UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

ON STATIONARY MARKOV PROCESSES WITH NON-INDEPENDENT INCREMENTS

TESIS

QUE PARA OPTAR POR EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

MICHELLE ANZARUT CHACALO

TUTOR PRINCIPAL

DR. RAMSÉS H. MENA CHÁVEZ, IIMAS-UNAM

COMITÉ TUTOR

DRA. MARÍA EMILIA CABALLERO ACOSTA, IMATE-UNAM

DR. MATTEO RUGGIERO, UNIVERSITÀ DI TORINO

MÉXICO, D.F., JULIO DE 2017

Abstract

This work studies a class of non-independent-increment, stationary, Markov processes. In particular, we aim at constructing processes in such a class with tractable transition probabilities, and with given but arbitrary invariant distributions. Designed with these features, the resulting processes provide several advantages. The non-independence of the increments presents a realistic approach in many fields of application where empirical data suggest the existence of complex dependence structures. The stationarity and Markovianity force the processes to a stable behavior that allows for prediction. The full control of the transition densities leads to efficient simulation and estimation methods, while the control over the invariant distributions provides flexibility and a clear trade-off between marginal and conditional properties, since one can have different dependence structures in a model while retaining the same stationary distribution. All of these advantages lead to a powerful class of processes.

The class may be applied in a wide spectrum of contexts. Specifically, we explore its usefulness in the context of stochastic volatility, and in the context of time-dependent density estimation models. The proposed approaches are illustrated with real financial datasets. Both approaches provide a good approximation to the observed data behavior, and prove not to compromise their generality when confronting them with other models available in the literature.

Resumen

Este trabajo estudia una clase de procesos de Markov, estacionarios con incrementos no independientes. En particular, nuestro objetivo es la construcción de procesos dentro de esa clase con probabilidades de transición manejables y con distribuciones invariantes dadas, pero arbitrarias. Diseñados con dichas características, los procesos resultantes proporcionan varias ventajas. La no independencia de los incrementos presenta un enfoque realista en muchos campos de aplicación en los que datos empíricos sugieren la existencia de estructuras complejas de dependencia; la estacionariedad y markovianiedad forzan a los procesos a un comportamiento estable que permite la predicción; el control total de las densidades de transición conduce a métodos de simulación y estimación eficientes, mientras que el control sobre las distribuciones invariantes ofrece flexibilidad y un claro equilibrio entre las propiedades marginales y condicionales, ya que uno puede tener diferentes estructuras de dependencia en un modelo conservando la misma distribución estacionaria. Todas estas ventajas conducen a una poderosa clase de procesos.

La clase se puede aplicar en un amplio espectro de contextos. Específicamente, se explora su utilidad en el contexto de volatilidad estocástica y en el contexto de modelos de estimación de densidades que varían en el tiempo. Los enfoques propuestos se ilustran con conjuntos de datos financieros reales. Ambos enfoques proporcionan una buena aproximación al comportamiento observado en los datos y demuestran no comprometer su generalidad al enfrentarlos con otros modelos disponibles en la literatura.

Contents

1 Introduction

The assumption of independent increments is a dominant characteristic within many stochastic modeling applications. However, in certain situations, empirical data suggest the existence of more complex dependence structures. For this reason, in various areas of application it is of interest to consider dependent increment processes that, while modeling better, remain statistically tractable. Diverse stability assumptions help to make this possible. Throughout this project, we chose to focus on non-independent-increment processes that are stationary and Markovian. In particular, we aim at constructing processes in such a class with tractable transition probabilities, and with given but arbitrary invariant distributions. This document presents the research advances we have achieved in such a class of processes and their applications.

The theory on stable behaviors of processes has been developed almost exclusively for Markov processes and, most importantly, the Markov property assures we can make predictions for the future of the process based solely on its present state. However, in some parts of the thesis we will relax this property, and discuss ways to deal with the prediction of the resulting processes, which need to have another kind of stability compensating the non-Markovianity. The stationarity assumption assures the random processes will not change their probabilistic structure with time. Hence, a long-enough path allows for prediction. With these two premises, a natural starting point is to look for transition mechanisms that retain a particular distribution of interest invariant over time. In fact, all Markovian process with a unique invariant distribution are stationary, and some of the drawbacks typically supporting the use of non-stationary models can be outmatched by increasing the flexibility of the invariant distribution.

Constructing Markov models with prescribed invariant distributions has been achieved repeatedly in the literature. A popular construction procedure is using a thinning argument, which implies decomposing the random variable at a fixed time as a thinned version of the immediate past plus an innovation term. For instance, Barndorff-Nielsen and Shephard (2001) use the property of self-decomposability to attain such thinning, characterizing a class of continuous-time stationary models termed Ornstein-Uhlenbeck type

processes. McKenzie (2003) uses thinning operators to define discrete-valued time series with Poisson, negative binomial, and geometric marginals. Joe (1996) and Jørgensen and Song (1998) present discrete-time Markov processes with invariant distributions in the infinitely divisible convolution-closed exponential family, and Bibby et al. (2005) explore one-dimensional stationary diffusion processes of the mean reverting type, by specifying a particular form of the diffusion coefficient modulating the process.

The above constructions are very appealing in statistical applications but, in most cases, no general expression for the transition probabilities is available, particularly in the continuous-time setting. However, full control of the transition probabilities driving a Markov process is always a desirable feature as estimation, simulation and prediction procedures become accessible. Other constructions leading to tractable transition probabilities include the one developed by Pitt et al. (2002), where stationary first order autoregressive processes with arbitrary marginal distributions are defined based on a Gibbs sampler representation, Contreras-Cristan et al. (2009); Mena and Walker (2005), where autoregressive processes are produced combining the idea of Pitt et al. (2002) with a non-parametric approach for modeling the transition density, and Mena and Walker (2009) where a continuous-time stationary Markov model is constructed based also on Pitt et al. (2002) latent representation of the transition probabilities.

In this work, we explore two simple yet powerful methods to construct stationary Markov processes with tractable transition probabilities and given invariant distributions. The first one arises from an extension of Harris recurrent Markov chains to continuous time being stationary, Feller, and Harris recurrent. Therefore, we named the resulting subclass of processes as stationary-Feller-Harris processes (SF-Harris processes). Most of the work on SF-Harris processes can be also found in Anzarut and Mena (2016). The second construction is based on a Poisson-type transform modulating the dependence structure, so we named this subclass as f-stationary Poisson-driven Markov processes, where f refers to the stationary density. Part of the work on f-stationary Poisson-driven Markov processes can be found in Anzarut et al. (2017).

The thesis is divided in six chapters. Chapter 2 defines some concepts and properties that are used along the work. The first part of the chapter deals with stability properties, while the second one treats distributions that are utilized either in the construction of SF-Harris and Poisson-driven Markov processes, or as invariant distributions to such processes in specific scenarios.

Chapter 3 directs the attention to the study of SF-Harris processes. The chapter starts with a motivation for such processes definition, it continues with the study of interesting mathematical and statistical properties, and it ends with the development of inference procedures, which are illustrated for some invariant distributions, and proved with robust simulation tests.

Chapter 4 features a first application of SF-Harris processes. A new stochastic volatility model, where the spot volatility follows a SF-Harris process, is proposed. The proposal allows for a simple tradeoff between independence and full dependence and, at the same time, possesses several stability properties and distributional flexibility.

A second application of SF-Harris processes, this time to a time-dependent density model, is presented over Chapter 5. The processes are used to model the particles and weights of a Bayesian nonparametric mixture model. The result is a simple and flexible strategy for time-dependent density estimation, allowing the construction of a great variety of models while preserving computational simplicity.

Chapter 6 focuses on f-stationary Poisson-driven Markov processes. Over this chapter, the processes definition is motivated, mathematical properties are studied, inference methods are thoroughly developed and tested, and a number of possible applications are presented. In the discrete-time case, special attention is placed to generalized inverse Gaussian stationary densities. In the continuous case, the focus is on Gamma distributions, which are then extended to cover other invariant distributions, such as the Generalized Extreme Value class.

We conclude the thesis in Chapter 7 with some final remarks about our current work on each of these topics, as well as future research directions. To perform all the reported calculations we used the **R** project for statistical computing.

2 Some concepts and properties

In this chapter, we present certain definitions that are used throughout the work, and may vary widely depending on the consulted literary source. First, we give a minimal description of a number stability properties. Afterwards, we introduce some probability distributions used either in the construction of stationary Markov processes with prescribed invariant distributions, or as invariant distributions of such constructed processes.

2.1 Stability properties

With 'stability properties' we mean a range of similar but very different conditions that assure stable behaviors of processes. The stability (or instability) of a process has both important performance engineering implications and fundamental mathematical consequences. Over the section we assume that $W = (W_t)_{t\geq 0}$ is a time-homogeneous Markov process with state space $(E, \mathcal{E})$, transition probabilities $P_t(w, A) = \mathbb{P}(W_t \in A | W_0 = w)$, and starting point $W_0 \sim \mu$, where μ denotes the initial distribution. Additionally, we assume that the state space E is a locally compact and separable metric space, and that $\mathcal{E}$ is the Borel field on E. Let us begin by recalling the definition of a Lévy process. For more on the theory of Lévy processes we refer to Bertoin (1998); Kyprianou (2006).

Definition 1. *The process W is said to be a **Lévy process** if it possesses the following properties:*

(i) The paths of W are almost surely right continuous with left hand limits.

(ii) It has stationary increments, meaning that for $0 \leq s \leq t$, $W_t - W_s$ is equal in distribution to W_{t-s}.

(iii) It has independent increments, meaning that for $0 \leq s \leq t$, $W_t - W_s$ is independent of $\{W_u : u \geq s\}$.

A Markov process possessing independent increments is said to be *spatially homogeneous.*

Much of the application of continuous time Markov processes is confined to the subclass of Lévy processes. The popularity of the Lévy class is in part due to the existing mechanisms to deal with them. Their theory is documented on thousands of articles; methods have been proposed that permit one to simulate paths of any Lévy process almost to an arbitrary level of accuracy (see e.g. Asmussen and Rosiński, 2001; Tankov, 2003); and various studies are devoted to their statistical estimation (see e.g. Comte et al., 2010; Figueroa-López and Houdré, 2006; Neumann et al., 2009; Nolan, 2003).

Nevertheless, on occasions, empirical data or theoretical considerations suggest the phenomena have a different dependence structure. Böttcher (2010) documents this issue well, providing examples in areas such as hydrology, geology, and mathematical finance, where state space dependent models provide a better fit. A natural extension to Lévy processes are Feller processes, which are Markov processes that include Lévy processes as a special case, but need not be spatially homogeneous.

Definition 2. *The* ***semigroup operator*** $\mathsf{T_t}$ *associated to the probability kernel* P_t *is given by*

$$\mathsf{T}_t f(x) = (\mathsf{T}_t f)(x) = \int P_t(x, dy) f(y),\ x \in E,\ t \geq 0.$$

where $f : E \to \mathbb{R}$ *is assumed to be measurable and either bounded or non-negative.*

Definition 3. *The process* W *is said to be a* ***Feller process*** *if its semigroup operator* T *meets the following conditions:*

(i) For all $t \geq 0$, $\mathsf{T_t}$ *maps* C_0 *into itself.*

(ii) For any $f \in C_0$ *and* $x \in E$, $\mathsf{T_t} f(x) \to f(x)$ *when* $t \downarrow 0$.

Where C_0 *denotes the space of real-valued continuous functions that vanish at infinity.*

The Chapman-Kolmogorov relation

$$P_{t+s}(x, A) = \int P_s(x, dy) P_t(y, A),$$

met for any $x \in E$ and $A \in \mathcal{E}$ by every Markov process, turns into the semigroup property $\mathsf{T}_s \mathsf{T}_t = \mathsf{T}_{s+t}$. This suggests a formal representation $\mathsf{T}_t = e^{tA}$ in terms of a generator A. Under the two Feller conditions we can define a generator A that describes the infinitesimal evolution of the underlying process W.

Definition 4. *The* ***infinitesimal generator*** *of a Feller process with semigroup* T *is given by the uniform limit*

$$\mathsf{A}f = \lim_{t \downarrow 0} \frac{\mathsf{T}_t - f}{t}.$$

A function $f \in C_0$ is said to be in the domain of the generator if the uniform limit exists in C_0.

Definition 5. *The* ***resolvent of order*** $\boldsymbol{\lambda}$ *of a Feller process with semigroup* T *is given by*

$$\mathsf{R}_\lambda f(x) = \int_0^\infty e^{-\lambda t} \mathsf{T}_t f(x) dt, \ x \in E.$$

These two key operators uniquely determine the semigroup. Over the work we will see which processes in our proposed construction are Feller, and some interesting implications this entails. For more on Feller processes see, e.g., Böttcher et al. (2013); Kallenberg (2002); Liggett (2010).

Let us now turn the attention to other stability properties, for which our main source is the book of Meyn and Tweedie (2012).

Definition 6. *An* ***invariant (or stationary) probability distribution*** *is a distribution* μ *over* $(E, \mathcal{E})$ *satisfying* $\int P_t(y, A)\mu(dy) = \mu(A)$ *for all* $A \in \mathcal{E}$ *and for all* $t \geq 0$.

Overall, μ does not have to be unique but, if it is, then is the *limit distribution*, meaning that $P_t(w, A) \to \mu(A)$ when $t \to \infty$ for all $A \in \mathcal{E}$. In general, it is difficult to find invariant distributions, and the following definition makes the work easier.

Definition 7. *W is* ***μ-reversible*** *if for all* $A, A' \in \mathcal{E}$ *and for all* $t \geq 0$ *it follows that* $\int_A P_t(y, A')\mu(dy) = \int_{A'} P_t(y, A)\mu(dy)$.

If a process is μ-reversible then, in particular, for any $A \in \mathcal{E}$ it follows that

$$\int P_t(y, A)\mu(dy) = \int_A P_t(y, E)\mu(dy) = \int_A \mu(dy) = \mu(A).$$

Hence, μ is an invariant distribution.

Definition 8. *W is **strongly stationary** when all its finite-dimensional distributions are invariant under translations of the indices. This is, if $t_1, ..., t_k \geq 0$, $A_1, ..., A_k \in \mathcal{E}$, and $\tau \geq 0$, then,*

$$\mathbb{P}(W_{t_1} \in A_1, ..., W_{t_k} \in A_k | W_0 = w) = \mathbb{P}(W_{t_1+\tau} \in A_1, ..., W_{t_k+\tau} \in A_k | W_\tau = w) \text{ for all } w \in E.$$

In this work, when we talk about stationary processes, we mean stationary in the strong sense. We will also assume that when a process has a stationary distribution, this one matches the initial distribution. As a consequence, any homogeneous Markov process with stationary distribution is stationary (Kallenberg, 2002, Theorem 7.11, page 25). Hence, it is simple to verify that the process has stationary increments.

Definition 9. *Let $\tau_B = \inf\{t \geq 0 : W_t \in B\}$. W is **μ-irreducible** if there exists a measure μ over $(E, \mathcal{E})$ such that for any $B \in \mathcal{E}$, $\mu(B) > 0$ implies*

$$\mathbb{P}(\tau_B < \infty | W_0 = w) > 0 \text{ for all } w \in E. \tag{2.1}$$

Moreover, we say that a process is recurrent if it is μ-irreducible and the average time spent on any $B \in \mathcal{E}$ is infinite.

Definition 10. *A Markov process W is **recurrent** if for all $B \in \mathcal{E}$ and $w \in E$,*

$$\mathbb{E}\left[\int_0^\infty \delta_{W_t}(B)dt \,\middle|\, W_0 = w\right] = \infty \text{ for all } w \in E. \tag{2.2}$$

The notion of recurrence can become even stronger by asking, rather than an infinite number of expected visits to any state, an infinite number of visits almost surely.

Definition 11. *A Markov process W is **Harris recurrent** if there exists a σ-finite measure μ in $(E, \mathcal{E})$ such that for any $B \in \mathcal{E}$, $\mu(B) > 0$ implies*

$$\mathbb{P}\left\{\int_0^\infty \delta_{W_t}(B)dt = \infty | W_0 = w\right\} = 1 \text{ for all } w \in E. \tag{2.3}$$

Getoor (1980) proves that every Harris recurrent process has a unique (up to constant multiples) invariant measure. A Harris recurrent process with finite invariant measure is called *positive Harris recurrent*. Note that the finite invariant measure can be normalized, so it reduces to the case of an invariant probability measure.

Let us denote by $\mu f = \int f(x)\mu(dx)$ for a measure μ and any measurable function $f : E \to \mathbb{R}$, then, when a process W is positive Harris recurrent with invariant distribution μ, it follows that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t g(W_s)ds \to \mu g$$

for any function $g : E \to \mathbb{R}$ such that $\mu|g| < \infty$ (see Azéma et al., 1969). The convergence can be made stronger by asking the process to be ergodic.

Definition 12. *A process is* ***ergodic*** *if it has an invariant probability measure μ and*

$$\lim_{t\to\infty} \left\{ \sup_{A\in\mathcal{E}} |P_t(w, A) - \mu(A)| \right\} = 0$$

The above result points out that the process will eventually converge to the invariant measure, but it does not detail the speed of convergence. A common measure used to analyze the speed of convergence is the order of the *total variation distance*, given by $\sup_{A\in\mathcal{E}} |P_t(w, A) - \mu(A)|$.

Definition 13. *Let μ be the invariant probability measure, if*

$$\sup_{A\in\mathcal{E}} |P_t(w, A) - \mu(A)| \leq M(w)\rho^t$$

for some finite function M and $\rho < 1$, we say that the process is ***geometrically (or exponentially) ergodic*** *with rate ρ. Furthermore, if M is bounded, we say that the process is* ***uniformly ergodic****.*

The term geometrically ergodic is usually applied for discrete-time processes while the term exponentially ergodic is used for continuous-time processes.

Finally, let us mention a few words about regeneration. Harris recurrence implies the existence of regeneration sets, so it is a type of regeneration that can occur in a process. More generally, a stochastic process is *regenerative in the classical sense* if there are random times where it starts afresh independently of the past, like a recurrent Markov

chain at the times of visits to a fixed state. The regeneration times form a renewal process, and split the stochastic process into a sequence of cycles that are independent and identically distributed. *Wide-sense regeneration* allows the future after regeneration to depend on the past, as long as it does not depend on the past regeneration times. In this case, the cycles are not necessarily independent. For more on the theory of regeneration we refer to Asmussen (2003); Sigman and Wolff (1993).

Contentions between the stability properties

We mentioned that Lévy process are a subclass of Feller processes, which themselves are a subclass of Markov processes. Regarding the other stability properties defined in Section 2.1 we have that, in continuous time, there is not a clear order between them. In discrete time, positive Harris recurrence is equivalent to wide-sense regeneration. In particular, when the chains are Markovian, these will additionally imply strong stationarity and, as a consequence, ergodicity. Therefore, a Venn diagram of the discrete-time properties in a Markovian setting is easy to picture (see Figure 2.1). The classification becomes very confuse as we move towards the continuous-time setting.

Some authors have managed to make such a classification assuming additional conditions. For example, Kunita (1997) proves that a strongly Feller irreducible process has at most one invariant measure and, when it has an invariant measure, such a measure is ergodic. A process with generator T is called *strongly Feller* if for every bounded measurable function f, $\mathsf{T}_t f$ is bounded and continuous. Strong Feller continuity is an interesting property, needed in many applications. However, it seems to be too restrictive.

Meyn and Tweedie (1993) identify a number of connections between the stability properties of a process and its *skeleton chains*, which they define as the chains sampled from the process at the times $(t_k)_{k\in\mathbb{N}}$ of a renovation process. For example, they prove that if a process is positive Harris recurrent with invariant probability measure μ, then the process is ergodic if and only if some skeleton chain is irreducible. Consequently, there are positive Harris recurrent processes which are not ergodic.

We can also prove that a process can be strongly stationary without being ergodic. For instance, if W is the strongly stationary process with invariant probability measure μ defined by the transition functions $P_t(w, A) = \delta_w(A)$ with $W_0 \sim \mu$, then W is not ergodic

since

$$\lim_{t\to\infty}\left\{\sup_{A\in\mathcal{E}}|P_t(w,A)-\mu(A)|\right\}=\sup_{A\in\mathcal{E}}|\delta_w(A)-\mu(A)|\neq 0.$$

On the other hand, an *autoregresive process of the first order* $W=(W_n)_{n\in\mathbb{N}}$, defined as

$$W_n=aW_{n-1}+\epsilon_n,\ W_0=w,\ a\in(-1,1),$$

with independent identically distributed standard Gaussian random variables $(\epsilon_n)_{n\in\mathbb{N}}$, is not strongly stationary since by induction, one easily verifies that

$$W_n=a^nx+\sum_{i=0}^{n-1}a^iW_{n-i}\sim N\left(a^nx,\frac{1-a^{2n}}{1-a^2}\right).$$

Therefore, its mean is dependent on n. The n-step transition probability is given by

$$P^n(w,A)=N\left(a^nx,\frac{1-a^{2n}}{1-a^2}\right).$$

As $n\to\infty$, $P^n(w,A)\to\mu$ in total variation, where

$$\mu=N\left(0,\frac{1}{1-a^2}\right)$$

is the unique stationary distribution. However, the convergence is geometric, but not uniform in w, since

$$\lim_{t\to\infty}\left\{\sup_{A\in\mathcal{E}}|P^n(w,A)-\mu(A)|\right\}=1$$

for any $n\in\mathbb{N}$, so the process is geometrically ergodic.

Wide-sense regenerative processes are also not necessarily stationary or ergodic, and the question about the equivalence between positive Harris recurrence and wide-sense regeneration has not been answered in continuous time (Glynn, 2011). We believe that a Venn diagram in continuous time would look as Figure 2.2. However, to fully understand how restrictive we are in choosing one or another stability property, a deep mathematical study of the contentions between these properties in continuous time is still missing in

the literature. Let us finish the section by recalling that in this work we are aiming at constructing a class of statistically tractable non-independent increment processes. Stability properties make the tractability of a process possible, and become important when working on the inference and application to real data sets.

2.2 On some probability distributions

This section deals with probability distributions that are used later on in the construction of stochastic processes or as invariant distributions in some scenarios. The common distributions, such as the Gamma, Inverse Gamma, Exponential, etc., are set out to clarify the parameterization we will use in the thesis, however, to dig into their theoretical properties we refer to Ross (2014). The more unconventional families of Generalized Inverse Gaussian and Generalized Extreme Value distributions are defined, along with a number of useful properties. Also, the Poisson Weighted Density is introduced.

• The *discrete Uniform distribution* with parameters $k_1, k_2, ..., k_n$ for some $n > 0$, is a discrete probability distribution with probability mass function given by

$$\mathsf{U}(k_i|k_1, k_2, ..., k_n) = \frac{1}{n},\ i \in \{1, ..., n\}$$

Therefore, a finite number of outcomes are equally likely to happen.

• The *continuous Uniform distribution* with parameters $a, b \in \mathbb{R}$, $a < b$, is a continuous probability distribution with density function given by

$$\mathsf{U}(x|a, b) = \frac{1}{b-a},\ x \in [a, b].$$

Hence, all intervals of the same length on the support are equally probable.

• The *Normal (or Gaussian) distribution* with parameters $\mu \in \mathbb{R}$, $\sigma^2 > 0$, is a continuous probability distribution with density function given by

$$\mathsf{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},\ x \in \mathbb{R}.$$

• The *Inverse Gaussian distribution* with parameters $\lambda, \mu > 0$, is a continuous probability

distribution with density function given by

$$\mathsf{IG}(x|\lambda,\mu) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right\},\ x>0.$$

• The *Gamma distribution* with parameters $a, b > 0$, is a continuous probability distribution with density function given by

$$\mathsf{Ga}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx},\ x>0.$$

• Let p be a positive integer. The *Wishart distribution* with parameters $\mathbf{V} \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$ positive definite, and $n > p-1$, is a continuous probability distribution with density function given by

$$\mathsf{W}_\mathsf{p}(\mathbf{X}|\mathbf{V},n) = |\mathbf{X}|^{(n-p-1)/2} \exp\left\{\mathsf{tr}(\mathbf{V}^{-1}\mathbf{X})/2\right\} \left\{2^{np/2}|\mathbf{V}|^{n/2}\Gamma_\mathsf{p}\left(\frac{n}{2}\right)\right\}^{-1},$$

where $\mathbf{X} \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$ is a positive definite matrix, Γ_p is the multivariate Gamma function, $|\mathbf{X}|$ denotes the determinant of $\mathbf{X}$, and tr denotes the trace function.

