x_{d,i})$. Conditional on the component indicator $z_{d,i}$, this gives

$$\begin{aligned}
\psi(x_{d,i}) &= \log \left[\frac{1}{2} \sum_{k=1}^{z_{d,i}} \alpha_{z_{d,i},k} \Big[\exp\left(-\lambda_{z_{d,i},k} e^{-x_{d,i}} (y_{d,i} - 1)\right) - \exp\left(-\lambda_{z_{d,i},k} e^{-x_{d,i}} (y_{d,i} + 1)\right) \Big] \right] \\
&= \log \left[\frac{1}{2} \sum_{k=1}^{z_{d,i}} \alpha_{z_{d,i},k} \left[g_k^-(x_{d,i}) - g_k^+(x_{d,i}) \right] \right]
\end{aligned}$$

where $g_k^{\pm}(x_{d,i}) = \exp(h_k^{\pm}(x_{d,i}))$ with $h_k^{\pm} = -\lambda_{z_{d,i},k} e^{-x_{d,i}} (y_{d,i} \pm 1)$ for $k = 1, \ldots, z_{d,i}$. We compute derivatives of $\psi(x_{d,i})$ bottom up. The steps are

1. Compute $h_k^{\pm}(x_{d,i})$ and its first five derivatives with respect to $x_{d,i}$, $k = 1, \ldots, z_{d,i}$:

$$\begin{aligned}
h_k^{\pm(1)}(x_{d,i}) &= h_k^{\pm(3)}(x_{d,i}) = h_k^{\pm(5)}(x_{d,i}) = -h_k^{\pm}(x_{d,i}) \\
h_k^{\pm(2)}(x_{d,i}) &= h_k^{\pm(4)}(x_{d,i}) = h_k^{\pm}(x_{d,i}).
\end{aligned}$$

2. Compute $g_k^{\pm}(x_{d,i})$ and first five derivatives with respect to $x_{d,i}$, using Fáa di Bruno's rule, for the exponential function composed with $h_k^{\pm}(x_{d,i})$.

3. Compute $p_1(y_{d,i} \,|\, x_{d,i}, z_{d,i})$ and first five derivatives with respect to $x_{d,i}$, using

$$\frac{\partial^{(r)}}{\partial x_{d,i}^{(r)}} p_1(y_{d,i} \,|\, x_{d,i}, z_{d,i}) = \sum_{k=1}^{z_{d,i}} \alpha_{z_{d,i},k} \Big[g_k^{-(r)}(x_{d,i}) - g_k^{+(r)}(x_{d,i}) \Big], \qquad r = 1, \ldots, 5.$$

4. Compute $\psi(x_{d,i})$ and first five derivatives with respect to $x_{d,i}$, using Fáa di Bruno's rule, for the logarithmic function composed with $p_1(y_{d,i} \,|\, x_{d,i}, z_{d,i})$, treated as a function of $x_{d,i}$. The first five derivatives of $q(z) = \ln z$ are

$$q'(z) = z^{-1}, \quad q''(z) = -z^{-2}, \quad q'''(z) = 2z^{-3}, \quad q^{(4)}(z) = -6z^{-4}, \quad q^{(5)} = 24z^{-5}(z).$$

C.2 Drawing artificial observations

Here we show how to draw observations from their conditional distribution $p(y \,|\, s, x, \beta, \pi)$. For the latent state process used for the Getting it right experiment (see equation (3.15)), there is no relationship between the duration and the dynamic of the latent state. The prior distribution also implies a fairly flat diurnal pattern, thus very little time variation in the unconditional mean of the latent state process. In this case, updating the vector of observations can be done efficiently with the following Metropolis-Hastings algorithm:

- For $d = 1, \ldots, D$, $i = 1, \ldots, n_d$, draw a proposal $y_{d,i}^* \sim p(y_{d,i} \,|\, s_{d,i}, x_{d,i}, \beta, \pi)$.
- For $d = 1, \ldots, D$, set $y_{d,0}^* = 0$ and construct the corresponding transaction times $t_{d,i}^* = \sum_{k=0}^{i-1} y_{d,k}^*$ to evaluate the B-spline function $m(t)$.
- Accept the proposal y^* with probability

$$\min \left\{ 1, \prod_{d=1}^{D} \prod_{i=1}^{n_d - 1} \frac{p(x_{d,i+1} \mid x_{d,i}, t_{d,i}, t_{d,i-1}, \sigma, \rho, \delta)}{p(x_{d,i+1} \mid x_{d,i}, t_{d,i}^*, t_{d,i-1}^*, \sigma, \rho, \delta)} \right\}.$$