The Wishart is in fact a generalization to multiple dimensions of the Gamma distribution.

• The *Exponential distribution* with parameter $\lambda > 0$, is a continuous probability distribution with density function given by

$$\mathsf{Exp}(x|\lambda) = \lambda e^{-\lambda x},\ x>0.$$

Notice that $\mathsf{Ga}(x|1,\lambda) = \mathsf{Exp}(x|\lambda)$.

• The *Inverse Gamma distribution* with parameters $a, b > 0$, is a continuous probability distribution with density function given by

$$\mathsf{IGa}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-bx^{-1}},\ x>0.$$

It is easy to see that if $X \sim \mathsf{Ga}(a,b)$, then $X^{-1} \sim \mathsf{IGa}(a,b)$.

• The *Generalized inverse Gaussian* (GIG) is a three-parameter family of continuous

probability distributions with density function

$$\mathsf{GIG}(x|\lambda,\delta,\gamma) = \frac{1}{2K_\lambda(\sqrt{\delta\gamma})}\left(\frac{\gamma}{\delta}\right)^{\lambda/2} x^{\lambda-1}\exp\left\{-\frac{1}{2}(\delta x^{-1}+\gamma x)\right\},\ x>0,$$

where $\lambda\in\mathbb{R}$, $(\lambda,\gamma)\in\Theta_\lambda$,

$$\Theta_\lambda = \begin{cases} \delta\geq 0, \gamma>0, & \text{if } \lambda>0,\\ \delta>0,\gamma>0, & \text{if } \lambda=0,\\ \delta>0,\gamma\geq 0, & \text{if } \lambda<0,\end{cases}$$

and K_ν is the modified Bessel function of the third kind with index ν.

Some distributions are special cases of the GIG family, for example,

$$\mathsf{GIG}(x|\lambda,0,\gamma)\propto x^{\lambda-1}\exp\left\{-\frac{1}{2}\gamma x\right\}.$$

Therefore, when $\lambda>0$, $\delta=0$, and $\gamma>0$, we have that X has a $\mathsf{Ga}(x|\lambda,\gamma/2)$ density. Now,

$$\mathsf{GIG}(x|\lambda,\delta,0)\propto x^{\lambda-1}\exp\left\{-\frac{1}{2}\delta x^{-1}\right\}.$$

Hence, if $\gamma=0$ and $\lambda<0$ then $\delta>0$ and we get the case of a $\mathsf{IGa}(x|\lambda,\delta/2)$ density. Also, when $\lambda=-1/2$ and $\delta,\gamma>0$, we have that

$$\begin{aligned}\mathsf{GIG}\left(x\mid -\frac{1}{2},\delta,\gamma\right) &\propto x^{-3/2}\exp\left\{-\frac{1}{2}(\delta x^{-1}+\gamma x)\right\}\\ &= x^{-3/2}\exp\left\{-\frac{1}{2x}\gamma\left(\delta\gamma^{-1}+x^2\right)\right\}\\ &\propto x^{-3/2}\exp\left\{-\frac{1}{2x}\gamma\left(\sqrt{\delta\gamma^{-1}}+x\right)^2\right\},\end{aligned}$$

which is the kernel of a $\mathsf{IG}\left(x|\gamma,\sqrt{\delta\gamma^{-1}}\right)$ random variable.

The GIG class is very flexible and allows to obtain various explicit results. For instance,

let us compute the moment generating function.

$$\begin{aligned}
\mathsf{M}_{\mathsf{GIG}(\lambda,\delta,\gamma)}(u) &= \int_0^\infty e^{ux}\left[\frac{1}{2K_\lambda(\sqrt{\delta\gamma})}\left(\frac{\gamma}{\delta}\right)^{\lambda/2} x^{\lambda-1}\exp\left\{-\frac{1}{2}(\delta x^{-1}+\gamma x)\right\}\right]dx \\
&= \frac{1}{2K_\lambda(\sqrt{\delta\gamma})}\left(\frac{\gamma}{\delta}\right)^{\lambda/2}\int_0^\infty x^{\lambda-1}\exp\left[-\frac{1}{2}\{\delta x^{-1}+(\gamma-2u)x\}\right]dx \\
&= \frac{1}{2K_\lambda(\sqrt{\delta\gamma})}\left(\frac{\gamma}{\delta}\right)^{\lambda/2} 2K_\lambda\left\{\sqrt{\delta(\gamma-2u)}\right\}\left(\frac{\delta}{\gamma-2u}\right)^{\lambda/2} \\
&= \frac{K_\lambda\left\{\sqrt{\delta(\gamma-2u)}\right\}}{K_\lambda(\sqrt{\delta\gamma})}\left(\frac{\gamma}{\gamma-2u}\right)^{\lambda/2},
\end{aligned}$$

With the restriction $2u<\gamma$. It follows that GIG distributions have moments of any order, in particular, if we take the first two derivatives with respect to u, and evaluate them at $u=0$ we obtain that for $X\sim\mathsf{GIG}(\lambda,\delta,\gamma)$,

(i) $\mathbb{E}[X]=\dfrac{K_{\lambda+1}(\sqrt{\delta\gamma})}{K_\lambda(\sqrt{\delta\gamma})}\left(\dfrac{\delta}{\gamma}\right)^{1/2}$,

(ii) $\mathsf{Var}(X)=\left[\dfrac{K_{\lambda+2}(\sqrt{\delta\gamma})}{K_\lambda(\sqrt{\delta\gamma})}-\left(\dfrac{K_{\lambda+1}(\sqrt{\delta\gamma})}{K_\lambda(\sqrt{\delta\gamma})}\right)^2\right]\dfrac{\delta}{\gamma}$.

In the following theorem we compute the expected value of some transformations of a random variable with GIG distribution. These transformations become important because the Kullback-Leibler divergence depends on them and, when making the estimation of the parameters of a GIG distribution, we use the divergence as a measure of the amount of information lost by working with the estimated model instead of the real one.

Theorem 2.2.1. *Let $X\sim\mathsf{GIG}(\lambda,\delta,\gamma)$, then*

(i) $\mathbb{E}\left[X^{-1}\right]=\dfrac{K_{\lambda+1}(\sqrt{\delta\gamma})}{K_\lambda(\sqrt{\delta\gamma})}\left(\dfrac{\gamma}{\delta}\right)^{1/2}-\dfrac{2\lambda}{\delta}$,

(ii) $\mathbb{E}[\log X]=\dfrac{1}{2}\log\left(\dfrac{\delta}{\gamma}\right)+\dfrac{\dfrac{d}{d\lambda}K_\lambda(\sqrt{\delta\gamma})}{K_\lambda(\sqrt{\delta\gamma})}$.

Proof. Let $Z = X^{-1}$, then denoting with $\hat{f}$ to the distribution function of Z we have that

$$\begin{aligned}\hat{f}(z) &\propto z^{-(\lambda-1)} \exp\left\{-\frac{1}{2}(\delta z + \gamma z^{-1})\right\} z^{-2} \\ &= z^{-\lambda+1} \exp\left\{-\frac{1}{2}(\delta z + \gamma^{-1})\right\},\end{aligned}$$

implying $Z \sim \mathsf{GIG}(-\lambda, \gamma, \delta)$. Using properties of the Bessel function it follows that

$$\begin{aligned}\mathbb{E}[Z] &= \frac{K_{-\lambda+1}(\sqrt{\delta\gamma})}{K_{-\lambda}(\sqrt{\delta\gamma})} \left(\frac{\gamma}{\delta}\right)^{1/2} \\ &= \frac{K_{\lambda-1}(\sqrt{\delta\gamma})}{K_{\lambda}(\sqrt{\delta\gamma})} \left(\frac{\gamma}{\delta}\right)^{1/2} \\ &= \frac{K_{\lambda+1}(\sqrt{\delta\gamma}) - \frac{2\lambda}{\sqrt{\delta\gamma}} K_{\lambda}(\sqrt{\delta\gamma})}{K_{\lambda}(\sqrt{\delta\gamma})} \left(\frac{\gamma}{\delta}\right)^{1/2} \\ &= \frac{K_{\lambda+1}(\sqrt{\delta\gamma})}{K_{\lambda}(\sqrt{\delta\gamma})} \left(\frac{\gamma}{\delta}\right)^{1/2} - \frac{2\lambda}{\delta}.\end{aligned}$$

Let us consider now the random variable $W = \log X$,

$$\begin{aligned}\mathbb{E}\left[e^{uW}\right] &= \mathbb{E}[X^u] \\ &= \frac{1}{2K_{\lambda}(\sqrt{\delta\gamma})} \left(\frac{\gamma}{\delta}\right)^{\lambda/2} \int_0^{\infty} x^{\lambda+u-1} \exp\left\{-\frac{1}{2}(\delta x^{-1} + \gamma x)\right\} dx \\ &= \frac{1}{2K_{\lambda}(\sqrt{\delta\gamma})} \left(\frac{\gamma}{\delta}\right)^{\lambda/2} 2K_{\lambda+u}(\sqrt{\delta\gamma}) \left(\frac{\delta}{\gamma}\right)^{(\lambda-u)/2} \\ &= \frac{K_{\lambda+u}(\sqrt{\delta\gamma})}{K_{\lambda}(\sqrt{\delta\gamma})} \left(\frac{\delta}{\gamma}\right)^{u/2}.\end{aligned}$$

Therefore,

$$\mathbb{E}[W] = \frac{\frac{d}{d\lambda} K_{\lambda}(\sqrt{\delta\gamma})}{K_{\lambda}(\sqrt{\delta\gamma})} + \frac{1}{2}\log\left(\frac{\delta}{\gamma}\right).$$

□

The following parametrization is sometimes useful for some GIG family calculations,

specifically for the development of estimation methods. Let us take $\delta = \kappa\eta$ and $\gamma = \kappa/\eta$, then $\sqrt{\delta\gamma} = \kappa$ and $\eta = \sqrt{\delta/\gamma}$ so it follows

$$\mathsf{GIG}(x|\lambda,\kappa,\eta) = \frac{1}{2K_\lambda(\kappa)}\left(\frac{1}{\eta}\right)^{\lambda} x^{\lambda-1}\exp\left\{-\frac{\kappa}{2}\left(\frac{\eta}{x}+\frac{x}{\eta}\right)\right\},\ x>0, \tag{2.4}$$

where $\lambda \in \mathbb{R}$, $(\lambda,\gamma) \in \Theta_\lambda$ with

$$\Theta_\lambda = \begin{cases} \kappa \geq 0, \eta > 0, & \text{if } \lambda \neq 0, \\ \kappa > 0, \eta > 0, & \text{if } \lambda = 0. \end{cases}$$

Making $R_\lambda(\kappa) = K_{\lambda+1}(\kappa)/K_\lambda(\kappa)$, it follows directly that for $X \sim \mathsf{GIG}(\lambda,\kappa,\eta)$,

(i) $\mathbb{E}[X] = R_\lambda(\kappa)\eta$.

(ii) $\mathbb{E}\left[X^{-1}\right] = \eta^{-1}\left(R_\lambda(\kappa) - 2\lambda\kappa^{-1}\right)$.

(iii) $\mathbb{E}[\log X] = \log(\eta) + \dfrac{d}{d\lambda}K_\lambda(\kappa)\left(K_\lambda(\kappa)\right)^{-1}$.

The advantage of this parametrization is that, as it is simple to verify, when $X \sim \mathsf{GIG}(\lambda,\kappa,\eta)$, then $X = \eta Z$, where $Z \sim \mathsf{GIG}(\lambda,\kappa,1)$.

The *Kullback-Leibler divergence* is a non-symmetric similarity measure between two probability distribution functions p and $\hat{p}$ sharing a support $\mathbb{X}$, defined as

$$\mathsf{KL}(p,\hat{p}) = \int_{\mathbb{X}} p(x)\log\frac{p(x)}{\hat{p}(x)}dx.$$

If $p(x) = \mathsf{GIG}(x|\lambda,\kappa,\eta)$ and $\hat{p}(x) = \mathsf{GIG}(x|\hat{\lambda},\hat{\kappa},\hat{\eta})$, it follows that

$$\begin{aligned}\frac{p(x)}{\hat{p}(x)} &= \frac{K_{\hat{\lambda}}(\hat{\kappa})}{K_\lambda(\kappa)}\frac{\hat{\eta}^{\hat{\lambda}}}{\eta^\lambda}x^{\lambda-\hat{\lambda}}\exp\left\{-\frac{\kappa}{2}\left(\frac{\eta}{x}+\frac{x}{\eta}\right)+\frac{\hat{\kappa}}{2}\left(\frac{\hat{\eta}}{x}+\frac{x}{\hat{\eta}}\right)\right\} \\ &= \frac{K_{\hat{\lambda}}(\hat{\kappa})}{K_\lambda(\kappa)}\frac{\hat{\eta}^{\hat{\lambda}}}{\eta^\lambda}x^{\lambda-\hat{\lambda}}\exp\left[-\frac{1}{2}\left\{(\kappa\eta-\hat{\kappa}\hat{\eta})\frac{1}{x}+\left(\frac{\kappa}{\eta}-\frac{\hat{\kappa}}{\hat{\eta}}\right)x\right\}\right].\end{aligned}$$

Table I: KL divergence between $p(x) = \mathsf{GIG}(x|0,3,4)$ and $\hat{p}(x) = \mathsf{GIG}(x|\hat{\lambda},\hat{\kappa},\hat{\eta})$

$\mathsf{KL}(p,\hat{p})$	0.06	0.36	2.48	6.05	49.75
$\hat{\lambda}$	3.5	-1	2	7	50
$\hat{\kappa}$	0.8	2	10	3	20
$\hat{\eta}$	0.6	10	6	4	4

Consequently,

$$\log\frac{p(x)}{\hat{p}(x)} = f(\lambda,\kappa,\eta,\hat{\lambda},\hat{\kappa},\hat{\eta}) + (\lambda-\hat{\lambda})\log x - \frac{1}{2}\left\{(\kappa\eta-\hat{\kappa}\hat{\eta})\frac{1}{x} + \left(\frac{\kappa}{\eta}-\frac{\hat{\kappa}}{\hat{\eta}}\right)x\right\},$$

where $f(\lambda,\kappa,\eta,\hat{\lambda},\hat{\kappa},\hat{\eta}) = \log K_{\hat{\lambda}}(\hat{\kappa}) - \log K_{\lambda}(\kappa) + \hat{\lambda}\log\hat{\eta} - \lambda\log\eta$.

Therefore, taking $X \sim p$,

$$\mathsf{KL}(p,\hat{p}) = f(\lambda,\kappa,\eta,\hat{\lambda},\hat{\kappa},\hat{\eta}) + (\lambda-\hat{\lambda})\mathbb{E}[\log X] - \frac{1}{2}\left\{(\kappa\eta-\hat{\kappa}\hat{\eta})\mathbb{E}\left[X^{-1}\right] + \left(\frac{\kappa}{\eta}-\frac{\hat{\kappa}}{\hat{\eta}}\right)\mathbb{E}[X]\right\}.$$

As an example, we set $\lambda = 0$, $\kappa = 3$, and $\eta = 4$, and compute the KL divergence for different approximations (Table I and Figure 2.3). We can observe that in certain situations, although the parameters are extremely different from the actual, the approximation is good. Because of this, the divergence is used to measure the fit, rather than an established measure for each parameter.

• The *Gumbel distribution* with parameters $\mu \in \mathbb{R}$, $\beta > 0$, is a continuous probability distribution with density function given by

$$\mathsf{Gum}(x|\mu,\beta) = \frac{1}{\beta}\exp\left[-\left\{\frac{x-\mu}{\beta} + \exp\left(-\frac{x-\mu}{\beta}\right)\right\}\right],\ x \in \mathbb{R}.$$

• The *Weibull distribution* with parameters $\lambda, k > 0$, is a continuous probability distribution with density function given by

$$\mathsf{Wei}(x|\lambda,k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp\left\{-\left(\frac{x}{\lambda}\right)^{k}\right\},\ x \geq 0.$$

• The *Fréchet distribution* with parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\alpha > 0$, is a continuous

probability distribution with density function given by

$$\mathsf{Fr}(x|\mu,\sigma,\alpha) = \frac{\alpha}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{-1-\alpha} \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}\right\},\ x > \mu.$$

The Fréchet is also known as the inverse Weibull distribution because when $X \sim \mathsf{Fr}(x|0,\sigma,\alpha)$, $X^{-1} \sim \mathsf{Wei}(x|\alpha,\sigma^{-1})$.

• The *Generalized Extreme Value (GEV) distribution* with parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\nu \in \mathbb{R}$, is a continuous probability distribution, often defined by its cumulative distribution function

$$\bar{\mathsf{G}}(x|\mu,\sigma,\nu) = \exp\left[-\left\{1+\frac{\nu(x-\mu)}{\sigma}\right\}^{-1/\nu}\right],$$

whose support $\mathbb{S} \subset \mathbb{R}$ is given by

$$\mathbb{S} = \begin{cases} \left(-\infty,\ \mu-\frac{\sigma}{\nu}\right), & \text{if } \ \nu < 0, \\ (-\infty,\ \infty), & \text{if } \ \nu = 0, \\ \left(\mu-\frac{\sigma}{\nu},\ \infty\right), & \text{if } \ \nu > 0. \end{cases}$$

We will save the notation $\mathsf{GEV}(x|\mu,\sigma,\nu)$ for the GEV density function with parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\nu \in \mathbb{R}$.

The Gumbel, Fréchet, and Weibull distributions can be derived from the GEV. In fact, the GEV family was developed within extreme value theory to combine the Gumbel, Fréchet and Weibull families also known as type I, II and III extreme value distributions. By the extreme value theorem, the GEV distribution is the only limit distribution of the normalized maxima of a sequence of independent and identically distributed random variables. Hence, the GEV is often used as an approximation to model the maxima of large sets of random variables.

For the Gumbel, notice that

$$\lim_{\nu\to 0} \bar{\mathsf{G}}(x|\mu,\sigma,\nu) = \exp\left\{-\exp\left\{\frac{x-\mu}{\sigma}\right\}\right\},$$

which is the cumulative distribution function of a $\mathsf{Gum}(x|\mu,\sigma)$ distribution. Now, if $X \sim \mathsf{GEV}(x|\mu,\sigma,1/\alpha)$ for some $\alpha > 0$, then $Y = \mu + \sigma + \frac{X-\mu}{\alpha} \sim \mathsf{Fr}(y|\mu,\sigma,\alpha)$ with $y > \mu$.

This implies that, $\tilde{Y} = \sigma + \frac{X}{\alpha} \sim \mathsf{Fr}(y|0,\sigma,\alpha)$, and $\tilde{Y}^{-1} \sim \mathsf{Wei}(x|\alpha,\sigma^{-1})$.

• The *Poisson distribution* with parameter $\lambda > 0$ is a discrete probability distribution with probability mass function given by

$$\mathsf{Po}(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0,1,2,...$$

• Let f be an absolutely continuous probability density function supported on $\mathbb{R}_+$. The *f-Poisson weighted distribution* with parameters $\phi > 0$, $y \in \{0,1,2,...\}$, introduced in Anzarut et al., is a continuous probability distribution with density function given by

$$\hat{f}(x|y,\phi) = \frac{x^y e^{-x\phi} f(x)}{\xi(y,\phi)}, \; x > 0, \tag{2.5}$$

where

$$\xi(y,\phi) := \int_{\mathbb{R}_+} z^y e^{-z\phi} f(z) dz.$$

The density function (2.5) is well defined since $\xi(y,\phi)$ can be seen as a moment of an exponentially tilted positive random variable, which always exists for $\phi > 0$. Moreover, when $\phi \downarrow 0$, the Poisson weighted density reduces to the size-biased density of f and, when $y = 0$, it reduces to the Esscher transform of f. The density (2.5) can also be seen as the posterior density of a Poisson distribution with parameter ϕx, denoted as $\mathrm{Po}(\phi x)$, with prior f on x. This latter aspect, combined with the general idea of weighted distributions introduced by Rao (1965), explains the name attributed to (2.5).

If X has density (2.5), then is direct that the moments of order r are given by

$$\mathbb{E}[X^r] = \int_{\mathbb{R}_+} \frac{x^y x^r e^{-x\phi} f(x)}{\xi(y,\phi)} dx = \frac{\xi(y+r,\phi)}{\xi(y,\phi)}, \tag{2.6}$$

and the Laplace transform is

$$\mathbb{E}[e^{-\lambda X}] = \int_{\mathbb{R}_+} \frac{x^y e^{-x\lambda} e^{-x\phi} f(x)}{\xi(y,\phi)} dx = \frac{\xi(y,\phi+\lambda)}{\xi(y,\phi)}.$$

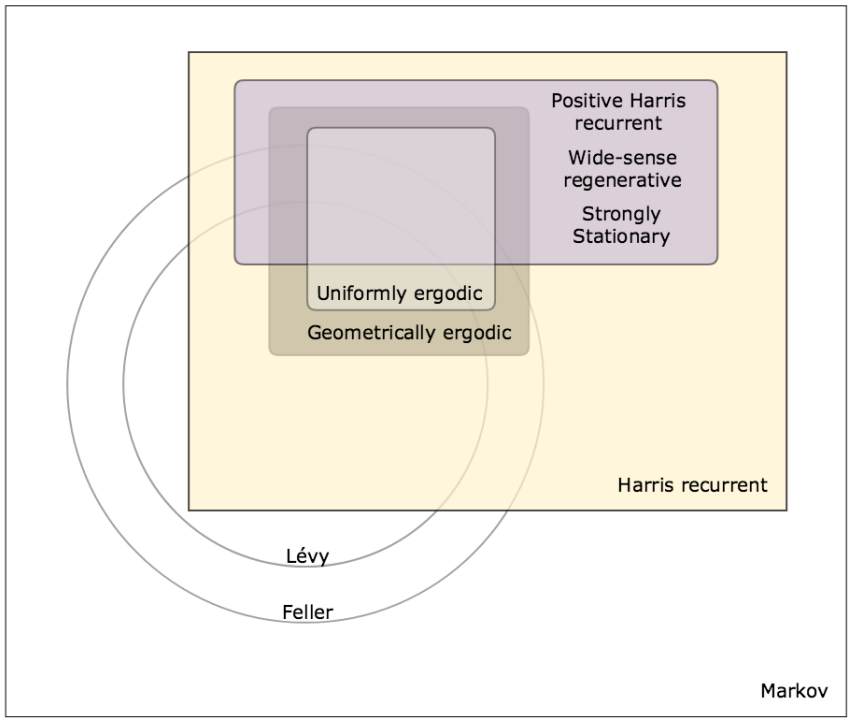


Figure 2.1: Venn diagram of the stability properties in discrete time.

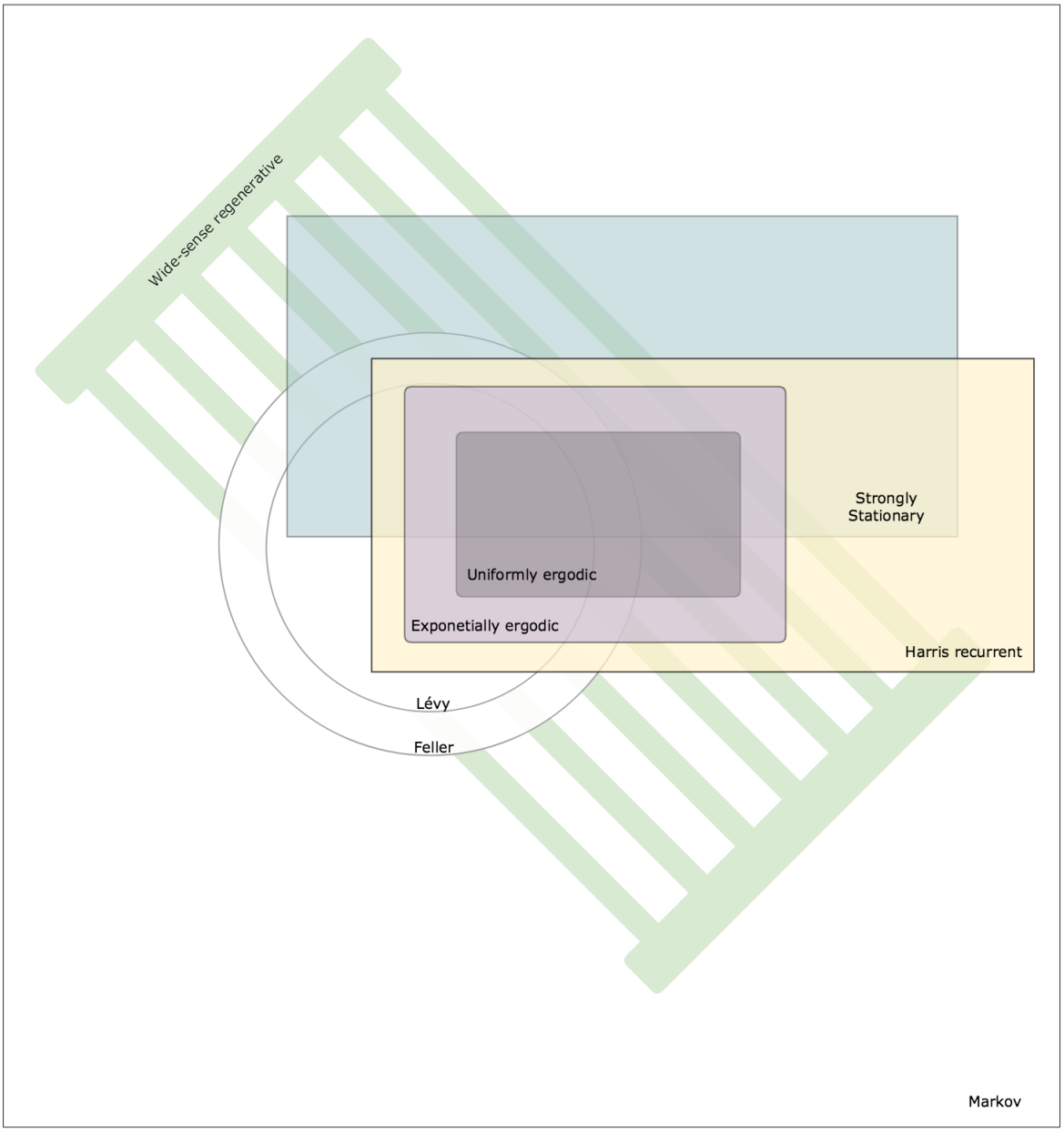


Figure 2.2: Venn diagram of the stability properties in continuous time.

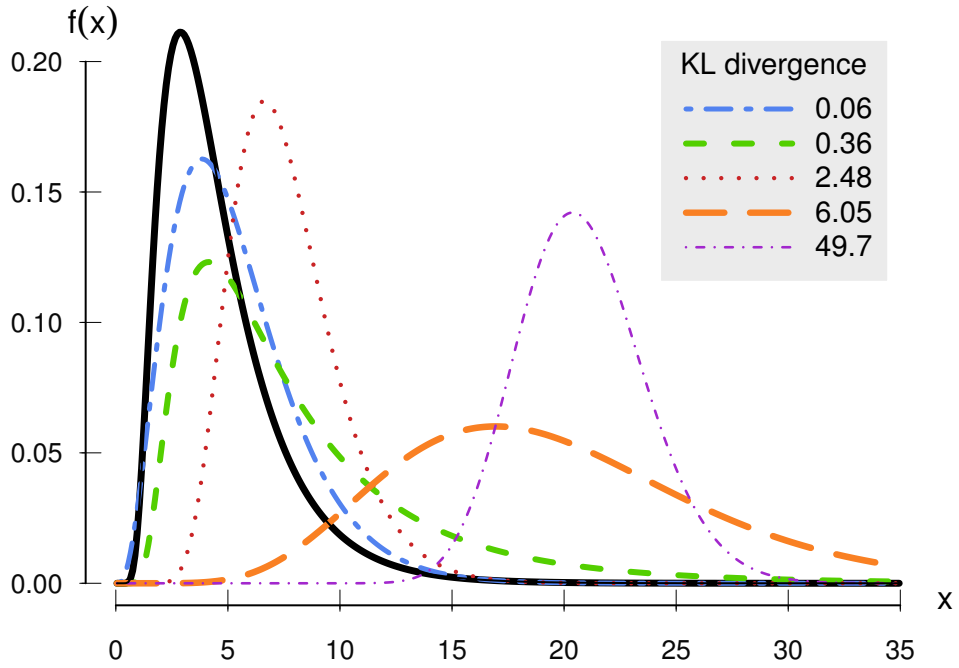


Figure 2.3: In solid appears the $\mathsf{GIG}(0, 3, 4)$ density function. The dotted lines are different approximations of the GIG class to such a distribution along with their KL divergence.

3 SF-Harris process

In this chapter we focus in a subclass of processes that, apart from belonging to our study class of stationary, non-independent increment Markov processes, are Feller, wide-sense regenerative, and Harris recurrent. The first natural question that raises up is if all of these stability conditions are highly restrictive. Should these be the minimal conditions we impose? We will start by motivating why to focus on such a subclass and discussing how flexible it is.

Working with Feller processes in general has not been very attractive in applications. This could be for various reasons. First, their literature is sparse compared to the one of Lévy processes. Second, their construction is in general difficult. The standard construction methods possess a number of technical conditions, and they have to deal with the problem of proving the very existence of the processes. (see e.g. Bass, 1988; Böttcher, 2010; Böttcher and Schnurr, 2011; Hoh, 1998; Jacob, 2005; Stroock, 2003). Third, simulating any Feller process is non-trivial (see e.g., Böttcher, 2010; Böttcher and Schnurr, 2011; Stroock, 2003). Last, to our knowledge, inference of the general class of Feller processes has not been fully resolved.

To be able to model more accurately some scenarios, while keeping at a range of tractable models, certain stability aspects can be assumed for the Feller class. The main motivation for the stability properties we chose to focus on emerged from what we consider to be an indispensable condition. For the processes to be tractable, their inference should be achievable from a single, sufficiently long sample. This follows because it is common in applications that a single realization of the processes is observed. Within the Lévy framework, the increments of such a realization form an independent sample of a probability distribution. As a consequence, inference is simple.

Outside the Lévy case, strong stationarity implies the random processes will not change their probabilistic structure with time; and ergodicity implies statistical sampling can be performed at one instant across a group of identical processes, or sampled over time on a single process with no change in the measured result. Hence, when these properties

are simultaneously assumed, inference can be performed from a single, sufficiently long sample.

In fact, there is a vast literature on strong stationarity and ergodicity for continuous-time Markov processes. Various central limit theorems have been established and prediction problems addressed (see e.g., Oodaira and Yoshihara, 1972; Urbanik, 1967; Volnỳ, 1993). The most applied theories have been developed for diffusion processes, or for processes with some form of regeneration sets, known as Harris recurrent processes (see e.g., Asmussen, 2003; Sigman and Wolff, 1993). Harris recurrent processes own the practical advantage of allowing limiting results, which serve as a tool for statistical analysis, while permitting a vast range of sample behaviors. Therefore, we may see them as a middle ground between the Lévy processes tractability, and the Feller processes generality.

Wide-sense regeneration implies the existence of certain times which allow the trajectories of the processes to be split into identically distributed cycles. As a consequence, it can be exploited in the development of numerical algorithms, specifically in Monte Carlo simulation-based methods. When a Harris recurrent strong Markov process has a recurrent atom it is possible to see it is wide-sense regenerative. However, for general Harris processes, at least without further assumptions, recurrent atoms do not exist. For discrete-time Harris chains, Nummelin (1978) gives a way of constructing a recurrent atom on an extended probability space, provided the transition operator of the chain satisfies a minorisation condition. This construction, called splitting, assures discrete-time Harris chains automatically exhibit a wide-sense regenerative structure.

The question of how to define an exhaustive class of Harris recurrent and wide-sense regenerative processes in continuous time remains open. To build a non-independent increment subclass of processes that are stationary, Feller, wide-sense regenerative, and Harris recurrent we can extend the chains obtained with the splitting technique, and then seek conditions to guarantee the membership to our research subclass is preserved. That is exactly how the SF-Harris processes are obtained.

Lastly, the answer to the question about these being the minimal stability conditions is nontrivial. It is hard to define minimal conditions since, as we saw in Section 2.1, in continuous time there is not a clear order between the properties that can be supposed to simplify the statistic work. The investigation on non-independent increment processes has been extensive in numerous directions, and there is still much to explore. In this chapter, we targeted a portion of these models which is narrow enough to assure us the existence

of simulation and inference methods, while being broad enough to suit different scenarios when applied to real data. In Chapter 6 we will study another subclass of processes satisfying these features. Alternative classes of processes such as diffusions, which must meet some conditions for the existence of stationary, ergodic or Harris recurrent solutions, are extensively used in applications.

3.1 Definition

In the discrete-time case, a Harris recurrent Markov process is known as a Harris chain (see, e.g., Asmussen and Rosiński, 2001). When an aperiodic Harris chain $X = (X_n)_{n\in\mathbb{N}}$ takes values in a separable metric space $(E, \mathcal{E})$, the transition probabilities can be written as a weighted sum between two probability measures, one depending on the starting point, and the other one independent. So for $\epsilon \in [0,1]$,

$$K(x, A) = (1-\epsilon)Q(A) + \epsilon\mu_x(A), \tag{3.1}$$

where $A \in \mathcal{E}$, $K(x, A) = \mathbb{P}(X_{n+1} \in A | X_n = x)$, Q is a probability measure, and μ_x is a probability for each $x \in E$.

Harris chains exhibit a wide-sense regenerative structure and the representation (3.1) has the advantage of explicitly marking it. The process will be dependent on the previous point x with probability ϵ, or it will take a new value, independent of x, with the complimentary probability, thus starting a new cycle.

The structure of (3.1) also clarifies that the chain flexibility relies in two aspects. One is the dependence structure μ_x, and the other is the model distributional properties, build upon Q. A trade-off between the generality of distributions μ_x and Q is usually made since, often, a flexible model can be obtained by focusing solely on the generality of one of them. For instance, by letting Q to be multimodal, it is possible to model with stationary processes trajectories typically associated to non-stationarity (see, e.g., Antoniano-Villalobos and Walker, 2016). Hence, a flexible enough Q allows a simple choice of μ_x. In the opposite direction, complex dependence structures may compensate simple choices of distributional features. Such is the case, for example, of some non-linear autorregresive processes.

Here, we will let Q to be arbitrary, and restrict ourselves to a simple dependence structure. The simplest kind we can think of is to set μ_x as a degenerate probability distribution in the value x, meaning that

$$P(x, A) = (1-\epsilon)Q(A) + \epsilon\delta_x(A), \tag{3.2}$$

where $P(x, A) = \mathbb{P}(Y_{n+1} \in A | Y_n = x)$, being $Y = (Y_n)_{n\in\mathbb{N}}$ a Harris chain, and $\delta_x(A)$ the Dirac measure, which equals one when $x \in A$, and zero otherwise. When $\epsilon = 0$ transitions (3.2) correspond to an independent process, and $\epsilon = 1$ results in a completely dependent process with constant paths.

We can extend a chain with transition functions (3.2) to a continuous-time Markov process $H = (H_t)_{t\geq 0}$ by making the parameter ϵ a function of time, $t \to \epsilon(t)$. Such an extension preserves entirely the chain virtues. Indeed, the resulting non-independent-increment process will be Harris recurrent, and will model in a simple manner the similarity between observations, while providing a flexible approach. Performing this extension, the Chapman-Kolmogorov equation leads to the condition $\epsilon(s+t) = \epsilon(s)\epsilon(t)$. Hence, $\epsilon(t) = e^{-\alpha t}$ for some $\alpha > 0$, and the Markov process transition probabilities $P_t(x, A) = \mathbb{P}(H_t \in A | H_0 = x)$ are given by

$$P_t(x, A) = (1-e^{-\alpha t})Q(A) + e^{-\alpha t}\delta_x(A). \tag{3.3}$$

As we saw in Section 2.1, by contrast to the discrete case, in continuous time, whether all Harris recurrent Markov processes have a wide-sense regenerative structure is an open problem. This has caused the definition of Harris processes to be somewhat combined in the literature. We will say a continuous-time stochastic process is Harris if it is strongly Markovian, Harris recurrent, and wide-sense regenerative. The extension of Harris chains to (3.3) falls in this class, since the wide-sense regenerative structure follows automatically. The process has piecewise constant paths, it stays in its current state an exponential time, before jumping to another state randomly sampled from Q.

Let us consider for a moment what kind of process we would obtain if we extend the wider class of chains (3.1) to continuous time. The transition probabilities

$$K_t(x, A) = (1-e^{-\alpha t})Q(A) + e^{-\alpha t}\mu_x(A) \tag{3.4}$$

are Markovian, since $K_t(x, A) = \int P_t(y, A)\mu_x(dy)$, where P_t denotes the transition proba-

bilities (3.3). It also follows the associated processes are wide-sense regenerative. However, if κ is an invariant measure for (3.4), then for every time t it follows that

$$Q(A)(1-e^{-\alpha t})+e^{-\alpha t}\int \mu_x(A)\kappa(dx)=\kappa(A)$$

Taking $t=0$ we get

$$\int \mu_x(A)\kappa(dx)=\kappa(A),$$

and making $t\to\infty$ we obtain $Q(A)=\kappa(A)$. Hence, only a subclass of (3.4) possesses a unique invariant measure, integrated by the kernels μ_x which keep Q invariant. Notice this is automatically fulfilled by the kernel δ_x. Hence, the SF-Harris process has a unique invariant measure and it is, as a consequence, strongly stationary.

Moreover, the semigroup operator $\hat{\mathsf{T}}$ corresponding to (3.4) is given by

$$\hat{\mathsf{T}}_t f(x)=(1-e^{-\alpha t})\int f(y)Q(dy)+e^{-\alpha t}\int f(y)\mu_x(dy). \tag{3.5}$$

Thus, it meets the Feller conditions only when $\mu_x=\delta_x$, in which case we return to (3.3). Due to this, the term *SF-Harris process* is justified.

Definition 14. *The* ***SF-Harris process*** $\{H_t\}_{t\geq 0}$ *(Anzarut and Mena, 2016) is a stochastic process taking values in* $(E,\mathcal{E})$*, evolving in continuous time, and driven by the transition probability functions* $P_t : E\times\mathcal{E}\to[0,1]$ *given by* (3.3)*, where* $P_t(x,A)=\mathbb{P}(H_t\in A|H_0=x)$*,* $(E,\mathcal{E})$ *is a measurable space,* Q *is a probability, and* $\alpha>0$*. We will assume a random starting point* x *with distribution* Q*.*

We let E be a Polish space, although this could be potentially relaxed. Figures 3.1 and 3.2 illustrate some SF-Harris process paths; we can observe the distribution Q modulates the marginal behavior at each time, and the parameter α sets the rate at which the process jumps. Let us emphasize that $(E,\mathcal{E})$ can be any measurable space. For example, we could let E be the space of positive definite matrices with real values. Setting Q as the Wishart distribution, we would obtain a stationary, non-independent increment, and easy to simulate process moving in this space of matrices.

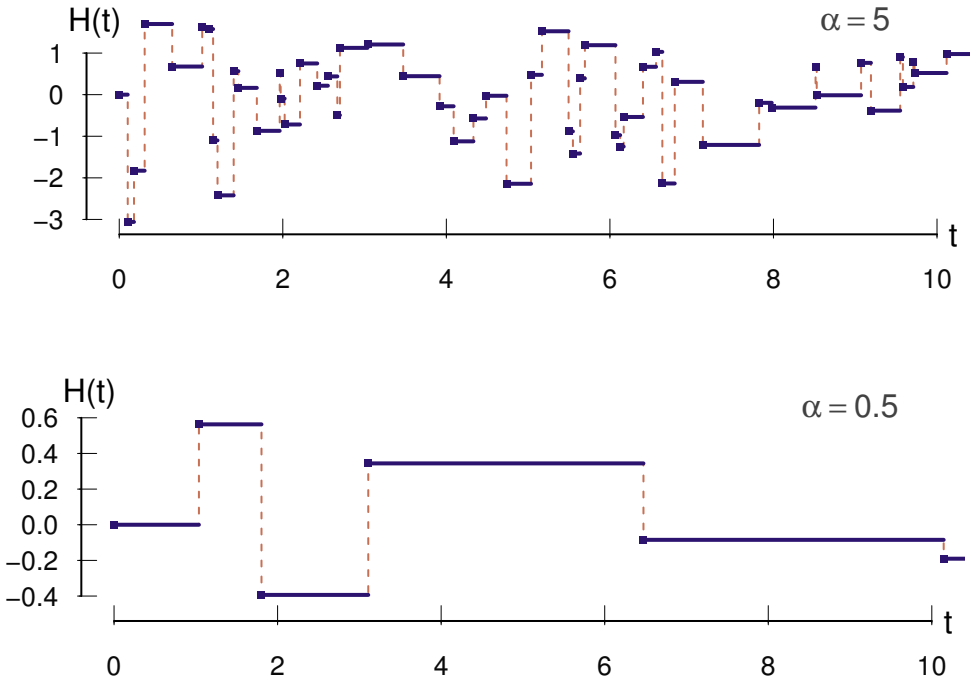


Figure 3.1: Trajectories of the SF-Harris process, where Q is a standard Normal.

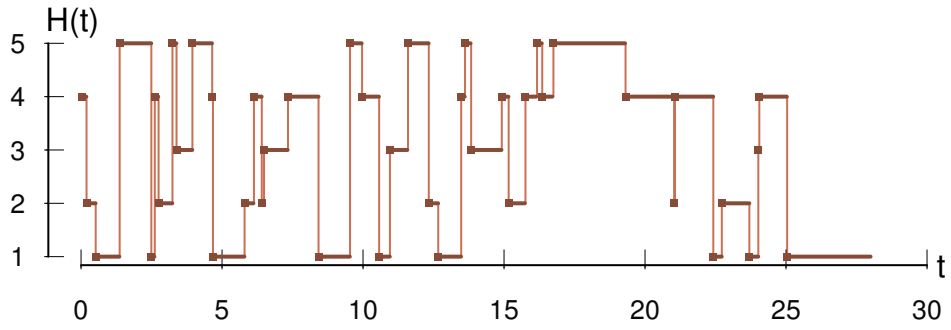


Figure 3.2: Trajectory of the SF-Harris process, where Q is a $\mathsf{U}(1,2,3,4,5)$ and $\alpha = 2$.

3.2 Properties

We stated that the SF-Harris process is Feller, Harris recurrent and it has stationary, non-independent increments. Once it is known that a process is Feller, many useful properties follow, such as the existence of right-continuous with left-hand limits paths modifications, the strong Markov property, and right-continuity of the filtration. Moreover, the resolvent R_λ of the process and the infinitesimal generator A are given by:

$$\begin{aligned}\mathsf{R}_\lambda f(x) &= \int_0^\infty e^{-\lambda t}\left\{(1-e^{-\alpha t})Qf(x)+e^{-\alpha t}f(x)\right\}dt \\ &= Qf(x)\int_0^\infty e^{-\lambda t}-e^{-(\lambda+\alpha)t}dt + f(x)\int_0^\infty e^{-(\lambda+\alpha)t}dt \\ &= Qf(x)\left(\frac{1}{\lambda}+\frac{1}{\lambda+\alpha}\right)+f(x)\frac{1}{\lambda+\alpha};\end{aligned}$$

and

$$\begin{aligned}\mathsf{A}f &= \lim_{t\downarrow 0}\frac{1-e^{-\alpha t}}{t}(Qf-f) \\ &= \alpha(Qf-f),\end{aligned}$$

where $f \in C_0$.

Using the semigroup operator is straightforward to check the SF-Harris process has constant mean and variance which match the ones of the distribution Q. Given this, we have that for any $t, h > 0$

$$\begin{aligned}\mathbb{E}[H_t H_{t+h}] &= \int\int\int yzP_t(x,dy)P_h(y,dz)Q(dx) \\ &= \int\int yzP_h(y,dz)Q(dy) \\ &= \int yQ(dy)\mathsf{T}_h y \\ &= \int yQ(dy)\left\{(1-e^{-\alpha h})Qy+e^{-\alpha h}y\right\} \\ &= (1-e^{-\alpha h})(Qy)^2+e^{-\alpha h}Qy^2 \\ &= (Qy)^2+e^{-\alpha h}\left\{Qy^2-(Qy)^2\right\}.\end{aligned}$$

Consequently, it follows that $\mathsf{Cor}(H_t, H_{t+h}) = e^{-\alpha h}$.

We defined the SF-Harris process through its transition probabilities but there exists a

unique Feller process with right-continuous with left-hand limits paths driven by transition probabilities (3.3). An explicit representation can be found by uniformizing the chain Y with transition probabilities (3.2). Let us expand on this. Assuming $\epsilon \in (0,1)$, consider the process $Y_N = (Y_{N_t})_{t\geq 0}$, where $N = (N_t)_{t\geq 0}$ is a Poisson process of rate $\lambda > 0$ stochastically independent of Y.

The 2-step transition functions of Y, $P^2(x,A) = \mathbb{P}(Y_2 \in A | Y_0 = x)$, are given by

$$\begin{aligned} P^2(x,A) &= \int P(x,dy)P(y,A) \\ &= (1-\epsilon^2)Q(A) + \epsilon^2 \delta_x(A); \end{aligned}$$

and indeed, the k-step transitions, $P^k(x,A) = \mathbb{P}(Y_k \in A | Y_0 = x)$, follow the recursion

$$P^k(x,A) = \int P^{k-1}(x,dy)P(y,A).$$

Hence, it is easy to prove inductively

$$P^k(x,A) = (1-\epsilon^k)Q(A) + \epsilon^k \delta_x(A).$$

Therefore, the transition functions of Y_N are given by

$$\begin{aligned} P_t(x,A) &= \sum_{k=1}^{\infty} P^k(x,A) \frac{(\lambda t)^k e^{-\lambda t}}{k!} \\ &= \left\{ 1 - e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t \epsilon)^k}{k!} \right\} Q(A) + \left\{ e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t \epsilon)^k}{k!} \right\} \delta_x(A) \\ &= \left(1 - e^{-\lambda t(1-\epsilon)}\right) Q(A) + e^{-\lambda t(1-\epsilon)} \delta_x(A). \end{aligned}$$

By making $\lambda = \alpha(1-\epsilon)^{-1}$, we recover (3.3). Consequently, we have the following stochastic representation.

Theorem 3.2.1. *Let $Y = (Y_n)_{n\in\mathbb{N}}$ be a Markov chain with transition functions given by (3.2), where $\epsilon \in [0,1)$, $Y_0 \sim Q$, and Q is a probability distribution. If $N = (N_t)_{t\geq 0}$ is a Poisson process with rate $\lambda = \alpha(1-\epsilon)^{-1}$ independent of Y, then the SF-Harris process, with marginal distribution Q and jump parameter α, has the stochastic representation $H_t = Y_{N_t}$ for all $t \geq 0$.*

Theorem 3.2.1 implies the SF-Harris process is a pseudo-Poisson Process, defined by Feller (1966) as a continuous-time process that can be obtained from discrete time Markov

chains by subordination with a Poisson process. Pseudo-Poisson Process are a special kind of Markov jump process. In fact, every Feller process can be approximated in a natural way by pseudo-Poisson processes (see Kallenberg, 2002, Chapter 10).

The case $\epsilon = 1$ in Theorem 3.2.1 corresponds to the limit $\alpha \to \infty$, where no jumps occur so the process has constant paths. The case $\epsilon = 0$ corresponds to independent identically distributed random variables $(Y_n)_{n\in\mathbb{N}}$, it is included in the theorem taking into account that $P^0(x, A) = \delta_x(A)$ for any $\epsilon \geq 0$.

Notice that the parameter $\lambda = \alpha(1-\epsilon)^{-1}$ ranges in the positive real line. Fixing a value for α, when ϵ grows λ gets smaller. In other words, when the dependence in the paths of the Harris chain Y grows, then the Poisson process N introduces less dependence, in order to get exactly the same dependence rate α in the uniformized process $H = Y_N$. For the construction or simulation of a SF-Harris process using Theorem 3.2.1 any value ϵ can be used. In particular, we stick to the simpler case $\epsilon = 0$ which results in $\lambda = \alpha$. The following immediate corollary is useful for some calculations.

Corollary 1. *Let Y be as in Theorem 3.2.1, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of random variables independent of Y. Suppose the increments $(S_n)_{n\in\mathbb{N}}$, defined as $S_n = T_n - T_{n-1}$, are exponential, independent, and identically distributed with mean $(1-\epsilon)\alpha^{-1}$ $\left(S_n \sim \mathsf{Exp}\left\{\alpha(1-\epsilon)^{-1}\right\}\right)$. Then the SF-Harris process, with marginal distribution Q and jump parameter α, has the stochastic representation*

$$H_t = x\mathbb{1}_{t<T_1} + \sum_{n=1}^{\infty} Y_n \mathbb{1}_{t\in[T_n, T_{n+1})}, \quad \textit{for all} \quad t \geq 0. \tag{3.6}$$

As of yet, we have three representations that give an intuition about the SF-Harris process, and facilitate the study of distinct features. For instance, using the transition probability functions (3.3), it can be easily checked the SF-Harris process is time-reversible with invariant measure Q, and so a strongly stationary process. On the other hand, with the representation in Corollary 1 can be shown that for any state B such that $Q(B) > 0$, we have that $\int_0^\infty \delta_{H_t}(B)dt$ is infinite almost surely. This means that when $Q(B) > 0$, the state B is visited infinitely often with probability one. Since Q is a finite measure, the process is positive Harris recurrent.

As a consequence of the Harris recurrence, the process will eventually converge to Q.

Moreover, the total variation distance is given by

$$\sup_{A\in\mathcal{E}}|P_t(x,A)-Q(A)| = e^{-\alpha t}\sup_{A\in\mathcal{E}}|\delta_x(A)-Q(A)| = e^{-\alpha t},$$

which translates to the process being uniformly ergodic. Lastly, the process regenerative property is evidenced in a natural way from Corollary 1 taking $\epsilon = 0$. The independent and identically distributed $(S_n)_{n\in\mathbb{N}}$ are inter-regeneration times were the process starts afresh. Hence, they form a renewal process, and split the stochastic process into a sequence of identically distributed cycles.

3.3 Integrated SF-Harris process

The features shown in the previous section about H allow us to derive in a simple manner similar properties for the integrated SF-Harris process denoted with $H^* = (H_t^*)_{t\geq 0}$, where $H_t^* = \int_0^t H_s ds$. For example, the stochastic representation in Corollary 1 provides the means to derive a similar representation for H^*.

Theorem 3.3.1. *Let Y, $(T_n)_{n\in\mathbb{N}}$, and $(S_n)_{n\in\mathbb{N}}$ be as in Corollary 1 . If $N_t = \max\{n\in\mathbb{N} : T_n \leq t\}$, then the integrated process $H^* = (H_t^*)_{t\geq 0}$ has the stochastic representation*

$$H_t^* = \sum_{n=0}^{N_t-1}(Y_n - Y_{N_t})S_{n+1} + Y_{N_t}t,$$

where we let Y_0 and T_0 be constant random variables equal to x and 0 respectively.

Proof. Due to Corollary 1

$$\begin{aligned} H_t^* &= \int_0^t \{x\mathbb{1}_{s<T_1} + \sum_{n=1}^{\infty} Y_n \mathbb{1}_{s\in[T_n,T_{n+1})}\}\, ds \\ &= x\int_0^{t\wedge T_1} ds + \sum_{n=1}^{\infty} Y_n \int_{t\wedge T_n}^{t\wedge T_{n+1}} ds. \end{aligned}$$

Since $Y_0 = x$ and $T_0 = 0$, it follows

$$H_t^* = \sum_{n=0}^{\infty} Y_n\{(t\wedge T_{n+1}) - (t\wedge T_n)\}.$$

Equivalently,

$$\begin{aligned} H_t^* &= \sum_{n=0}^{N_t-1} Y_n(T_{n+1}-T_n) + Y_{N_t}(t-T_{N_t}) \\ &= \sum_{n=0}^{N_t-1} Y_n S_{n+1} + Y_{N_t}\left(t - \sum_{n=0}^{N_t-1} S_{n+1}\right) \\ &= \sum_{n=0}^{N_t-1} (Y_n - Y_{N_t}) S_{n+1} + Y_{N_t} t. \end{aligned}$$

□

Note that, once again, N is a Poisson process of rate α. Let us proceed by highlighting some second order properties about the integrated process.

Proposition 1. *Denote with $\xi = \mathbb{E}[H_t]$ and $\sigma^2 = \mathsf{Var}(H_t)$. Then, for any $t, h > 0$, it follows that*

(i) $\mathbb{E}[H_t^*] = \xi t$.

(ii) $\mathsf{Var}(H_t^*) = \dfrac{2\sigma^2}{\alpha^2}\left(e^{-\alpha t} + \alpha t - 1\right)$.

(iii) $\mathsf{Cov}(H_t^*, H_{t+h}^*) = \dfrac{\sigma^2}{\alpha^2}\left(e^{-\alpha(t+h)} + e^{-\alpha t} - e^{-\alpha h} + 2\alpha t - 1\right)$.

Proof. Property (i) is direct. Property (ii) is a consequence of

$$\begin{aligned} \mathsf{Var}(H_t^*) &= \int_0^t \int_0^t \mathsf{Cov}(H_u, H_v) du dv \\ &= \int_0^t \int_0^t \sigma^2 e^{-\alpha|u-v|} du dv \\ &= 2\sigma^2 \int_0^t \int_0^v e^{-\alpha u} du dv; \end{aligned}$$

and property (iii) arises from

$$\begin{aligned} \mathsf{Cov}(H_t^*, H_{t+h}^*) &= \int_0^t \int_0^{t+h} \mathsf{Cov}(H_u, H_v) du dv \\ &= \mathsf{Var}(H_t^*) + \int_0^t \int_t^{t+h} \mathsf{Cov}(H_u, H_v) du dv \\ &= \frac{2\sigma^2}{\alpha^2}\left(e^{-\alpha t} + \alpha t - 1\right) + \int_0^t \int_t^{t+h} \sigma^2 e^{-\alpha(u-v)} du dv. \end{aligned}$$

□

Two interesting observations that follow from Proposition 1 are that $\mathsf{Cor}(H_t^*, H_{t+h}^*)$ is always positive, and it does not depend on σ^2. Thus, it is the same for any marginal distribution Q. Further, the ergodicity of H implies that

$$\lim_{t\to\infty} \frac{H^*(t)}{t} = \xi$$

almost surely.

3.4 A semi-Markovian extension

The SF-Harris process is a short-memory process because its correlation decays exponentially at a rate determined by one of the parameters. However, long-memory structures are often observed in real data, examples can be found in fields as econometrics, hydrology or telecommunications (see, e.g. Samorodnitsky, 2007, for a review). To capture these dependence structures we could consider to model the time between jumps with a heavy-tailed distribution, rather than an exponential one. That is, taking Yas in Theorem 3.2.1, we can define a more general process $\xi = (\xi_t)_{t\geq 0}$, given by

$$\xi_t = x\mathbb{1}_{t<R_1} + \sum_{n=1}^{\infty} Y_n \mathbb{1}_{t\in[R_n, R_{n+1})}, \tag{3.7}$$

where $(R_n)_{n\in\mathbb{N}}$ is a sequence of random variables independent of Y whose increments $(V_n)_{n\in\mathbb{N}}$, defined as $V_n = R_n - R_{n-1}$, are positive with probability one, independent, identically distributed, and with an arbitrary cumulative distribution function G.

When the increments $(V_n)_{n\in\mathbb{N}}$ are exponential, we recover the SF-Harris process, in which case the lack of memory property of the exponential distribution causes the process to be Markovian. Otherwise, the process will be non-Markovian. However,

$$\begin{aligned} P(Y_{n+1} \in A, R_{n+1} \leq t | Y_m = y_m, R_m = r_m, m \leq n) \\ = P(Y_{n+1} \in A, R_{n+1} - R_n \leq t - r_n | Y_n = y_n). \end{aligned}$$

This implies that the process ξ is a particular case of a semi-Markov processes.

Semi-Markov processes were introduced by Lévy (1954) and Smith (1955) almost simultaneously. For these processes, the Markov property is not fulfilled in general. However, the processes can be embedded in a Markov process on a higher dimensional state space. The class of semi-Markov processes has been thoroughly developed and applied in many fields. The majority of authors study the case of finite or countable state space. Given that the case of interest in this study is possibly in uncountable state space, we refer to the work of Vega-Amaya (2012).

The characteristic feature of the process (3.7) and, in fact, of any semi-Markov process, is a set of intervals of constancy in their trajectory. Since this structure is similar to the SF-Harris process we can deduct a number of properties. For example, the process (3.7) has non-independent increments, its mean and variance match the ones of the distribution Q, while its autocorrelation function is given by $r(t) = 1 - G(t)$. We also have that for any $A \in \mathcal{E}$,

$$\mathbb{P}_x(\xi_t \in A) = G(t)Q(A) + \{1 - G(t)\}\delta_x(A).$$

Therefore, the process is wide-sense regenerative, it stays in its current state for a time depending on distribution G, before jumping to another state sampled from Q.

It is immediate to see as well that $\mathbb{P}_x(\xi_t \in A) \to Q(A)$ when $t \to \infty$. This implies Q is the limit distribution. Even more,

$$\sup_{A\in\mathcal{E}} |\mathbb{P}_x(\xi_t \in A) - Q(A)| = \{1 - G(t)\} \sup_{A\in\mathcal{E}} |\delta_x(A) - Q(A)| \leq 1 - G(t).$$

Hence,

$$\lim_{t\to\infty} \left\{ \sup_{A\in\mathcal{E}} |\mathbb{P}_x(\xi_t \in A) - Q(A)| \right\} = 0$$

and the process is always ergodic. If G is a light-tailed distribution, meaning its tail is exponentially bounded, then the process is uniformly ergodic.

When $Q(B) > 0$ it is easy to verify that

$$\mathbb{P}_x \left\{ \int_0^\infty \delta_{\xi_t}(B)dt = \infty \right\} \geq \mathbb{P}_x \left\{ \sum_{n=1}^{\infty} V_n \delta_{Y_n}(B) = \infty \right\} = 1.$$

Therefore, the process (3.7) is positive Harris recurrent. Additionally, we have the following results, whose proofs are analogous to the corresponding SF-Harris process case.

Theorem 3.4.1. *Let $\xi = (\xi_t)_{t\geq 0}$ be as in (3.7). If $K_t = \max\{n \in \mathbb{N} : R_n \leq t\}$, and $K_0 = 0$, then $K = (K_t)_{t\geq 0}$ is stochastically independent of Y and $\xi_t = Y_{K_t}$ for all $t \geq 0$.*

Notice that K is a renewal counting process representing the number of arrivals of the process (3.7) in the interval $(0, t]$.

Theorem 3.4.2. *Let $\xi = (\xi_t)_{t\geq 0}$ be as in (3.7); and let K be as in Theorem 3.4.1. The integrated process, $\xi^* = (\xi_t^*)_{t\geq 0}$, where $\xi_t^* = \int_0^t \xi_s ds$, is given by*

$$\xi_t^* = \sum_{n=0}^{K_t-1} (Y_n - Y_{K_t})V_{n+1} + Y_{K_t} t,$$

defining $Y_0 = x$.

Proposition 2. *Let $\xi = (\xi_t)_{t\geq 0}$ be as in (3.7). Denote with $\hat{\xi} = \mathbb{E}[\xi_t]$ and $\hat{\sigma}^2 = \mathsf{Var}(\xi_t)$. Then, the integrated process defined in Theorem 3.4.2 has the following second order properties.*

For any $t, h > 0$,

(i) $\mathbb{E}[\xi_t^*] = \hat{\xi} t$.

(ii) $\mathsf{Var}(\xi_t^*) = \hat{\sigma}^2 \int_0^t \int_0^t G(u - v) du dv$.

(iii) $\mathsf{Cov}(\xi_t^*, \xi_{t+h}^*) = \mathsf{Var}(\xi_t^*) + \hat{\sigma}^2 \int_0^t \int_t^{t+h} G(u - v) du dv$.

3.5 Inference

3.5.1 Estimation methods

In this section we develop estimation methods for the SF-Harris process parameters. We deal with the one-dimensional case, although all methods can be extended in a natural way to higher dimensions. We assume that $x_1, ..., x_n$ is a discrete realization of the process at times $\hat{t}_1 < \cdots < \hat{t}_n$, and develop four different methods.

The no difference no jump method (NDNJ)

In this method we assume that when two observations are different it occurs because the process jumped exactly at that moment, and when they are the same, it occurs because the process did not jump. Following this reasoning, we equate the expected value of α with its sample mean, and then estimate the parameters of Q with the observations in the sample that are distinct. This natural method is likely to show problems when Q has a high probability of repeated values, or when few observations are available.

Maximum likelihood estimation (MLE)

The MLE consists in the maximization of the likelihood or, equivalently, the log-likelihood function, which in this case is given by

$$\log L(\alpha, \beta) = \log Q(x_0|\beta) + \sum_{i=1}^{n} \log\left\{(1 - e^{-\alpha t_i})Q(x_i|\beta) + e^{-\alpha t_i}\delta_{x_i}(x_{i-1})\right\},$$

where β denotes the set of parameters of Q, $t_i = \hat{t}_i - \hat{t}_{i-1}$ for $i = 1, ..., n$, and $t_0 = 0$. We maximize it numerically.

Expectation-maximization algorithm (EM)

The EM algorithm is an iterative procedure to compute maximum likelihood estimates of parameters in statistical models depending on unobserved latent variables, or in problems which can be posed in a similar form. The EM iteration alternates between performing an expectation (E) step, which creates a function for the conditional expectation of the augmented likelihood, evaluated using the current estimate for the parameters; and a maximization (M) step, which computes parameters maximizing the expectation found on the E step. These parameter estimates are then used to determine the distribution of the latent variables in the next E step. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration.

In this case, we introduce the latent variables $z_0, ..., z_n$, where z_i equals one if x_i comes

from Q, or zero otherwise. Including these variables the augmented likelihood results in

$$\begin{aligned} L^a(\alpha,\beta) &= \prod_{i=1}^{n} f(x_i, z_i|\alpha,\beta) \\ &= Q(x_0|\beta) \prod_{i=1}^{n} \left\{(1-e^{-\alpha t_i})Q(x_i|\beta)\right\}^{z_i} \left\{e^{-\alpha t_i}\delta_{x_i}(x_{i-1})\right\}^{1-z_i}. \end{aligned}$$

Therefore, as $z_0 = 1$,

$$\log L^a(\alpha,\beta) = \sum_{i=1}^{n} \left\{ z_i \log(1-e^{-\alpha t_i}) + (1-z_i)(-\alpha t_i)\right\} + \sum_{i=0}^{n} z_i \log Q(x_i|\beta).$$

Now $f(z_i|x_i,\alpha,\beta) \propto f(x_i,z_i|\alpha,\beta)$ so it follows that $z_i|(x_i,\alpha,\beta) \sim \mathsf{Ber}(p_i)$ with

$$p_i = \frac{(1-e^{-\alpha t_i})Q(x_i|\beta)}{(1-e^{-\alpha t_i})Q(x_i|\beta) + e^{-\alpha t_i}\delta_{x_i}(x_{i-1})}.$$

Properties of the Bernoulli distribution imply

$$\mathsf{E}_{z_i|x_i,\alpha^{(k-1)},\beta^{(k-1)}}[z] = \frac{\left(1-e^{-\alpha^{(k-1)}t_i}\right) Q\left(x_i|\beta^{(k-1)}\right)}{\left(1-e^{-\alpha^{(k-1)}t_i}\right) Q\left(x_i|\beta^{(k-1)}\right) + e^{-\alpha^{(k-1)}t_i}\delta_{x_i}(x_{i-1})}.$$

Denoting by $p_i^{(k-1)} = \mathsf{E}_{z_i|x_i,\alpha^{(k-1)},\beta^{(k-1)}}[z]$, at each step of the EM we maximize the function $G\left\{(\alpha,\beta)\,|\,\left(\alpha^{(k-1)},\beta^{(k-1)}\right)\right\}$, given by

$$\sum_{i=1}^{n} \left\{ p_i^{(k-1)} \log(1-e^{-\alpha t_i}) + (1-p_i^{(k-1)})(-\alpha t_i)\right\} + \sum_{i=0}^{n} p_i^{(k-1)} \log Q(x_i|\beta),$$

either numerically or analytically; depending on the difficulty of the distributional form of Q.

Gibbs sampler

The Gibbs sampler is a Markov chain Monte Carlo algorithm for obtaining a sequence of observations from a specified probability distribution when direct sampling is difficult. The idea is to construct a Markov chain that has the desired distribution as its equilibrium distribution. Since the algorithm generates a sequence of observations, when comparing it to the other methods we use the modes as estimators.

We assign prior distributions $\pi(\alpha), \pi(\beta)$ for α and β. Since α is a positive real number we choose $\pi(\alpha) = \mathsf{Exp}(c)$ for some $c > 0$. On the other hand, the choice of $\pi(\beta)$ will depend on the form of Q. Consequently, the joint posterior distribution is

$$\pi(\alpha, \beta | x_1, ..., x_n) \propto \prod_{i=1}^{n} \Big\{ (1 - e^{-\alpha t_i}) Q(x_i|\beta) + e^{-\alpha t_i} \delta_{x_i}(x_{i-1}) \Big\} e^{-\alpha c} \pi(\beta).$$

The joint posterior distribution has 2^n terms. Rather than dealing with it, we work again with the variables $z_1, ..., z_n$. Under this scheme, denoting with $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{z} = (z_1, ..., z_n)$, we have that $f(\mathbf{x}, \mathbf{z}, \alpha, \beta) \propto L^a(\alpha, \beta)\pi(\beta)$. As a result

$$\pi(\beta | \alpha, \mathbf{x}, \mathbf{z}) \propto \prod_{i=1}^{n} \Big\{ (1 - e^{-\alpha t_i}) Q(x_i|\beta) \Big\}^{z_i} \pi(\beta),$$

so the choice of $\pi(\beta)$ can cause the full conditional distribution of β to have a tractable form. To simulate from the full conditional distribution of α we consider two options:

(a) (Gibbs-a) We apply Adaptive Rejection Metropolis Sampling (ARMS) (Gilks et al., 1995) to simulate at each iteration from

$$\log \pi(\alpha | \beta, \mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \Big\{ z_i \log(1 - e^{-\alpha t_i}) + (1 - z_i)(-\alpha t_i) \Big\} - c\alpha.$$

(b) (Gibbs-b) We use an additional set of artificial random variables. Let m be the number of $z_i = 1$, that is $m = \sum_{i=0}^{n} z_i$, then $j_0, ..., j_m$ will be the ordered times where the variables z_i equal one. Since $j_1 - j_0, j_2 - j_1, ..., j_m - j_{m-1}$ are independent, identically distributed variables with distribution $\mathsf{Exp}(\alpha)$, denoting by $\mathbf{j} = (j_0, ..., j_m)$ we have that

$$\begin{aligned} \pi(\alpha | \mathbf{j}) &= \prod_{i=1}^{m} \alpha \exp\{-\alpha(j_i - j_{i-1})\} \exp\{-\alpha c\} \\ &= \alpha^m \exp\{-\alpha \sum_{i=1}^{m} j_i - j_{i-1}\} \exp\{-\alpha c\} \\ &= \alpha^m \exp\{-\alpha j_m\} \exp\{-\alpha c\} \\ &= \alpha^m \exp\{-\alpha(j_m + c)\}. \end{aligned}$$

Consequently, $\alpha | \mathbf{j} \sim \mathsf{Gamma}(m + 1, j_m + c)$.

3.5.2 Simulation study

We run a simulation study to account the performance of the four estimation methods. As we will see, the main characteristics which affect the performance are the number of available observations, and the probability of repeated values in Q. Consequently, the best method to apply will depend on the context. To exemplify this, we vary the number of observations on the samples, and test the methods with two marginal distributions.

The first distribution is a discrete uniform, $\mathsf{U}(1,2,3,4,5)$. This distribution serves as a good example for two reasons. First, it makes easy to follow each estimation method, since the only parameter to estimate is α. Second, it points to the fact that with a high probability of repeated values the methods performance differs from the one obtained when Q is continuous. The second distribution is a GIG. This specific continuous distribution is chosen since it will later be used in the SV model.

Roughly, the testing procedure comprises two steps:

I. We choose 100 parameter values randomly. This is done by simulating from a continuous uniform distribution at reasonable intervals for the parameters.

II. For the sample sizes $k = 20$, 100, 500, and 1000:

   i. We simulate a process sample of length k for each parameter value.

   ii. We estimate each parameter value with the four methods.

   iii. We calculate an estimation error for each method using the accuracy for the 100 parameters estimation.

A testing procedure of this kind has a number of advantages. First, it proves the estimation is working for a wide range of parameter sets and for different combinations of them. Second, it assures that any user can run the method without problems since the starting conditions are computed at random. Third, it provides a highly accurate estimation error because there is no subjectivity involved neither in which set of parameters is selected to display, nor in which starting point to use for each run.

Q is a discrete uniform on {1,2,3,4,5}

Let us start with the discrete uniform case. Figure 3.2 illustrates a 28-day trajectory. It is natural to expect that the NDNJ method and the Gibbs-b, which is its Bayesian equivalent, underestimate the value of α. This is because the process is likely to jump and fall back to the previous state. However, in the NDNJ or Gibbs-b methods, if two consequent observations are equal, is assumed that it occurs because the process did not jump. Following the testing procedure confirms this forecast. In step I, we draw the parameters $\alpha_1, ..., \alpha_{100}$ from a uniform distribution over $(0, 30)$. In step II, we compute the error as $E_\alpha = \frac{1}{100} \sum_{i=1}^{100} \left| \frac{\alpha_i - \hat{\alpha}_i}{30} \right|$. The results are shown in Table II. Gibbs-a outperforms all methods.

Table II: Estimation errors for the SF-Harris process parameters in the case of Q being $\mathsf{U}(1,2,3,4,5)$

Sample size	E_α^{NDNJ}	E_α^{MLE}	E_α^{EM}	$E_\alpha^{Gibbs-a}$	$E_\alpha^{Gibbs-b}$
20	0.45	0.94	0.75	0.32	0.44
100	0.40	0.12	0.10	0.15	0.39
500	0.28	0.03	0.04	0.04	0.26
1000	0.22	0.03	0.03	0.02	0.18

Q is a generalized inverse Gaussian

We will use the GIG alternative parameterization developed in Section 2.2, given by equation (2.4). To illustrate the process dynamics we simulate a 40-day trajectory, appearing in Figure 3.3. The estimation with the four methods may be developed further, using the shape and properties of the GIG family.

NDNJ method

The algorithm we follow is:

1. We locate $J = \{j : x_j \neq x_{j-1}\}$,
2. If $J = \emptyset$ then $\hat{\alpha} = 0$, otherwise $\hat{\alpha} = \left(\frac{1}{|J|} \sum_{j \in J} t_j \right)^{-1}$,

3. To obtain $(\hat{\lambda}, \hat{\kappa}, \hat{\eta})$ we maximize numerically $L(\lambda, \kappa, \eta) = \sum_{j \in J} \log \mathsf{GIG}(x_j | \lambda, \kappa, \eta)$ using the BFGS quasi-Newton algorithm (named after Broyden, Fletcher, Goldfarb, and Shanno). Such an algorithm has proven to have good performance for non-smooth optimizations (see, e.g., Nocedal and Wright, 2006).

Maximum likelihood estimation

The corresponding log-likelihood is four-dimensional. Because of the convoluted form of its gradient we apply the Nelder-Mead method (Nelder and Mead, 1965) to maximize it.

Expectation-maximization algorithm

It remains to expand the expression $\sum_{i=0}^{n} p_i^{(k-1)} \log Q(x_i|\beta)$, which in this case is proportional to

$$\sum_{i=0}^{n} p_i^{(k-1)} \left[-\log\{K_\lambda(\kappa)\} - \lambda \log(\eta) + (\lambda - 1)\log(x_i) - \frac{\kappa}{2}\left(\frac{\eta}{x_i} + \frac{x_i}{\eta}\right)\right]$$

or, alternatively,

$$-m^{(k-1)} \log\{K_\lambda(\kappa)\} - m^{(k-1)} \lambda \log(\eta) + (\lambda - 1) S_1^{(k-1)} - \frac{\kappa}{2}\left(\eta S_2^{(k-1)} + \frac{1}{\eta} S_3^{(k-1)}\right),$$

where $m^{(k-1)} = \sum_{i=1}^{n} p_i^{(k-1)}$, $S_1^{(k-1)} = \sum_{i=1}^{n} p_i^{(k-1)} \log(x_i)$, $S_2^{(k-1)} = \sum_{i=1}^{n} p_i^{(k-1)}/x_i$, and $S_3^{(k-1)} = \sum_{i=1}^{n} p_i^{(k-1)} x_i$.

We maximize numerically at each iteration, using BFGS algorithm.

Gibbs sampler

We assume $\pi(\lambda,\kappa,\eta)=\pi(\lambda)\pi(\kappa)\pi(\eta)$, with $\lambda\sim\mathsf{N}(\mu_\lambda,\sigma_\lambda^2)$, $\kappa\sim\mathsf{Ga}(a_\kappa,b_\kappa)$, and $\eta\sim\mathsf{Ga}(a_\eta,b_\eta)$. The final log-joint density function is

$$\log\pi(\lambda,\kappa,\eta|\alpha,\mathbf{x},\mathbf{z})=\sum_{i=1}^{n}z_i\log Q(x_i|\beta)+\log\pi(\lambda)+\log\pi(\kappa)+\log\pi(\eta).$$

With an analogous calculation to the one done in the EM method we obtain

$$\sum_{i=0}^{n}z_i\log Q(x_i|\beta)=-m\log\{K_\lambda(\kappa)\}-m\lambda\log(\eta)+(\lambda-1)S_1-\frac{\kappa}{2}\left(\eta S_2+\frac{1}{\eta}S_3\right),$$

where $m=\sum_{i=1}^n z_i$, $S_1=\sum_{i=1}^n z_i\log(x_i)$, $S_2=\sum_{i=1}^n z_i/x_i$, and $S_3=\sum_{i=1}^n z_i x_i$. Therefore, the full conditional log-distributions are:

$$\begin{aligned}
\log\pi(\lambda|\alpha,\kappa,\eta,\mathbf{x},\mathbf{z})&=-n\log\{K_\lambda(\kappa)\}-\lambda n\log(\eta)+(\lambda-1)S_1-\frac{1}{2\sigma_\lambda^2}(\lambda-\mu_\lambda)^2,\\
\log\pi(\kappa|\alpha,\lambda,\eta,\mathbf{x},\mathbf{z})&=-n\log\{K_\lambda(\kappa)\}-\frac{\kappa}{2}\left(\eta S_2+\frac{1}{\eta}S_3\right)+(a_\kappa-1)\log(\kappa)-b_\kappa\kappa,\\
\log\pi(\eta|\alpha,\lambda,\kappa,\mathbf{x},\mathbf{z})&=-\lambda n\log(\eta)-\frac{\kappa}{2}\left(\eta S_2+\frac{1}{\eta}S_3\right)+(a_\eta-1)\log(\eta)-b_\eta\eta.
\end{aligned}$$

They have no known form so we use the ARMS method to simulate them.

We program these methods, and follow the testing procedure again. In step I, we simulate the values $\alpha_1,...,\alpha_{100}$, $\lambda_1,...,\lambda_{100}$, $\kappa_1,...,\kappa_{100}$, and $\eta_1,...,\eta_{100}$ from a uniform distribution over the intervals $(0,30)$, $(-5,5)$, $(0,50)$, and $(0,4)$ respectively. In step II, the errors are calculated as:

$$E_\alpha=\frac{1}{100}\sum_{i=1}^{100}\left|\frac{\alpha_i-\hat{\alpha}_i}{30}\right|\text{ and }E_Q=\frac{1}{100}\sum_{i=1}^{100}\mathsf{KL}\left\{\mathsf{GIG}(\lambda_i,\kappa_i,\eta_i),\mathsf{GIG}(\hat{\lambda}_i,\hat{\kappa}_i,\hat{\eta}_i)\right\},$$

The results are presented in Table III. On this occasion, the NDNJ method is the one that provides the best results. The MLE, EM, and Gibbs-a methods are poor for the estimation of α, while the Gibbs-b performs well for large samples.

Table III: Estimation errors for the SF-Harris process parameters in the case of Q being $\mathsf{GIG}(\lambda, \kappa, \eta)$

Sample size	E_α^{NDNJ}	E_α^{MLE}	E_α^{EM}	$E_\alpha^{Gibbs-a}$	$E_\alpha^{Gibbs-b}$
20	0.51	1.76	1.49	0.33	0.51
10	0.44	1.59	1.29	0.15	0.44
500	0.28	1.56	1.29	1.09	0.25
1000	0.19	1.51	1.29	1.86	0.17
	E_Q^{NDNJ}	E_Q^{MLE}	E_Q^{EM}	$E_Q^{Gibbs-a}$	$E_Q^{Gibbs-b}$
20	0.09	0.09	0.1	0.77	0.84
100	0.02	0.02	0.03	0.31	0.31
500	0.01	0.04	0.07	0.16	0.18
1000	0.01	0.05	0.09	0.09	0.12

3.5.3 Prediction procedure

A procedure for obtaining a set of m trajectories in new times $t_{n+1} < \cdots < t_{n+k}$ is available by, first, simulating m parameter values from the posterior distributions, and, second, for each one of the m values, simulating a realization of H starting on x_n at time t_n and parameterized by such a value.

Given a SF-Harris process trajectory in the case where Q is $\mathsf{U}(1,2,3,4,5)$, we draw a large number of realizations in future times employing the Gibbs-a method, obtaining a set from which we can make predictions. We could calculate highest posterior density intervals (HPD intervals), but that would not be truly meaningful since, after the process jumps, it will randomly fall in some state $1, 2, 3, 4, 5$. Hence, the most interesting aspect is when the next jump will occur, or the probability of the first jump being before certain time. Indeed, the mean value of the first jump time in the Figure 3.2 trajectory is 28.60, and the probability of the first jump being before one day is 0.79. In the GIG density case, it does make sense to apply the prediction procedure to find HPD intervals. As an example, we took away the last ten days from the trajectory in Figure 3.3 and performed this; simulating 1000 realizations using the Gibbs-b method and calculating with them HPD intervals.

3.6 Concluding remarks

SF-Harris processes arise from an extension of Harris recurrent Markov chains to continuous time being stationary, Feller, and Harris recurrent. Such processes exhibit a wide-sense regenerative structure, they have a unique invariant distribution which can be tailored to fit different scenarios, and they are uniformly ergodic. The structure of their transition probabilities makes it easy to develop several representations and accurate inference methods. In addition, the transitions offer a clear trade-off between independence and full dependence.

SF-Harris processes may be applied in a wide spectrum of contexts, we explored two of them, stochastic volatility modeling, which comes in the next chapter, and time-dependent density estimation, which comes in the chapter following the next.

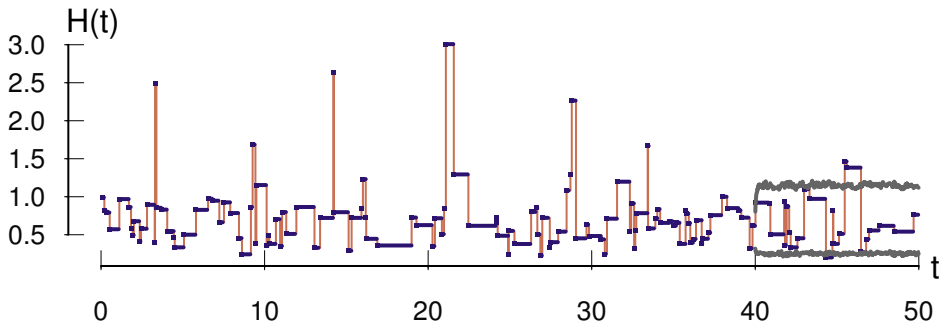


Figure 3.3: Trajectory of the SF-Harris process, where Q is a $\mathsf{GIG}(-2, 4, 1)$ and $\alpha = 3$; plus 0.9-HPD intervals computed with the first 40 days.

4 | A stochastic volatility model

Over this chapter we develop a novel stochastic volatility (SV) model where the spot volatility follows a SF-Harris process. Continuous-time models offer the natural framework for theoretical option pricing. Hence, they have dominated the literature since mid-1980s; being the major process used a Brownian motion timed changed by a random clock representing the volatility. Assuming friction-less markets, a weak no-arbitrage condition implies the asset log-price will be a semimartingale. This leads to the formulation $Y_t = A_t + B_{\tau_t^*}$, where A is a finite variation process and $B_{\tau_t^*}$ is a local martingale. A popular choice for A is $A_t = \mu t + \beta\tau_t^*$ for a pair of constants μ and β, often referred to as the drift and the risk premium. The time change τ^* is a random clock, which we define as a real-valued process with non-negative and non-decreasing sample paths.

At the outset, τ^* was assumed to be a Lévy subordinator (Clark, 1973). However, empirical data suggested that the amplitude of returns is positively autocorrelated in time. Surveys on this matter are given by Bollerslev et al. (1994); Ghysels et al. (1996); and Shephard (1996). To deal with the non-independence of the increments new continuous-time models have arisen. A popular class among them is the diffusion-based models. In this approach, the volatility is given as the solution to a univariate stochastic differential equation (see, e.g., Hull and White, 1987; Wiggins, 1987).

Because of the Dambis-Dubins-Schwarz theorem, the diffusions are in fact special cases of time-changed Brownian motions, where the time change is an integrated process, $\tau_t^* = \int_0^t \tau_s ds$. The process τ, identified as the spot volatility, is assumed to have almost surely locally square integrable sample paths, while being positive and stationary. The adoption of integrated processes makes sense theoretically, since they are a natural choice to consider the unobserved volatility period which appears when working with a time discretization. The Brownian motion is a martingale and $\tau_t^* = \int_0^t \tau_s ds$ is continuous and non-decreasing. Hence, for the process Y to be a semimartingale, $\mathbb{E}\Big[\sqrt{\tau_t}\Big] < \infty$, since this is necessary and sufficient to ensure that $\mathbb{E}\Big[\mid B_{\tau_t^*} \mid\Big] < \infty$.

Several models may originate by assuming different spot volatility processes τ. In

particular, the SV model introduced here lies in this class, where τ is the SF-Harris process in the case of Q being a positive distribution. Other examples are the standard Black-Scholes model, which is recovered when τ is a positive constant, and the notorious models proposed in Barndorff-Nielsen and Shephard (2001) (which we will term the BNS models), where τ follows an Ornstein-Uhlenbeck-type equation. An overview on SV models is given by Shephard and Andersen (2009).

4.1 The GIG-Harris SV model

To implement the proposed model, we will restrict our attention to the subclass in which the spot volatility has a one-dimensional marginal GIG distribution. This restriction enables us to perform a deeper study in the following sections; in which we develop its estimation, prediction, and its application to a real data set. Formally,

Definition 15. *We term the* ***GIG-Harris SV model*** *to the process*

$$Y_t = \mu t + \beta \tau_t^* + B_{\tau_t^*}, \tag{4.1}$$

where $\tau_t^* = \int_0^t \tau_s ds$, $\mu, \beta \in \mathbb{R}$, $B = (B_t)_{t \geq 0}$ *is a Brownian motion independent of* $\tau = (\tau_t)_{t \geq 0}$, *and* τ *is the SF-Harris process when* Q *is a GIG distribution.*

Although any positive marginal distribution Q defines a spot volatility model, we choose to use GIG distributions for various reasons. First, special cases, such as the Gamma, Inverse Gamma, or Inverse Gaussian, have been extensively used as spot volatilities in different kinds of models. Second, they have been proven to accurately fit real data (see, e.g., Gander and Stephens, 2007). Third, they can adjust to different scenarios since the choice of parameters will change their shape, skewness, and tail weight. Finally, GIG distributions meet some desirable properties; among them the property of being self-decomposable and, as a consequence, infinitely divisible.

Over Section 3.3 we developed properties for the integrated SF-Harris process, which is the process of integrated volatility in the GIG-Harris SV model. Similar properties may be derived for the return process. For example, second order results can be found in Barndorff-Nielsen and Shephard (2001), where they deal with the general case of the process τ in (4.1) being any second order or covariance stationary process. Also, we

saw that as $t \to \infty$, $t^{-1}H^*(t) \to \xi$ almost surely. As a consequence, the returns tend to normality. This desirable result is known as aggregational Gaussianity (Barndorff-Nielsen and Shephard, 2003).

4.2 Comparison with the BNS model

The BNS model and the GIG-Harris SV model share a number of attractive properties and differ in a few important ones. Let us start with the ones they share. In the BNS model, the spot volatility follows a stationary process of Ornstein-Uhlenbeck type (OU type); with the restriction that the background driving Lévy process (BDLP) has positive increments and no drift. This implies that the spot volatility jumps when the BDLP does it, and decays exponentially in-between. Therefore, in both models the spot volatility is a stationary and ergodic process, with positive jumps, and with an exponential autocorrelation function.

Moreover, in the BNS model, given a self-decomposable distribution, there is a unique BDLP that will generate that specific marginal distribution for the volatility. Hence, the modeler is allowed to choose the marginal distribution, in the same manner as in the SF-Harris process. In particular, when Q is self-decomposable, we can choose a BDLP in a way that the OU type and the SF-Harris process share also the mean and variance. As a consequence, the integrated volatility and the return processes will also be equivalent up to second order. Many of the powerful results which have been developed for estimating and forecasting returns rely only on second moments so they can be equally applied to the GIG-Harris SV model. Examples of these results may be found in Barndorff-Nielsen and Shephard (2001) or Sørensen (2000).

Now, among the similarities we stated that both processes are ergodic, meaning they will eventually converge to their invariant measure. However, the order of the speed of convergence is not the same. The SF-Harris process is uniformly ergodic, while the OU type processes are exponentially ergodic. Uniform ergodicity is a special case of exponential ergodicity, where the convergence to the invariant distribution is bounded independently of the starting point, so it leads to stronger ergodic theorems (in the terms of total variation convergence). We refer to Down et al. (1995) for more on this results.

Finally, an important difference between both models is the simplicity of their infer-

ence. Estimating the BNS model has not been easy. Novel simulation strategies have been developed over the years which allow us to test a part of them in real data (see, e.g., Gander and Stephens, 2007; Griffin and Steel, 2006, 2010). Nevertheless, we believe such difficulties have prevented the wide spread of this class of SV models in applications. For this reason, an easily tractable yet flexible model is of significant importance. The simple form of the transition functions of the SF-Harris process allows to develop the estimation methods of Section 3.5. These methods serve a basis to propose an efficient estimation procedure for the GIG-Harris SV model.

4.3 Extending the memory of the model

We mentioned that in both the SF-Harris and the OU type processes the spot volatility correlation decays exponentially. This will imply the resulting SV models, GIG-Harris and BNS, are short-memory models. Various authors have found empirical evidence suggesting that, in certain data sets, the dependence on the volatility structure decays at a hyperbolic rate for shorter lags; which is much slower than the exponential time-decay (see, e.g., Andersen and Bollerslev, 1997a).

Two streams of extensions for short-range models have been mainly applied to induce long-memory. The first one is to alter directly the short-memory volatility process by using superpositions. Generalizations of this kind are developed in Barndorff-Nielsen and Shephard (2003), where the BNS model is extended employing a weighted sum of independent OU type processes. A first case is to consider a finite number of independent processes. A new process σ^2 is defined as a superposition of OU type processes, $\sigma_1^2, ..., \sigma_m^2$, with different persistence rates $\lambda_1, ..., \lambda_m$, and independent BDLP $Z_1, ..., Z_m$ through

$$\sigma^2(t) = \sum_{j=1}^{m} w_j \sigma_j^2(t), \tag{4.2}$$

where the positive weights $w_1, ..., w_m$ add up to one. The autocorrelation function

$$r(t) = \sum_{j=1}^{m} w_j e^{-\lambda_j t} \tag{4.3}$$

is a weighted sum of exponentials. Thus, some of the volatility components may represent short term variations, while others represent long term movements. Many results from the

case $m = 1$ carry over to the superposition process. In particular, the integrated process can be derived explicitly.

In general, it is difficult to find a suitable value of m in equation (4.2). An alternative, given by Griffin (2010), is to assume an infinite number of components of which only a finite number have non-negligible weight. Now, while this aggregation mechanism provides a possible explanation of the long-range dependence in a time series, the models are still of short-memory. Formally, we shall say that a stationary stochastic process exhibits long-memory if its autocorrelation function has an asymptotic power-like behavior.

A second option to extend the BNS model that may lead to long-memory is to use an infinite number of independent OU type process. A new process σ^2 is defined as a superposition of OU type processes, σ^2_η, with different persistence rates λ_η, and independent BDLP Z_η through

$$\sigma^2(t) = \int_{\mathcal{R}} \sigma^2_\eta(t) F(d\eta)$$

where F is a probability distribution over $\mathcal{R} \subset \mathbb{R}^+$. For example, when F is a $\mathsf{Ga}\{2(1-H), 1\}$ law, the autocorrelation function is given by

$$r(t) = (1+t)^{-2(1-H)} \tag{4.4}$$

with $H \in (\frac{1}{2}, 1)$ being the long memory parameter.

Aiming for similar extension results with the SF-Harris process, we consider the following special case of the process ξ defined in (3.7).

Definition 16. *We define the* ***mixture SF-Harris process*** $\chi = (\chi_t)_{t\geq 0}$ *as a process given by* (3.7)*, where we let* Y *be as in Theorem 3.2.1,* $(R_n)_{n\in\mathbb{N}}$ *be a sequence of random variables independent of* Y*. We suppose the increments* $(V_n)_{n\in\mathbb{N}}$*, defined as* $V_n = R_n - R_{n-1}$*, are exponentially distributed,* $V_n \sim \mathsf{Exp}(\rho_n)$*, and* $\{\rho_n\}_{n\in\mathbb{N}}$ *are independent identically distributed random variables with some distribution* F*.*

It is direct to verify that

$$\mathbb{P}_x(\chi_t \in A) = \int_0^\infty \left\{(1 - e^{-\rho_1 t})Q(A) + e^{-\rho_1 t}\delta_x(A)\right\} F(d\rho_1).$$

Hence, the transition functions of the mixture SF-Harris process are a mixture of the

SF-Harris process transitions.

Now, the autocorrelation function of the mixture SF-Harris process matches the generating function of F since

$$r(t) = \mathbb{P}(V_1 > t) = \int_0^\infty e^{-t\rho_1} F(d\rho_1).$$

Therefore, subject to the dependence structure in a time series, an adequate function F can be chosen. For instance, when F is a degenerated distribution on the value α, the mixture SF-Harris process reduces to the SF-Harris process case, where $r(t) = e^{-\alpha t}$. When F is a discrete distribution that takes the value λ_j with probability w_j for $j = 1, ..., m$, we recover the autocorrelation (4.3) of the finite superposition of OU type processes. Even more, if F is a discrete distribution that may take a countably infinite number of values we get the extension of Griffin (2010). Last, when F is a $\mathsf{Ga}\{2(1-H), t+1\}$, then the autocorrelation function is given by (4.4), so it matches the autocorrelation of the infinite superposition of OU type processes.

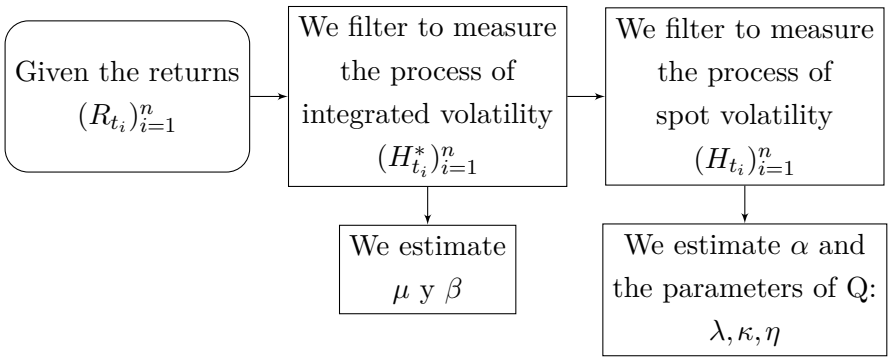
More general long-memory models can be obtained by using any heavy-tailed distribution F, and, as we saw in Section 3.4, a number of stability properties fulfilled by the SF-Harris process, such as wide-sense regeneration, ergodicity, and positive Harris recurrence, follow directly for the general model. Also, the integrated process can be derived explicitly. However, the process is neither uniformly nor exponentially ergodic. This makes sense since observations far away in the past remain correlated.

The second stream of extensions consists in leaving the volatility process unaltered while changing the shape of equation (4.1). This has been achieved by replacing the Brownian motion in the equation with a fractional Brownian motion; which is a generalization of Brownian motion that can have correlated increments (see, e.g., Comte and Renault, 1998). Indeed, fractional Brownian motion exhibits long-range dependence for certain parameters set (when the Hurst index $H \in (\frac{1}{2}, 1)$). Should the observed data need it, this construction could be further explored to extend our proposal of SV model. This would consist in replacing the Brownian motion by a normal approximation to fractional Brownian motion, assuring the no-arbitrage condition is maintained (see, e.g., Gander and Stephens, 2007).

4.4 Inference

4.4.1 Estimation method

In this section we develop a method to simulate from the posterior distributions of the parameters of the GIG-Harris SV model. Suppose we observe a high-frequency, discretized realization of the log-price process Y at times $t_1 < \cdots < t_n$. Then, the model is adjusted to returns, defined as $R_i = Y_{t_i} - Y_{t_{i-1}}$, for $i = 1, ..., n$, where we assume $Y_{t_0} = 0$. Given such returns, a series of filters are required to obtain a set of observations of the SF-Harris process. These are summarized in the following diagram.



The measurement of the integrated volatility process is performed using a common procedure based on the semimartingales quadratic variation. The procedure, found in Barndorff-Nielsen et al. (2002), consists in approximating the quadratic variation of Y with the realized variance. If S is a local semimartingale right continuous with left limits, then it follows that $S = A + M$, where A is a predictable process of locally bounded variation, and M is a local continuous martingale. The quadratic variation of S, denoted by $[S]$, is defined as

$$[S](t) = \lim_{M \to \infty} \sum_{j=0}^{M-1} \{S(t_{j+1}) - S(t_j)\}^2,$$

where $\{t_1, ..., t_M\}$ is any partition of the interval $[0, t]$ in M points. Indeed, when A is continuous, then $[S] = [M]$. Therefore, is direct to verify the quadratic variation of the GIG-Harris SV model matches the Brownian motion random clock, H^*

Assuming h is a reasonable division of the time in which we observed the trajectory, for

example, days, let M be the number of daily observations. The intra-day high frequency observations are defined as

$$r_{j,i} = R\left\{(i-1)h + \frac{hj}{M}\right\} - R\left\{(i-1)h + \frac{h(j-1)}{M}\right\},$$

and the realized variance as

$$[R_M]_i = \sum_{j=1}^{M} r_{j,i}^2.$$

The procedure of Barndorff-Nielsen et al. (2002) is based on the fact that

$$[R_M]_i \to [R]_i = [R](hi) - [R]\{h(i-1)\}$$

in probability when $M \to \infty$. Hence, the realized variance is a good estimator of the integrated volatility when M is large. This method has proven to work well for irregular time intervals, and when the random clock is continuous, as is the integrated SF-Harris process case. Now, given the integrated volatility observations, we perform the measurement of the spot volatility process by applying the common right-hand side derivative approximation.

$$H_{t_i} \approx \frac{H^*_{t_i} - H^*_{t_{i-1}}}{t_i - t_{i-1}}.$$

In the literature, there are multiple alternatives to filter both the integrated and spot volatility. In order to choose the best method to apply, we tried out a number of these alternatives, comparing them by computational time, and by the estimation errors obtained, which are computed with 100 simulated trajectories and add up the errors of all of the estimation steps (see Section 4.4.2). The errors show that the path properties, such as the jump rate and the values it ranges in, are well preserved with these simple filtering methods. If, however, we are interested in using the filtration for other purpose than the estimation of the SF-Harris process parameters, other methods could be tried out.

For filtering univariate integrated volatility, other methods include the bipower realized variation, which is used later on for the jump detection procedure, the two-time scales and the multi-scale estimator (Zhang, 2006; Zhang et al., 2005), and the realized kernel

estimators of Barndorff-Nielsen et al. (2008). Regarding the spot volatility, examples of alternative methods are the estimation by rolling and block sampling filters (Foster and Nelson, 1996), the kernel-weighted estimators (Kanaya and Kristensen, 2016) and the nonparametric kernel methods of Bandi and Renò (2009) which are robust to the presence of jumps.

Having the spot volatility observations, we proceed using such an approximation to estimate α and the parameters of Q, by means of the Gibbs-b method of Section 3.5.1.

Finally, for the estimation of μ and β, we have that

$$R_i \sim N(\mu(t_i - t_{i-1}) + \beta H_i^*, H_i^*),$$

where $H_i^* = H_{t_i}^* - H_{t_{i-1}}^*$. To facilitate the calculations display we assume that times are equidistant and h denotes the distance between them. Assigning $\pi(\mu,\beta) = \pi(\mu)\pi(\beta)$, with $\mu \sim \mathsf{N}(m_\mu, \sigma_\mu^2)$, and $\beta \sim \mathsf{N}(m_\beta, \sigma_\beta^2)$, the joint posterior density is given by

$$\pi(\mu,\beta|R_1,...,R_{n-1}) = f(R_1,...,R_{n-1}|\mu,\beta)\pi(\mu,\beta),$$

where

$$\begin{aligned} f(R_1,...,R_{n-1}|\mu,\beta) &= f(R_1,...,R_{n-1}|H_1^*,...,H_{n-1}^*,\mu,\beta) f(H_1^*,...,H_{n-1}^*|\mu,\beta) \\ &\propto \exp\left\{ \sum_{i=1}^{n-1} -\frac{1}{2H_i^*}(R_i - \mu h - \beta H_i^*)^2 \right\}, \end{aligned}$$

since $R_1|H_1^*, ..., R_{n-1}|H_{n-1}^*$ are independent, and $H_1^*, ..., H_{n-1}^*$ do not depend on μ or β.

After some algebra we obtain that, making

$$\bar{R} = \frac{1}{n-1}\sum_{i=1}^{n-1} R_i,\ \bar{R_2} = \frac{1}{n-1}\sum_{i=1}^{n-1} \frac{R_i}{H_i^*},\ \bar{H}^* = \frac{1}{n-1}\sum_{i=1}^{n-1} H_i^*,\ \text{and}\ \bar{H}_2^* = \frac{1}{n-1}\sum_{i=1}^{n-1} \frac{1}{H_i^*},$$

the last expression becomes

$$f(R_1,...,R_{n-1}|\mu,\beta) \propto \exp\left\{ -\frac{1}{2}\left(\mu^2 h^2 \bar{H}_2^* - 2\mu h \bar{R}_2 + 2h\mu\beta + \beta^2 \bar{H}^* - 2\beta\bar{R} \right) \right\}.$$

Since

$$\pi(\mu) \propto \exp\left\{-\frac{1}{2\sigma_\mu^2}\left(\mu^2 - 2\mu m_\mu\right)\right\} \text{ and } \pi(\beta) \propto \exp\left\{-\frac{1}{2\sigma_\beta^2}\left(\beta^2 - 2\beta m_\beta\right)\right\},$$

taking $A = h^2\bar{H}_2^* + (\sigma_\mu^2)^{-1}$, $B = h\bar{R}_2 + (m_\mu)(\sigma_\mu^2)^{-1}$, $C = h$, $D = \bar{H}^* + (\sigma_\beta^2)^{-1}$, and $E = \bar{R} + (m_\beta)(\sigma_\beta^2)^{-1}$, we get

$$\pi(\mu, \beta | R_1, ..., R_{n-1}) \propto \exp\left\{-\frac{1}{2}\left(\mu^2 A - 2\mu B + 2\mu\beta C + \beta^2 D - 2\beta E\right)\right\}.$$

Now marginalizing,

$$\begin{aligned}
\pi(\mu | R_1, ..., R_{n-1}) &\propto \exp\left\{-\frac{1}{2}\left(\mu^2 A - 2\mu B\right)\right\} \int \exp\left[-\frac{1}{2}\left\{\beta^2 D - 2\beta(E - \mu C)\right\}\right] d\beta \\
&\propto \exp\left\{-\frac{1}{2}\left(\mu^2 A - 2\mu B\right)\right\} \exp\left\{-\frac{1}{2}\left(-\frac{\mu^2 C^2 - 2\mu EC}{D}\right)\right\} \\
&\propto \exp\left[-\frac{1}{2}\left(A - \frac{C^2}{D}\right)\left\{\mu - \left(\frac{DB - EC}{AD - C^2}\right)\right\}^2\right].
\end{aligned}$$

So if we set $F = AD - C^2$, then $\mu|(R_1, ..., R_{n-1}) \sim N(\tilde{m_\mu}, \tilde{\sigma_\mu^2})$, with $\tilde{m_\mu} = (DB - EC)F^{-1}$, and $\tilde{\sigma_\mu^2} = DF^{-1}$.

With an analogous calculation it can be proven that $\beta|(R_1, ..., R_{n-1}) \sim N(\tilde{m_\beta}, \tilde{\sigma_\beta^2})$, with $\tilde{m_\beta} = (EA - BC)F^{-1}$, and $\tilde{\sigma_\beta^2} = AF^{-1}$.

4.4.2 Simulation study

To validate the estimation procedure described in the previous section we run a simulation study. Notice that the stochastic representation from Theorem 3.3.1 makes possible to simulate trajectories of the integrated SF-Harris process in a simple way. The testing procedure we employ is similar to the one presented in Section 3.5.2. We begin by drawing 100 values for μ, β, α, λ, κ, and η from a uniform distribution over the intervals $(-2, 2)$, $(-2, 2)$, $(0, 30)$, $(-5, 5)$, $(0, 50)$, and $(0, 4)$ respectively. With these parameter values, we proceed with the simulation of 100 series of spot volatility, integrated volatility, and returns. Subsequently, we run all of the estimation steps, and calculate the estimation

errors as:

$$E_\mu = \frac{1}{100}\sum_{i=1}^{100}\left|\frac{\mu_i - \hat{\mu}_i}{4}\right|,\ E_\beta = \frac{1}{100}\sum_{i=1}^{100}\left|\frac{\beta_i - \hat{\beta}_i}{4}\right|,\ E_\alpha = \frac{1}{100}\sum_{i=1}^{100}\left|\frac{\alpha_i - \hat{\alpha}_i}{30}\right|,\ \text{and}$$
$$E_Q = \frac{1}{100}\sum_{i=1}^{100}\mathsf{KL}\left\{\mathsf{GIG}(\lambda_i,\kappa_i,\eta_i),\mathsf{GIG}(\hat{\lambda}_i,\hat{\kappa}_i,\hat{\eta}_i)\right\}.$$

Such estimation errors are shown in Table IV. The errors of μ, β, and α are easily interpretable since they are relative errors, ponderated by each simulation interval length. However, that is not the case of the KL divergence, which can take any non-negative value. To get a sense of the measure of similarity that the KL divergence represents we refer to Figure 2.3; in which we fix a GIG distribution, and calculate the KL divergence for different approximations.

Table IV: Estimation errors for the GIG-Harris SV model parameters

E_μ	0.18
E_β	0.10
E_α	0.22
E_Q	0.20

4.4.3 Empirical analysis

Data and cleaning procedure

In this section, we apply the GIG-Harris SV model to the stock prices of IBM (International Business Machines). Let us start with an explanation of the data and the cleaning procedure applied. The three-year series we use, obtained from Kibot (2015), covers the period from January 2012 until December 2014. It records at every minute the open, high, low, and close prices, and the volume of IBM stocks; in the regular times of the US trading session, between 9:30 AM and 4:00 PM on workdays. The data are provided adjusted, using appropriate split and dividend multipliers adhering to the Center for Research in Security Prices standards.

To clean the data we implemented the step-by-step procedure proposed by Barndorff-

Nielsen et al. (2009). Part of the procedure is to delete (or replace) entries with a repeated time stamp, a zero transaction price, or a time stamp outside the exchange hours. No entries were found in any of these cases. Hence, we proceeded by creating a single variable that represents the stock price. We computed the standard deviation between the open and close price for each observation, finding it is less than 0.1 for 95 percent of the sample. This suggests that creating such a new variable by averaging the open and close prices makes sense, we call it *average price*.

Two remaining steps of the cleaning procedure were run using the average price. The first one is to delete entries for which the average price exceeds by more than 50 times the median on that day. Once again no entries were removed. The second step is to replace entries for which the average price deviated by more than 10 mean absolute deviations from a rolling centered median (excluding the observation under consideration) of 50 observations (25 observations before and 25 after). This was performed in a slightly different way, restricting the observations from the rolling median to be on the same day. A total of 154 entries was replaced in this step, which is around 0.05 percent of the sample size.

In Section 4.4.2 we proved we have the means to estimate and predict the GIG-Harris SV model satisfactorily so, at this point, we could have fitted the model to the cleaned data. However, performing an initial exploratory analysis we noticed that, as is common in practice, the estimation accuracy should be better when introducing a jump and a periodic component to the model. For this reason, in the next subsections we continue as follows. First, we introduce a jump component to the returns, and approximate the integrated and spot volatility processes. Next, we add a periodic component to the spot volatility. Taking into account both components, we then proceed with full estimation of the process, and, last, we conclude by testing such an estimation.

Adding a jump component

Figure 4.1 displays the trajectories of the log-average price and the returns. Sudden changes in the price level can be observed, which result in extremely large or extremely small returns, compared to the rest of them. To reflect this behavior, a customary practice is to model log-prices $Y = (Y_t)_{t\geq 0}$ generalizing the semimartingale (4.1) by adding a finite

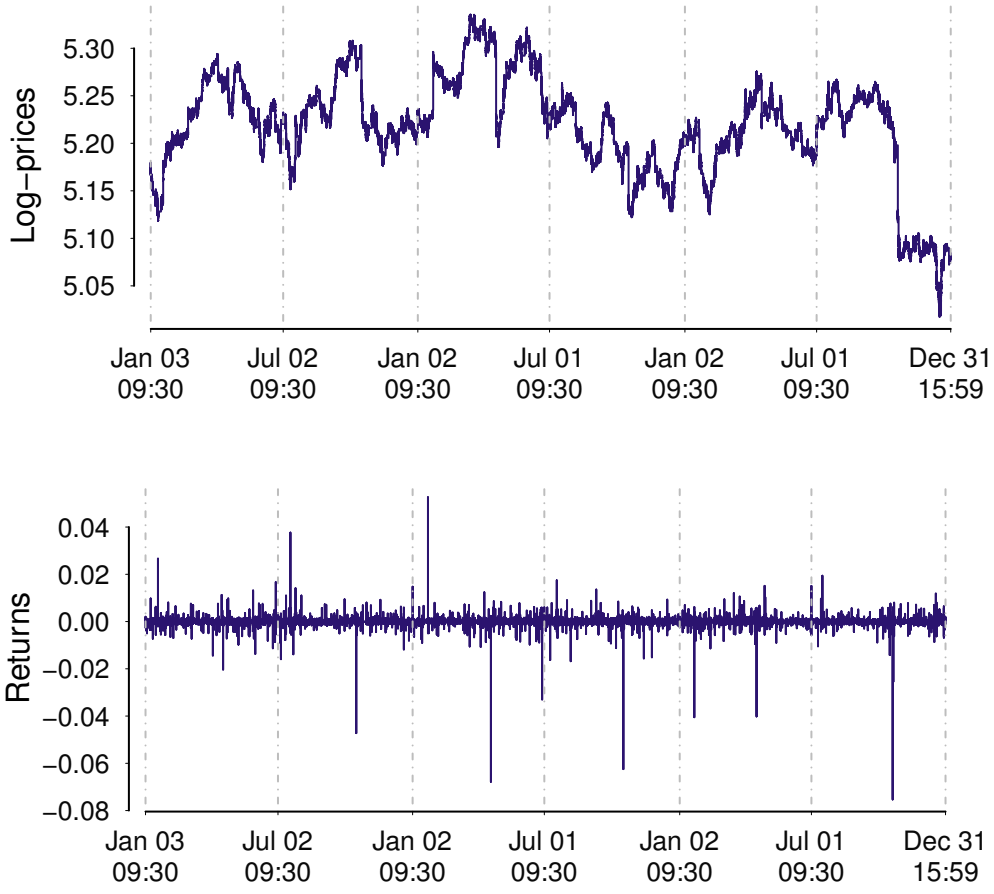


Figure 4.1: Log-prices and returns for IBM from 2012 to 2014.

activity jump process $J = (J_t)_{t\geq 0}$. That is

$$Y_t = \mu t + \beta\tau_t^* + B_{\tau_t^*} + J_t, \text{ where } \tau_t^* = \int_0^t \tau_s ds, \text{ and } J_t = \sum_{j=1}^{N_t} C_j, \tag{4.5}$$

The jump process J is assumed to be independent of B and τ. The process $N = (N_t)_{t\geq 0}$ counts the number of jumps that have occurred in the interval $[0, t]$; and $C = (C_t)_{t\geq 0}$ is a process such that for all t, (i) $C_t < \infty$, and (ii) $\sum_{j=1}^{N_t} C_j^2 < \infty$. These properties ensure the quadratic variation of Y is finite. For reviews of process (4.5) see, for instance, Andersen et al. (2007). As done in the GIG-Harris SV model, we suppose τ is the SF-Harris process in the case of Q being a GIG distribution.

The change in the model affects the measurement of the integrated volatility process, as the realized variance approximates the quadratic variation of Y; which in this case is given by $\tau^*(t) + \sum_{j=1}^{N_t} C_j^2$. To deal with this, we first implement a jump-detection procedure, and then perform the measurement of the integrated volatility without taking the jumps into consideration. When deleting the jumps from the sample, the remaining data can be modeled with the GIG-Harris SV model, where the quadratic variation matches the integrated SF-Harris process. Since realized variance approximates quadratic variation also when the observations are not equally spaced, there is no need to replace the jump entries, the crucial part is to detect them.

Many options to detect jumps are available, we explored a few of them, finding the best results were obtained when using the bipower variation, with which we detected around 90 percent of the jumps. The bipower variation, introduced in Barndorff-Nielsen and Shephard (2004), equals the quadratic variation of the continuous component and, in a range of cases, it produces an estimator of integrated volatility in the presence of jumps. It can be consistently estimated, and the estimator is called realized bipower variance.

The difference between the realized variance and the realized bipower variance is, consequently, a good indicator for the squared jumps; but nothing prevents this difference from becoming negative in a given finite sample. Thus, following Barndorff-Nielsen and Shephard (2004) suggestion, we calculate at every fifteen minutes the maximum value between such a difference and zero. The top 0.1 percent values are considered intervals with jumps. We explore individually those fifteen minute intervals, marking the entry that differs the most from the interval mean as a jump. Notice we mark exactly one jump in each interval. This could be restrictive and is fixed by running the procedure a few

times. We found that in these data it suffices to run it twice, so around 0.2 percent of the sample observations were considered as jumps and deleted.

Next, we proceeded with the measurement of the integrated and spot volatility processes, H^* and H, as described in Section 4.4.1. In doing so, the realized variance was computed based on fifteen minutes returns.

Adding a periodic component

Recurring events, such as opening, lunch, and closing of financial markets, cause the return volatility to vary systematically over the trading days and weeks. Taking into account this periodic structure may improve the volatility modeling. We present a general procedure to extract the periodic component of a spot volatility process τ; and prove it works for the special case of the SF-Harris process.

The main idea is based on Boudt et al. (2011). Following them, we make a partition of the data time interval in smaller time intervals of length d, called local windows; and consider the time transformation $c(s)$ indicating the position of s in the periodicity cycle. So, accordingly, $c(s) = s \bmod L$, where the cycle repeats itself every L days. Next, we define the periodicity function $f : [0, d] \to \mathbb{R}^+$ as

$$f(t) = \mathbb{E}\left[\frac{\tau_t}{\frac{1}{d}\int_0^d \tau_s ds}\right],$$

and the periodicity factor for each time t as $f(c(t))$. Finally, the process $\hat{\tau}_t = \dfrac{\tau_t}{f\{c(t)\}}$ is called the *periodically adjusted volatility*.

Notice two important properties. First, when τ is a strongly stationary process, for any $t^r \in [(r-1)d, rd]$,

$$\mathbb{E}\left[\frac{\tau(t^r)}{\frac{1}{d}\int_{(r-1)d}^{rd} \tau_s ds}\right] = \mathbb{E}\left[\frac{\tau\{c(t)\}}{\frac{1}{d}\int_0^d \tau_s ds}\right].$$

Therefore, it suffices to define the function f in the first local window, $[0, d]$, and then extend it to the rest of local windows. Second, by definition, f is a measurable function

and

$$\frac{1}{d}\int_0^d f(s)ds = 1. \tag{4.6}$$

To clarify the meaning of f let us think of local windows as days. Then f is the expected value of the spot volatility τ, divided by the average daily volatility $\frac{1}{d}\int_0^d \tau_s ds$. That is, when removing the daily volatility, we would expect that the changes in the remainder volatility were due to the periodicity. This is evident considering τ_t as a periodic factor depending on t multiplied by a constant effect of the volatility within each day; meaning that the spot volatility, after filtering out the periodicity, is approximately constant over each day. That is the main premise in the studies of Andersen and Bollerslev (1997b); and Boudt et al. (2011); and, with such a premise, is direct to verify that

$$\mathbb{E}\left[\frac{\tau_t}{\frac{1}{d}\int_0^d \tau_s ds}\right] = f(t). \tag{4.7}$$

In the case of the GIG-Harris SV model, the spot volatility after filtering out the periodicity is the SF-Harris process. This implies that, instead of assuming the periodically adjusted volatility constant within each day, we are thinking of it as a piece-wise constant process; with jumps determined by independent and identically distributed exponential random variables. Because of the Theorem 3.2.1, the following proposition ensures equality (4.7) holds in this case.

Proposition 3. *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of positive, finite mean, and exchangeable random variables, stochastically independent of a Poisson process N. Let $d > 0$, and $f : [0, d] \to \mathbb{R}^+$ be a measurable function such that $\frac{1}{d}\int_0^d f(s)ds = 1$. Then, for $t \in [0, d]$, it follows that*

$$\mathbb{E}\left[\frac{f(t)Y_{N_t}}{\frac{1}{d}\int_0^d f(s)Y_{N_s}ds}\right] = f(t).$$

The following lemma will be used in the proof of Proposition 3

Lemma 1. *For any sequence of exchangeable random variables $X_1, ..., X_n$, and any $a_1, ..., a_n \in$*

$\mathbb{R}$ *whose sum is nonzero, it holds that*

$$\mathbb{E}\left[\frac{X_j}{\sum_{i=1}^n a_i X_i}\right] = \frac{1}{\sum_{i=1}^n a_i}, \quad j = 1, ..., n.$$

Proof of Lemma 1. Let

$$k = \mathbb{E}\left[\frac{X_j}{\sum_{i=1}^n a_i X_i}\right]$$

for all $j = 1, ..., n$, then,

$$1 = \mathbb{E}\left[\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i X_i}\right] = \sum_{i=1}^n a_i \mathbb{E}\left[\frac{X_i}{\sum_{i=1}^n a_i X_i}\right] = k \sum_{i=1}^n a_i.$$

□

Proof of Proposition 3. The case $N_d = 0$ is trivial, hence, we assume $N_d > 0$. We will treat a simpler case first. If f is a measurable simple function, meaning that

$$f(t) = \sum_{j=1}^m B_j \mathbb{1}_{(b_{j-1}, b_j]}(t),$$

where $0 = b_0 < b_1 < \cdots < b_{m-1} < b_m = d$, and $B_1, ..., B_m \in \mathbb{R}$. Then, it follows that, taking $\mathcal{F}_d = \sigma(N_s : s \leq d)$,

$$\mathbb{E}\left[\frac{Y_{N_t}}{\frac{1}{d}\int_0^d f(s) Y_{N_s} ds}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{Y_{N_t}}{\frac{1}{d}\int_0^d f(s) Y_{N_s} ds} \,\middle|\, \mathcal{F}_d\right]\right].$$

Now, denoting with $S_1, S_2, ...$ the time between jumps of N; given $\mathcal{F}_d$, let $n = N_d$. We define $J_0 = 0$, $J_i = \sum_{k=0}^{i-1} S_k$, for $i = 1, ..., n-1$, and $J_n = d$. Next, we create a new partition of the interval $[0, d]$ by taking the common refinement of partitions $0 = b_0 < \cdots < b_m = d$ and $0 = J_0 < \cdots < J_n = d$ (consisting of all different points from the two partitions renamed in order). Suppose we get the partition $0 = v_0 < \cdots < v_r = d$. There is a representation of f in terms of such a partition,

$$f(t) = \sum_{j=1}^r \hat{B}_j \mathbb{1}_{(v_{j-1}, v_j]}(t),$$

and, consequently,

$$\int_0^d f(s)Y_{N_s}ds = \sum_{j=1}^{r} \hat{B}_j Y_j (v_j - v_{j-1}).$$

Moreover, $N_t \in \{0, ..., N_d\}$, so $S_0, ..., S_{N_d-1}$, N_t, and N_d are $\mathcal{F}_d$-measurable. Therefore, applying Lemma 1 we get

$$\mathbb{E}\left[\frac{Y_{N_t}}{\frac{1}{d}\int_0^d f(s)Y_{N_s}ds} \,\middle|\, \mathcal{F}_d \right] = \frac{1}{\frac{1}{d}\sum_{j=1}^{r} \hat{B}_j (v_j - v_{j-1})},$$

but

$$\frac{1}{d}\sum_{j=1}^{r} \hat{B}_j (v_j - v_{j-1}) = \frac{1}{d}\int_0^d f(s)ds = 1.$$

So we obtain the desired equality for measurable simple functions.

Now, if f is any measurable function, then there is an increasing sequence $(f_n)_{n\in\mathbb{N}}$ of measurable simple functions which converge to f (almost surely). Thus, as is customary in proofs, the desired equality follows using the case proven for measurable simple functions, and applying the Monotone Convergence Theorem. □

The most natural method for the periodic component estimation is to approximate the expected value with the sample mean,

$$\hat{f}(t) = \frac{1}{|G(t)|} \sum_{r\in G(t)} \frac{\tau_r}{\frac{1}{d}\int_{w(r)} \tau_s ds}, \tag{4.8}$$

with $G(t) = \{r : c(r) = t\}$, and $w(r)$ the local window which contains r; standardizing such an approximation so it meets condition (4.6). Nonetheless, estimator (4.8) may be biased in the presence of jumps. Since we have eliminated the sample jumps, we can apply it without such a concern. Still if, for a given t, the standard deviation of the $G(t)$ values exceeded the standard deviation of the sample; we computed the mean with those values lying below the 0.9-quantile. It is noteworthy that, when evidence of jump presence is found, robust estimation methods can be further explored; extending the ideas exposed in Boudt et al. (2011) to the GIG-Harris process.

We set d to one day and L to five, meaning the cycle repeats itself every week of five trading days, as is common in practice. The resulting periodic component graphic appears in Figure 4.2. After this, we continued with the measurement of the periodically adjusted volatility, to which we adjusted the SF-Harris process. In view of such a process having piece-wise constant paths; when two observations differed by less than a small value ϵ (we fixed $\epsilon \approx 10^{-5}$), we considered this as measurement noise, and assumed them to be equal to their average.

Results

To test the modeling we broke the sample into an estimation period (first 80 percent of the data), and a subsequent forecasting period (last 20 percent of the data). Next, we predicted probability intervals for the forecasting period with 1000 simulated trajectories. Repeating this for different values of interval probabilities, we computed Table V. If the procedure works correctly, as is occurring, the percentage of the original trajectory that falls within the prediction intervals should be similar to the interval probability. Therefore, the empirical results suggest the GIG-Harris SV model, including a jump and a periodic component, may be useful for forecasting.

Table V: Percentage of the original trajectory of IBM stock prices that falls within HPD intervals of probability p

p	%
0.25	25
0.50	51
0.75	75
0.85	84
0.90	89
0.95	93
1.00	97

Numeric comparison to the BNS model

We mentioned before that SF-Harris processes are equivalent up to second order to OU type processes. As an example, we developed for the IBM data the case of a Gamma stationary distribution. Here, there is a direct inference algorithm for the OU type process

since the BDLP is a compound Poisson process. Using this, we estimated point-wise the IBM data parameters with maximum likelihood estimation. As for the Gamma-SF-Harris process, we applied the EM method for the estimation. The empirical KL divergence from the estimated stationary Gamma distribution in the SF-Harris process and the one in the OU type process is 0.0064. Hence, the stationary distributions are very similar. For many other choices of marginal distribution, the BDLP of the OU type process will be a Lévy process with infinite rather than finite activity, so the estimation becomes harder and the comparison drags subjectivity since it depends on the applied estimation method.

4.5 Concluding remarks

In this chapter, we formulated a SV model based on SF-Harris processes. The model shares various attractive properties with the BNS model (Barndorff-Nielsen and Shephard, 2001), it generalizes Boudt et al. (2011) model (in which volatility is approximately constant over each day) when considering periodicity, and it provides a good approximation to observed market behavior when the marginal distributions belong to the GIG class. In this case, a Bayesian estimation technique was proposed and rigorously tested, and an example with IBM data was developed.

There are numerous possible extensions to the proposed SV model. Among the work we mentioned, (i) considering the microstructure noise, (ii) inducing long memory either using the mixture of SF-Harris processes or replacing the Brownian motion in the equation (4.1) by a normal approximation to fractional Brownian motion, and (iii) exploring jump robust estimation methods for the periodic component. In addition, all the methodologies can be equally applied to any marginal distribution so a different family of marginal distributions may be explored. It should also be noted that this study dealt with the one-dimensional case, yet the SV model definition, and its respective research, can be extended to multidimensional cases.

In the following chapter we will introduce another application of SF-Harris processes, this time to time-dependent density models.

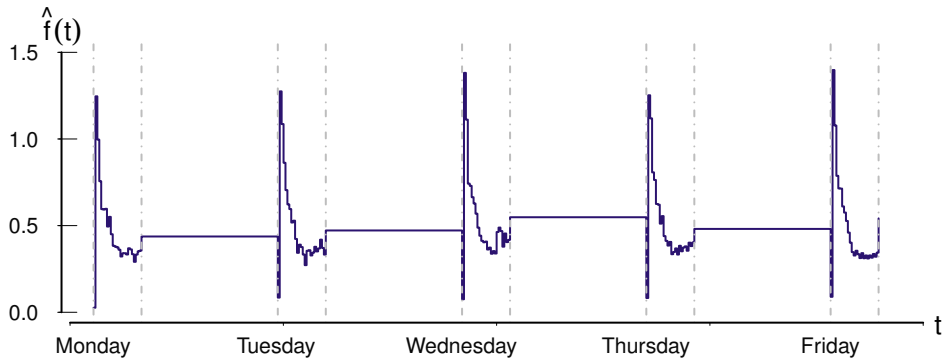


Figure 4.2: Approximation of the periodicity function of IBM stock prices. Average of the spot volatility standardized by daily volatility over all the values which have the same position in the periodicity cycle.

5 Time dependent density estimation

This chapter introduces the reader to the construction of an infinite mixture model in a non-parametric Bayesian paradigm. Based on SF-Harris processes, we introduce a new class of nonparametric prior distributions on the space of continuously varying densities.

A key problem in statistical modeling is model selection, how to choose a model at an appropriate level of complexity. This problem appears in many settings, one of them is choosing the number of clusters in mixture models. Traditional finite mixture models group data into a fixed number of latent clusters. A common solution to address the problem of choosing the number of clusters is, first, to fit several models with different numbers of clusters, and then to choose one using model selection metrics. Bayesian non-parametric methods side-step this issue by allowing the data to determine the complexity of the model. The approach consists in fitting a single mixture model that can adapt its complexity to the data. This is done by allowing the mixture to have infinite terms, setting a prior over the mixing distribution and a prior over the cluster parameters.

A simple case of a nonparametric mixture is when we are interested in estimating a single distribution from an independent and identically distributed sample $y_1, ..., y_m$, and we assume the observations come from a convolution

$$y_t \sim \int k(\cdot|\phi) G(d\phi),\ t = 1, ..., m,$$

where $k(\cdot|\phi)$ is a density function with parameter ϕ, and G is a mixing distribution, which is assigned a flexible prior.

We say that G follows a stick-breaking prior, with centering measure G_0, and shape measure H on $(0, 1)$, if and only if it admits a representation of the form:

$$G(\cdot) = \sum_{l=1}^{\infty} w_l \delta_{\theta_l}(\cdot),$$

where the *particles* $\{\theta_l\}_{l=1}^{\infty}$ are independent and identically distributed from G_0 and the

weights $\{w_l\}_{l=1}^{\infty}$ are defined by

$$w_l = v_l \prod_{r<l}(1-v_r),$$

For a set of *sticks* $\{v_l\}_{l=1}^{\infty}$ independently distributed from H. For example, assuming that G follows a Dirichlet process prior leads to the well known Dirichlet process mixture models, introduced by Lo (1984).

To model the situation where the underlying mixing distribution evolves with time, the simplest mixture model may be modified by making the distribution G time dependent. Hence, we are interested in estimating a set of distributions $\{G_t\}_{t\geq 0}$ from a time dependent sample $y_1, ..., y_m$, and we assume the observations come from a convolution

$$y_t \sim \int k(\cdot|\phi) G_t(d\phi),\ t = 1, ..., m, \tag{5.1}$$

where $k(\cdot|\phi)$ is a density function with parameter ϕ, and for each t, G_t is a random density given by

$$G_t(\cdot) = \sum_{l=1}^{\infty} w_l(t)\delta_{\theta_l(t)}(\cdot), \tag{5.2}$$

where

$$w_l(t) = v_l(t) \prod_{r<l}\{1-v_r(t)\}, \tag{5.3}$$

and $\{\{\theta_l(t)\}_{l=1}^{\infty}\}_{t\geq 0}$ and $\{\{v_l(t)\}_{l=1}^{\infty}\}_{t\geq 0}$ are independent collections of independent stochastic processes each taking values on $\mathbb{X} \subset \mathbb{R}$ and $(0,1)$ respectively. Indeed, with this definition, $\sum_l w_l(t) = 1$ for any $t \geq 0$.

Moreover, the observations need not be equidistant and there may be multiple data points available at every time. Therefore, we can assume more generally that we have p independent random realizations of the time dependent variables $y_{t_1}, \ldots, y_{t_m}$, where $t_1 \leq \ldots \leq t_m$ are not necessarily equally spaced. The data structure can be organized in a $p \times m$ matrix Y,

$$\mathsf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{p1} & \cdots & y_{pm} \end{bmatrix},$$

where each row represents a time dependent trajectory, and each column represents independent identically distributed realizations. Such data structure is habitually found for example in longitudinal data analysis, where patients or units samples are observed repeatedly over time and each patient or unit can be assumed to be independent of each other.

The design of such models is a challenging task because, on one hand, there is a need for priors with large enough support and, on the other, the prior choice may easily lead to identifiability issues or fall within the class of ill-posed problems. Hence, the collections of stochastic processes generating the particles and the weights need to have ability to share information between times, flexibility to capture the changes in the density, and simplicity to allow for estimation algorithms. A class of stochastic models satisfying those conditions is the one of SF-Harris processes.

5.1 SF-Harris BMP

We propose a novel approach to construct rich and flexible families of nonparametric priors, inducing the time dependence through both the weights and particles with independent collections of SF-Harris processes. The fact that the transition densities are known explicitly and, even more, tractable, makes SF-Harris processes an attractive alternative.

Definition 17. *The SF-Harris Bayesian mixture process (**SF-Harris BMP**) is given by Equations* (5.1)-(5.3)*, where* $\{\{\theta_l(t)\}_{l=1}^{\infty}\}_{t\geq 0}$ *and* $\{\{v_l\}_{l=1}^{\infty}\}_{t\geq 0}$ *are independent collections of independent SF-Harris process, meaning that, for each* $l = 1, 2, ...,$

$$S_t^l(\theta_0, B) := \mathbb{P}(\theta_l(t) \in B | \theta_l(0) = \theta_0) = (1 - \alpha_\theta^t) G_0(B) + \alpha_\theta^t \delta_{\theta_0}(B),$$
$$T_t^l(v_0, B) := \mathbb{P}(v_l(t) \in B | v_l(0) = v_0) = (1 - \alpha_v^t) \pi_v(B) + \alpha_v^t \delta_{v_0}(B),$$

where $\alpha_\theta, \alpha_v \in (0,1)$*, and* G_0 *and* π_{v} *are probability distributions over* $\mathbb{R}$ *and* $(0,1)$ *respectively.*

The transition probabilities in definition 17 match the SF-Harris process tranistions by making $\alpha_\theta = e^{-\nu_\theta}$ and $\alpha_v = e^{-\nu_v}$ for some $\nu_\theta, \nu_v > 0$.

Notice this proposal implies that the jump process for the sticks and particles induces

positive correlation. As an example, if π_v coincides with a $\mathsf{Be}(1, M)$ distribution, then the invariant distribution of G_t is a Dirichlet process centered at $\mathbb{E}[G_t] = G_0$. We can also prove the SF-Harris BMP is non-Markovian. However, the Markov property is retained with respect to the filtration generated by $(w(\cdot), \theta(\cdot))$, implying that the process is a particular case of a semi-Markov processes, property which makes prediction possible after the inference.

The nonparametric Bayesian mixture model approach has been studied by various authors using different strategies. Of particular interest for the purposes of this work are the developments of dependent processes moving in time, meaning that the random probability measures are as in (5.2), time dependent. In this regard we mention, among others, Dunson (2006), who models the dependent process as an autoregression with Dirichlet distributed innovations; Griffin and Steel (2011), who construct strictly stationary measures (5.2) with stick-breaking marginals by reducing the innovation to a single atom sampled from the centering measure; Caron et al. (2008), who model the noise in a dynamic linear model with a Dirichlet process mixture; Rodriguez and Ter Horst (2008), who induce the dependence in time only via the particles making them a random walk; Rodriguez and Dunson (2011), who construct the weights of the process as probit transformations of normal random variables; and Mena and Ruggiero (2016), who use one-dimensional Wright-Fisher diffusions for the sticks. Our proposal will add to this literature by introducing a model with the time dependence in both the weights and the particles that still remains statistically tractable.

5.2 Correlation between the random measures

Part of the ability of time dependent density models to capture the changes in the shape of the density across time is given by the r-lagged correlation between the random measures. The following proposition provides a general expression for such a correlation.

Proposition 4. *Let $G_t(\cdot)$ be the random measure of the SF-Harris BMP defined in* (5.2) *then*

$$\mathsf{Corr}[G_t(B), G_{t+r}(B)] = \frac{\{2\mu_v - \mu_v^{(2)}\}(\alpha_v^r \sigma_v + \mu_v^2)\alpha_\theta^r}{\mu_v^{(2)}(2\mu_v - \mu_v^2 - \alpha_v^r \sigma_v)} \tag{5.4}$$

where, taking $v \sim \pi_v$, $\mu_v := \mathbb{E}[v]$, $\mu_v^{(2)} := \mathbb{E}\left[v^2\right]$, and $\sigma_v := \mathsf{Var}(v)$.

Proof. To simplify the notation let $t' = t + r$, then,

$$\mathsf{Cov}[G_t(B), G_{t'}(B)] = \mathbb{E}[G_t(B)\, G_{t'}(B)] - G_0^2(B),$$

where

$$\begin{aligned}
\mathbb{E}\left[G_t(B)\, G_{t'}(B)\right] &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\mathbb{E}[w_k(t)w_l(t')\delta_{\theta_k(t)}(B)\delta_{\theta_l(t')}(B)] \\
&= \sum_{k=1}^{\infty}\mathbb{E}[w_k(t)w_k(t')]\mathbb{E}[\delta_{\theta_k(t)}(B)\delta_{\theta_k(t')}(B)] \\
&+ \sum_{k=1}^{\infty}\sum_{l=k+1}^{\infty}\mathbb{E}[w_k(t)w_l(t')]\mathbb{E}[\delta_{\theta_k(t)}(B)\delta_{\theta_l(t')}(B)] \\
&+ \sum_{l=1}^{\infty}\sum_{k=l+1}^{\infty}\mathbb{E}[w_k(t)w_l(t')]\mathbb{E}[\delta_{\theta_k(t)}(B)\delta_{\theta_l(t')}(B)].
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}[w_k(t)w_k(t')] &= \mathbb{E}\left[v_k(t)\prod_{j=1}^{k-1}\{1-v_j(t)\}v_k(t')\prod_{j=1}^{k-1}\{1-v_j(t')\}\right] \\
&= \mathbb{E}\left[v_k(t)v_k(t')\right]\prod_{j=1}^{k-1}\mathbb{E}\left[1-v_j(t')-v_j(t)+v_j(t)v_j(t')\right] \\
&= \varphi_{tt'}\prod_{j=1}^{k-1}(1-2\mu_v+\varphi_{tt'}) \\
&= \varphi_{tt'}(1-2\mu_v+\varphi_{tt'})^{k-1},
\end{aligned}$$

where $\varphi_{tt'} := \mathbb{E}\left[v_k(t)v_k(t')\right]$ and $\mu_v := \mathbb{E}[v_k(t)] = \mathbb{E}[v_k(t')]$.

On the other hand, we have that

$$\begin{aligned}
\mathbb{E}[\delta_{\theta_k(t)}(B)\delta_{\theta_k(t')}(B)] &= \mathbb{E}\left[\mathbb{E}[\mathbb{1}_B\{\theta_k(t)\}\mathbb{1}_B\{\theta_k(t')\} \mid \theta_k(t) = \theta_0\}]\right] \\
&= \mathbb{E}\left[\mathbb{E}[\mathbb{1}_B\{\theta_0\}\mathbb{1}_B\{\theta_k(t')\} \mid \theta_k(t) = \theta_0\}]\right] \\
&= G_0(B)\{(1-\alpha_\theta^r)G_0(B) + \alpha_\theta^r\} \\
&= (1-\alpha_\theta^r)G_0^2(B) + \alpha_\theta^r G_0(B).
\end{aligned}$$

For $l > k$, we have

$$\begin{aligned}
\mathbb{E}[w_k(t)w_l(t')] &= \mathbb{E}\left[v_k(t)v_l(t')\{1-v_k(t')\}\prod_{j=1}^{k-1}\{1-v_j(t)\}\{1-v_j(t')\}\prod_{s=k+1}^{l-1}\{1-v_s(t')\}\right] \\
&= \mathbb{E}\left[\{v_k(t)v_l(t')-v_k(t)v_l(t)v_k(t')\}\prod_{j=1}^{k-1}\{1-v_j(t')-v_j(t)+v_j(t)v_j(t')\}\right] \\
&\times \mathbb{E}\left[\prod_{s=k+1}^{l-1}\{1-v_s(t')\}\right] \\
&= \Big\{\mathbb{E}[v_l(t')]\mathbb{E}[v_k(t)]-\mathbb{E}[v_l(t)]\mathbb{E}[v_k(t)v_k(t')]\Big\}\prod_{j=1}^{k-1}\mathbb{E}[1-v_j(t')-v_j(t)+v_j(t)v_j(t')] \\
&\times \prod_{s=k+1}^{l-1}\mathbb{E}[1-v_s(t')] \\
&= \mu_v(\mu_v-\varphi_{tt'})(1-2\mu_v+\varphi_{tt'})^{k-1}(1-\mu_v)^{l-k-1},
\end{aligned}$$

and, since θ_l and θ_k are independent processes,

$$\mathbb{E}[\delta_{\theta_k(t)}(B)\delta_{\theta_l(t')}(B)] = \mathbb{P}\{\theta_k(t)\in B\}\mathbb{P}\{\theta_l(t')\in B\} = G_0^2(B).$$

Finally, for $l < k$ we obtain

$$\begin{aligned}
\mathbb{E}[w_k(t)w_l(t')] &= \mathbb{E}\left[v_k(t)v_l(t')\{1-v_k(t')\}\prod_{j=1}^{l-1}\{1-v_j(t)\}\{1-v_j(t')\}\prod_{s=l+1}^{k-1}\{1-v_s(t')\}\right] \\
&= \mathbb{E}\left[\{v_k(t)v_l(t')-v_k(t)v_l(t)v_k(t')\}\prod_{j=1}^{l-1}\{1-v_j(t')-v_j(t)+v_j(t)v_j(t')\}\right] \\
&\times \mathbb{E}\left[\prod_{s=l+1}^{k-1}(1-v_s(t'))\right] \\
&= \Big\{\mathbb{E}[v_l(t')]\mathbb{E}[v_k(t)]-\mathbb{E}[v_l(t)]\mathbb{E}[v_k(t)v_k(t')]\Big\} \\
&\times \prod_{j=1}^{l-1}\mathbb{E}[1-v_j(t')-v_j(t)+v_j(t)v_j(t')]\prod_{s=l+1}^{k-1}\mathbb{E}[1-v_s(t')] \\
&= \mu_v(\mu_v-\varphi_{tt'})(1-2\mu_v+\varphi_{tt'})^{l-1}(1-\mu_v)^{k-l-1},
\end{aligned}$$

and

$$\mathbb{E}[\delta_{\theta_k(t)}(B)\delta_{\theta_l(t')}(B)] \;=\; G_0^2(B).$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[G_t(B)\,G_{t'}(B)\right] &= \sum_{k=1}^{\infty}\varphi_{tt'}(1-2\mu_v+\varphi_{tt'})^{k-1}[(1-\alpha_\theta^r)G_0^2(B)+\alpha_\theta^r G_0(B)] \\
&+ \sum_{k=1}^{\infty}\sum_{l=k+1}^{\infty} G_0^2(B)\mu_v(\mu_v-\varphi_{tt'})(1-2\mu_v+\varphi_{tt'})^{k-1}(1-\mu_v)^{l-k-1} \\
&+ \sum_{l=1}^{\infty}\sum_{k=l+1}^{\infty} G_0^2(B)\mu_v(\mu_v-\varphi_{tt'})(1-2\mu_v+\varphi_{tt'})^{l-1}(1-\mu_v)^{k-l-1} \\
&= \{(1-\alpha_\theta^r)G_0^2(B)+\alpha_\theta^r G_0(B)\}\varphi_{tt'}\sum_{k=1}^{\infty}(1-2\mu_v+\varphi_{tt'})^{k-1} \\
&+ G_0^2(B)\mu_v(\mu_v-\varphi_{tt'})\sum_{k=1}^{\infty}(1-2\mu_v+\varphi_{tt'})^{k-1}\sum_{l=k+1}^{\infty}(1-\mu_v)^{l-k-1} \\
&+ G_0^2(B)\mu_v(\mu_v-\varphi_{tt'})\sum_{l=1}^{\infty}(1-2\mu_v+\varphi_{tt'})^{l-1}\sum_{k=l+1}^{\infty}(1-\mu_v)^{k-l-1} \\
&= \frac{\{(1-\alpha_\theta^r)G_0^2(B)+\alpha_\theta^r G_0(B)\}\varphi_{tt'}}{2\mu_v-\varphi_{tt'}} + 2\frac{G_0^2(B)(\mu_v-\varphi_{tt'})}{2\mu_v-\varphi_{tt'}} \\
&= \frac{\{(1-\alpha_\theta^r)G_0^2(B)+\alpha_\theta^r G_0(B)\}\varphi_{tt'}+G_0^2(B)(2\mu_v-2\varphi_{tt'})}{2\mu_v-\varphi_{tt'}} \\
&= \frac{\alpha_\theta^r G_0(B)\varphi_{tt'}+G_0^2(B)(2\mu_v-\varphi_{tt'}-\alpha_\theta^r\varphi_{tt'})}{2\mu_v-\varphi_{tt'}}.
\end{aligned}$$

This implies the following equality for the covariance

$$\begin{aligned}
\mathsf{Cov}[G_t(B),G_{t'}(B)] &= \mathbb{E}[G_t(B)\,G_{t'}(B)]-G_0^2(B) \\
&= \frac{\alpha_\theta^r G_0(B)\varphi_{tt'}+G_0^2(B)(2\mu_v-\varphi_{tt'}-\alpha_\theta^r\varphi_{tt'}-2\mu_v+\varphi_{tt'})}{2\mu_v-\varphi_{tt'}} \\
&= \frac{\alpha_\theta^r G_0(B)\varphi_{tt'}+G_0^2(B)(-\alpha_\theta^r\varphi_{tt'})}{2\mu_v-\varphi_{tt'}} \\
&= \frac{\alpha_\theta^r\varphi_{tt'}G_0(B)\{1-G_0(B)\}}{2\mu_v-\varphi_{tt'}}.
\end{aligned}$$

Consequently the variance is given by

$$\mathsf{Var}[G_t(B)] = \frac{\varphi_{tt} G_0(B)\{1 - G_0(B)\}}{2\mu_v - \varphi_{tt}},$$

and, using that $\varphi_{tt} = \varphi_{t't'}$, we have that the correlation is given by

$$\mathsf{Corr}[G_t(B), G_{t'}(B)] \;=\; \frac{\varphi_{tt'}(2\mu_v - \varphi_{tt})\alpha_\theta^r}{\varphi_{tt}(2\mu_v - \varphi_{tt'})}.$$

Replacing the value of $\varphi_{tt'}$ and φ_{tt} in the last expression we obtain 5.4. □

Corollary 2. *From Proposition* (4) *we have the following*

a) *When* $\alpha_v = \alpha_\theta = 0$ *then*

$$S_t^l(\theta, B) = \mathsf{G}_0(B),$$
$$T_t^l(v, B) = \pi_v(B).$$

This corresponds to the case of complete time independence, meaning that $G_t(B)$ *and* $G_{t+r}(B)$ *are independent for any* $r > 0$ *and state* B*. As a consequence,* $y_{t_1}, ..., y_{t_n}$ *in 5.1 are independent. Clearly, in this case* $\mathsf{Corr}[G_t(B), G_{t+r}(B)] = 0$.

b) *When* $\alpha_v = \alpha_\theta = 1$

$$S_t^l(\theta, B) = \delta_\theta(B),$$
$$T_t^l(v, B) = \delta_v(B).$$

which implies that for any $t > 0$

$$G_t(B) = \sum_{l=1}^{\infty} w_l \, \delta_{\theta_l}(B),$$

so it reduces to the simplest mixture model described.

This corresponds to the case of complete time dependence, meaning that $G_t(B) = G_{t+r}(B)$ *for any* B*. As a consequence* $\mathsf{Corr}[G_t(B), G_{t+r}(B)] = 1$*, and* $y_{t_1}, ..., y_{t_n}$ *in 5.1 are identically distributed.*

c) *When $\alpha_\theta = 1$, $\alpha_v \in (0,1)$, and π_v has positive variance,*

$$G_t(B) = \sum_{l=1}^{\infty} w_l(t)\, \delta_{\theta_l}(B). \tag{5.5}$$

Hence, the particles are random but do not vary with time, and the dependence is induced only via the weights, which is the case considered, for example, by Mena and Ruggiero (2016). In this case, as $r \to \infty$

$$\mathsf{Corr}[G_t(B), G_{t+r}(B)] \;\; \to \;\; \frac{\mu_v^2\{2\mu_v - \mu_v^{(2)}\}}{\mu_v^{(2)}(2\mu_v - \mu_v^2)}. \tag{5.6}$$

Notice Equation (5.6) *is less than one if and only if*

$$2\mu_v\mu_v^2 < 2\mu_v\mu_v^{(2)},$$

which follows since $\sigma_v = \mu_v^{(2)} - \mu_v^2 > 0$.

Now, if $v \sim \pi_v$ we have that $0 < v < 1$. Hence, $0 < v^2 < v$ and, consequently, $0 < \mu_v^{(2)} < \mu_v$. Therefore,

$$2\mu_v - \mu_v^2 > 2\mu_v - \mu_v^{(2)} > 2\mu_v - \mu_v = \mu_v > 0,$$

and Equation (5.6) *is always positive. This implies the existence of an upper bound for the correlation of all dependent processes whose particles are fixed, which could be an important restriction in some applications.*

d) *When $\alpha_\theta \in (0,1)$ and $\alpha_v = 1$,*

$$G_t(B) = \sum_{l=1}^{\infty} w_l\, \delta_{\theta_l(t)}(B), \tag{5.7}$$

meaning that the time dependence is induced only via the particles. This is the case studied, for example, by Rodriguez and Ter Horst (2008). In this setting, as $r \to \infty$

$$\mathsf{Corr}[G_t(B), G_{t+r}(B)] \to 0.$$

Therefore, the correlation does not have a lower bound. However, there are a number of theoretical and practical advantages of having time-dependent weights. For example, models with time-dependent weights have richer support. In fact, when

the weights are not time-dependent the models cannot generate a set of independent measures (MacEachern, 2000).

e) *When* $\alpha_v \in (0,1)$ *and* $\alpha_\theta \in (0,1)$, *as* $r \to \infty$

$$\mathsf{Corr}[G_t(B), G_{t+r}(B)] \to 0.$$

Consequently, under our proposal, the correlation between the random measures reaches the limits 0 and 1 for some conditions, avoiding the disadvantage of working with a lower bound as in constant-particles models, and providing richer support than constant-weights models.

As an example, let us present the correlation of a dependent Dirichlet process presented in MacEachern (2000). In this process, the stick-breaking weights are defined through independent realizations $\{v_k(t)\}_{t\geq 0}$ for $k = 0, 1, ...$ from a stochastic process with marginals $v_k(t) \sim \text{Beta}(1, M)$ for $M > 0$. Notice that for any fixed t, such a construction yields a Dirichlet process prior distribution for G_t. For the $\text{Beta}(1, M)$ distribution we have that $\mu_v = \frac{1}{1+M}$, $\mu_v^{(2)} = \frac{2}{(1+M)(2+M)}$ and $\sigma_v = \frac{M}{(1+M)^2(2+M)}$. Hence, using (5.4), we obtain

$$\begin{aligned}
\mathsf{Corr}[G_t(B), G_{t+r}(B)] &= \frac{\left\{\frac{2}{1+M} - \frac{2}{(1+M)(2+M)}\right\}\left\{\alpha_v^r \frac{M}{(1+M)^2(2+M)} + \frac{1}{(1+M)^2}\right\}\alpha_\theta^r}{\frac{2}{(1+M)(2+M)}\left\{\frac{2}{1+M} - \frac{1}{(1+M)^2} - \alpha_v^r \frac{M}{(1+M)^2(2+M)}\right\}} \\
&= \frac{\left\{\frac{2}{2+M}\right\}\left\{\frac{M\alpha_v^r+2+M}{(1+M)^2(2+M)}\right\}\alpha_\theta^r}{\frac{2}{(1+M)^2(2+M)}\left\{2 - \frac{1}{(1+M)} - \alpha_v^r \frac{M}{(1+M)(2+M)}\right\}} \\
&= \frac{\left\{\frac{M\alpha_v^r+2+M}{(1+M)^2(2+M)}\right\}\alpha_\theta^r}{\frac{1}{(1+M)^2}\left\{\frac{2(1+M)(2+M)-(2+M)-\alpha_v^r M}{(1+M)(2+M)}\right\}} \\
&= \frac{(1+M)(M\alpha_v^r+2+M)\alpha_\theta^r}{2(1+M)(2+M)-(2+M)-\alpha_v^r M} \\
&= \frac{(1+M)(M\alpha_v^r+2+M)\alpha_\theta^r}{(2+M)(1+2M)-\alpha_v^r M}
\end{aligned}$$

5.3 Current research

In this chapter we explored the use of SF-Harris processes in time-dependent density models. We proposed an approach to construct nonparametric priors for BMP inducing the time dependence through the weights and particles with independent collections of SF-Harris processes. These implies the correlation between the random measures reaches the limits zero and one for some conditions, avoiding the disadvantage of working with a lower bound as in constant-particles models, and providing richer support than constant-weights models.

The posterior computation and a real-data application are still under development. The main difficulty with a Gibbs sampler is that there are an infinite number of random variables that need to be sampled. To circumvent this problem, Walker (2007) proposes to use slice sampling (Neal, 2003) to truncate the representation adaptively to a finite number of features. A couple of years later Kalli et al. (2011) made a proposal to make the algorithm more efficient. We will base our strategy for posterior computation on this latest modification.

We denote with $v_k(t) := \{v_l(t)\}_{l=1}^k$, $\theta_k(t) := \{\theta_l(t)\}_{l=1}^k$ for $k = 1, ..., \infty$, and $f_t\{\cdot \mid \mathsf{v}_\infty(t), \theta_\infty(t)\} := \int k(\cdot \mid \phi) G_t(d\phi)$ which, using (5.2), can be written as

$$f_t\{y \mid \mathsf{v}_\infty(t), \theta_\infty(t)\} = \sum_{l=1}^{\infty} w_l(t)\, k\{y \mid \theta_l(t)\}. \tag{5.8}$$

Now, we apply the slice algorithm in Kalli et al. (2011), augmenting the density (5.8) to

$$f_t\{y, u, s \mid \mathsf{v}_\infty(t), \theta_\infty(t)\} = \frac{1}{\psi_s} \mathbb{1}_{(u<\psi_s)}\, w_s(t)\, k\{y \mid \theta_s(t)\},$$

where $s \mapsto \psi_s$ is a $\mathbb{N}-$valued decreasing function with known inverse ψ^*. The latent variable s indexes the specific kernel $k\{\cdot|\theta_s(t)\}$ that better captures the mass at y and, given s, $u \sim \mathsf{U}(0, \psi_s)$.

Next, denoting with $v_k^{(m)} := \{v_k(t_j)\}_{j=1}^m$, $\theta_k^{(m)} := \{\theta_k(t_j)\}_{j=1}^m$, for $k = 1, ..., \infty$,

$$\mathsf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pm} \end{bmatrix}, \text{ and } \quad \mathsf{U} = \begin{bmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pm} \end{bmatrix},$$

the conditional augmented likelihood is given by

$$\mathcal{L}\left\{\mathsf{Y}, \mathsf{U}, \mathsf{S} \mid v_\infty^{(m)}, \theta_\infty^{(m)}\right\} = \prod_{i=1}^{p} \prod_{j=1}^{m} \frac{1}{\psi_{s_{ij}}} \mathbb{1}_{(u_{ij} < \psi_{s_{ij}})} \left[v_{s_{ij}}(t_j) \prod_{k < s_{ij}} \{1 - v_k(t_j)\} \right] k\{y_{ij} \mid \theta_{s_{ij}}(t_j)\}.$$

Setting $\mathsf{N} := \max\{\lfloor \psi^*(u_{1,t_1}) \rfloor, \ldots, \lfloor \psi^*(u_{p,t_1}) \rfloor, \ldots, \lfloor \psi^*(u_{1,t_m}) \rfloor, \lfloor \psi^*(u_{p,t_m}) \rfloor\}$ (where $\lfloor p \rfloor$ denotes the largest integer not greater than p) we have that

$$\begin{aligned} \mathcal{L}\left\{v_\infty^{(m)}, \theta_\infty^{(m)} \mid \mathsf{Y}, \mathsf{U}, \mathsf{S}\right\} &\approx \mathcal{L}\left\{v_{\mathsf{N}}^{(m)}, \theta_{\mathsf{N}}^{(m)} \mid \mathsf{Y}, \mathsf{U}, \mathsf{S}\right\} \\ &\propto \mathcal{L}\left\{\mathsf{Y}, \mathsf{U}, \mathsf{S} \mid v_{\mathsf{N}}^{(m)}, \theta_{\mathsf{N}}^{(m)}\right\} \mathcal{L}\{v_{\mathsf{N}}^{(m)}) \, \mathcal{L}(\theta_{\mathsf{N}}^{(m)}\}, \end{aligned}$$

where

$$\mathcal{L}\{v_{\mathsf{N}}^{(m)}\} = \prod_{l=1}^{\mathsf{N}} \mathcal{L}\{v_l^{(m)}\} = \prod_{l=1}^{\mathsf{N}} \pi_v\{v_l(h_1)\} \prod_{j=2}^{m} T_{h_j}^l\{v_l(t_{j-1}), v_l(t_j)\}$$

and

$$\mathcal{L}(\theta_{\mathsf{N}}^{(m)}\} = \prod_{l=1}^{\mathsf{N}} \mathcal{L}\{\theta_l^{(m)}\} = \prod_{l=1}^{\mathsf{N}} G_0\{\theta_l(h_1)\} \prod_{j=2}^{m} S_{h_j}^l\{\theta_l(t_{j-1}), \theta_l(t_j)\},$$

denoting with $h_1 = t_1$ and $h_j = t_j - t_{j-1}$ for $j = 2, ..., m$.

Next, we need to worry about the updating of the particles processes, the weights processes, the slice and membership variables, and the jump parameters. In theory, this should not be hard since the particles and weights processes are Markovian with known transitions. For the slice variable it follows that

$$\mathcal{L}(u_{ij} \mid \cdots) = \mathsf{U}(0, \psi_{s_{ij}}),$$

and for the membership variable

$$\mathcal{L}(s_{ij} = k \mid \cdots) \propto \frac{1}{\psi_k} \mathsf{w}_k(t_j) \, k\{y_{ij} \mid \theta_k(t_j)\} \, \mathbb{1}_{\left(s_{ij} \in \left\{k : \psi_{s_{ij}} > u_{ij}\right\}\right)}.$$

Notice that since $\left\{k : \psi_{s_{ij}} > u_{ij}\right\}$ is a finite set, this latter distribution is easy to sample from $s_{ij} = 1, \ldots, \lfloor \psi^*(u_{ij}) \rfloor$.

For the updating of the jump parameters we have the methods presented in Section 3.5. However, as we saw in that section, the estimation of the dependence parameters is extremly sensitive, certain properties of smoothness of the trajectories must be kept under the approximation for the estimation to be good. Even though for some particular parameter values we obtain reasonable results, the simulation-based tests we are using among the work are exhaustive, based on thousands of simulated parameters and, overall, the previous algorithm still fails to give us satisfactory-enough results regarding the jump parameters. Therefore, a revision of the algorithm is still under development.

6 Poisson driven Markov processes

This chapter is a continuation of the model initiated by Nava (2013), and part of it is presented in Anzarut et al. (2017). Here, we consider another stationary Markov process, focusing again on transition mechanisms that retain a particular distribution of interest invariant over time. The process is constructed exploiting a symmetry induced by a Poisson-type transform. Such a construction works for prescribed arbitrary invariant distributions supported on $\mathbb{R}_+$ but it may be extended, by means of simple transformations, to processes with invariant distributions supported on $\mathbb{R}$ or other state-spaces, while preserving the appealing transition probability tractability.

In the discrete-time case, attention is focused on the class of GIG distributions. In the continuous-time setup, we use the Gamma distribution as basic building block and obtain, via suitable transformations, a richer class of diffusion processes with known transition density. These include, for instance, diffusions with GEV invariant distributions which, to our knowledge, have not been derived before.

In order to perform Bayesian estimation for such processes, we derive a Gibbs sampling algorithm, based on slice sampler techniques. Later on, we implement the algorithm in a simulation study, and in the analysis of three financial datasets.

6.1 The construction in discrete time

Let $\hat{f}(\cdot|y,\phi)$ denote the Poisson weighted distribution with parameters $\phi > 0, y \in \{0,1,2,...\}$, previously defined in Section 2.2. The time-homogeneous one-step ahead Markovian tran-

sition

$$P(x_{n-1},x_n) = \sum_{y=0}^{\infty} \hat{f}(x_n|y,\phi)\,\mathsf{Po}(y|x_{n-1}\phi)$$
$$= \exp\{-\phi(x_n+x_{n-1})\} f(x_n) \sum_{y=0}^{\infty} \frac{(x_n x_{n-1}\phi)^y}{y!\xi(y,\phi)}, \tag{6.1}$$

satisfies the *detailed balance condition*

$$P(x_{n-1},x_n)f(x_{n-1}) = P(x_n,x_{n-1})f(x_n)$$

for all $x_{n-1}, x_n \in \mathbb{R}_+$. Therefore, it leads automatically to a strongly stationary f-reversible Markov process, where f is the invariant distribution.

Definition 18. *The stationary Markov process, driven by transition density* (6.1) *and with stationary density f, is termed* **f*-stationary Poisson-driven Markov process***.

Constructions of strongly stationary, reversible Markov processes of this type where introduced in continuous time by Mena and Walker (2009). The "latent" representation of the transition density, as given in (6.1), provides with an instrumental way of dealing with the law of the process useful for the implementation of efficient estimation procedures.

The combination of the moments equation (2.6) and equation (6.1), leads to the following conditional moments for the f-stationary Poisson-driven Markov process.

$$\mathbb{E}[X_n^r \mid X_{n-1} = x_{n-1}] = \sum_{y=0}^{\infty} \frac{\xi(y+r,\phi)}{\xi(y,\phi)} \mathsf{Po}(y|x_{n-1}\phi).$$

Consequently, provided that f admits second moment, the autocorrelation can be expressed as

$$\mathsf{Corr}(X_n, X_{n-1}) = \frac{1}{\sigma_f^2}\left[\sum_{y=0}^{\infty}\left\{\frac{\xi(y+1,\phi)^2}{\xi(y,\phi)}\right\}\frac{\phi^y}{y!} - \mu_f^2\right], \tag{6.2}$$

where μ_f and σ_f^2 denote the mean and variance of f. For example if f is chosen to be a $\mathsf{Ga}(a,b)$ distribution, correlation (6.2) reduces to $\dfrac{\phi}{b+\phi}$.

Once the form of f is chosen, the dependence in the model is driven by the parameter ϕ. See, for example, Figure 6.1. In particular, when ϕ goes to infinity, the correlation

tends to one. As a consequence, the function f can be selected by the nature of the phenomenon under study, while the dependence in the data is modeled trough ϕ.

Next, we focus on densities f belonging to the family of GIG distributions. We denote with $A(\lambda,\delta,\gamma) = (\delta/\gamma)^{\lambda}\, 2K_{\lambda}(\delta\gamma)$ to the normalizing constant of a $\mathsf{GIG}(x|\lambda,\delta,\gamma)$ distribution. Developing further the construction, we obtain

$$\begin{aligned}\xi(y,\phi) &= \int_{\mathbb{R}_+} z^y e^{-z\phi}\ \mathsf{GIG}\,(z|\lambda,\delta,\gamma)\,dz \\ &= \frac{\delta^y}{(\gamma^2+2\phi)^{y/2}}\,\frac{\gamma^{\lambda}}{(\gamma^2+2\phi)^{\lambda/2}}\,\frac{K_{y+\lambda}\left(\delta\sqrt{\gamma^2+2\phi}\right)}{K_{\lambda}(\delta\gamma)} \\ &= \frac{A(\lambda+y,\delta,\sqrt{\gamma^2+2\phi})}{A(\lambda,\delta,\gamma)}.\end{aligned}$$

It is then straightforward to see that,

$$\hat{f}(x|y,\phi) = \mathsf{GIG}(x|\lambda+y,\delta,\sqrt{\gamma^2+2\phi}).$$

Hence, the Poisson weighted distribution generated by a GIG density is also GIG, meaning that the GIG family is closed under Poisson weighted transformations, an appealing feature for simulation and estimation purposes. Given this, the corresponding transition density (6.1) is

$$\begin{aligned}P(x_{n-1},x_n) &= \sum_{y=0}^{\infty} \mathsf{GIG}\left(x_n|\lambda+y,\delta,\sqrt{\gamma^2+2\phi}\right)\mathsf{Po}(y|x_{n-1}\phi) \\ &= x_n^{\lambda-1}\exp\left\{-\phi(x_n+x_{n-1}) - \frac{1}{2}\left(\frac{\delta^2}{x_n}+\gamma^2 x_n\right)\right\}\sum_{y=0}^{\infty}\frac{(x_{n-1}x_n\,\phi)^y}{y!\ A(\lambda+y,\delta,\sqrt{\gamma^2+2\phi})}.\end{aligned}$$

Some particular cases offering further simplifications are at hand. For example, for $\lambda = -\frac{1}{2}$, we obtain the inverse Gaussian distribution, meaning that $\mathsf{IG}(\delta,\gamma) = \mathsf{GIG}(-1/2,\delta,\gamma)$. So for the IG–stationary Poisson-driven Markov process

$$\begin{aligned}\xi(y,\phi) &= \frac{A(y-\frac{1}{2},\delta,\sqrt{\gamma^2+2\phi})}{A(-\frac{1}{2},\delta,\gamma)} \\ &= \sqrt{\frac{2}{\pi}}\,e^{\delta\gamma}\,\delta^{y+\frac{1}{2}}\left(\sqrt{\gamma^2+2\phi}\right)^{\frac{1}{2}-y} K_{y-\frac{1}{2}}\left(\delta\sqrt{2\phi+\gamma^2}\right),\end{aligned}$$

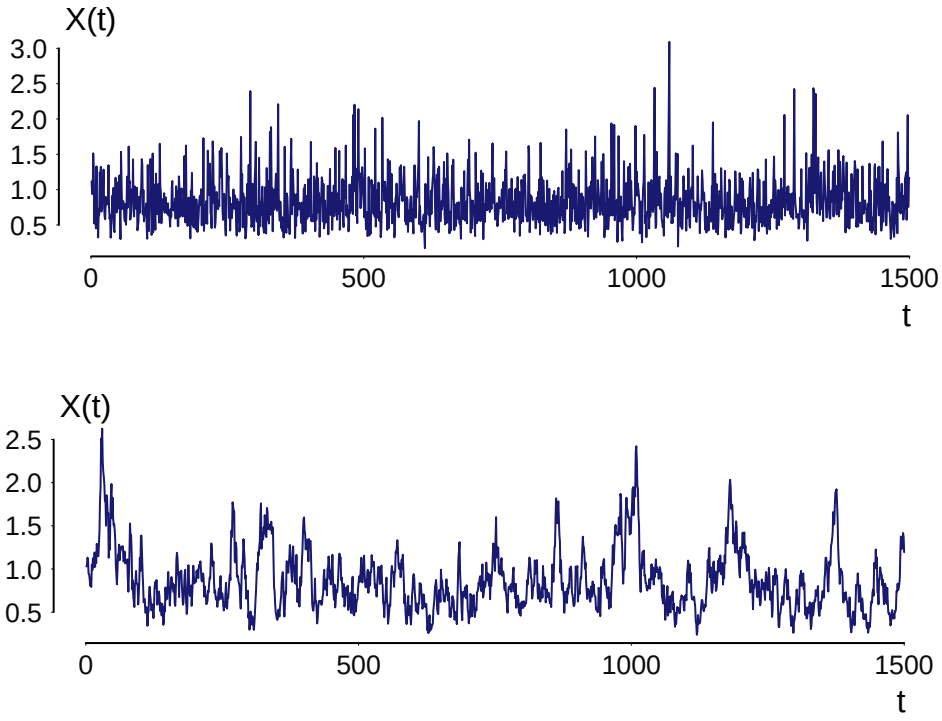


Figure 6.1: Simulation of $\mathsf{GIG}(1, 2, 3)$-stationary Poisson-driven Markov processes. The top panel displays 1500 simulated data with $\phi = 1$. The bottom repeats the simulation for $\phi = 80$.

and the corresponding transition density is

$$P(x_{n-1},x_n) = \exp\left\{-\phi(x_{n-1}+x_n) - \frac{1}{2}\left(\delta^2 x_n^{-1} + \gamma^2 x_n\right)\right\} x_n^{-\frac{3}{2}} \left(\frac{\delta}{\sqrt{\gamma^2+2\phi}}\right)^{\frac{1}{2}} \times \sum_{y=0}^{\infty} \frac{\left(\phi\ x_{n-1}\ x_n\sqrt{\gamma^2+2\phi}\ \delta^{-1}\right)^y}{y!\ 2K_{y-\frac{1}{2}}(\delta\sqrt{\gamma^2+2\phi})}.$$

For $\delta = 0$, the Gamma distribution is recovered, i.e. $\mathsf{Ga}(\lambda,\beta) = \mathsf{GIG}(\lambda,0,\gamma)$ where $\beta = \gamma^2/2$ resulting in

$$\xi^{\mathsf{Ga}}(y,\phi) = \frac{\beta^\lambda}{\Gamma(\lambda)} \frac{\Gamma(\lambda+y)}{(\beta+\phi)^{y+\lambda}}, \tag{6.3}$$

with corresponding transition density

$$P(x_{n-1},x_n) = \frac{\exp\{-[\phi(x_n+x_{n-1})+\beta x_n]\}}{(\phi+\beta)^{-(\lambda+1)/2}\,\phi^{(\lambda-1)/2}} \left(\sqrt{\frac{x_n}{x_{n-1}}}\right)^{\lambda-1} K_{\lambda-1}\left(2\sqrt{x_{n-1}x_n\phi(\phi+\beta)}\right),$$

where K_ν is the modified Bessel function of the third kind with index ν.

Also, the IGa-stationary Poisson-driven Markov process can be recovered noticing that for $\gamma = 0$, a GIG random variable reduces to the inverse distribution, i.e. $\mathsf{GIG}(-\lambda,\delta,0) = \mathsf{IGa}(a,b)$ with $a = -\lambda, b = \frac{\delta^2}{2}$. This leads to

$$\xi(y,\phi) = 2\,\frac{b^{\frac{1}{2}(a+y)}\,\phi^{\frac{1}{2}(a-y)}}{\Gamma(a)}\,K_{y-a}(2\sqrt{b\phi}),$$

and

$$P(x_{n-1},x_n) = x_n^{-a-1}\exp\left\{-\phi(x_n+x_{n-1}) - \frac{b}{x_n}\right\}\left(\frac{b}{\phi}\right)^{\frac{a}{2}} \sum_{y=0}^{\infty} \frac{\left(\phi\ x_{n-1}\ x_n\ \sqrt{\frac{\phi}{b}}\right)^y}{y!2\,K_{y-a}(2\sqrt{b\phi})}.$$

The particular case of $a = 1/2$ corresponds to the positive $1/2$ – stable distribution.

6.2 Extension to continuous time

As we did in the definition of SF-Harris processes in Section 3.1, we can extend a chain with transition functions (6.1) to a continuous-time Markov process $X = (X_t)_{t\geq 0}$ by making the parameter ϕ a function of time, $t \to \phi(t)$, and then verifying which functions $\phi(t)$ meet the Chapman-Kolmogorov equation.

In terms of Laplace transforms the Chapman-Kolmogorov equation is given by

$$\mathcal{L}_{X_{t+s}|X_0}(\lambda) = \mathbb{E}\Big[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \,\Big|\, X_0\Big], \tag{6.4}$$

where

$$\mathcal{L}_{X_t|X_0}(\lambda) = \mathbb{E}\Big[e^{-\lambda X_t} \,\Big|\, X_0\Big] = \sum_{y=0}^{\infty} \mathsf{Po}\{y|x_0\phi(t)\}\frac{\xi\{y,\phi(t)+\lambda\}}{\xi\{y,\phi(t)\}}.$$

Therefore, provided $\phi(t)$ satisfies (6.4), the transition functions

$$P_t(x_0,x_t) = \exp\{-\phi(t)(x_0+x_t)\}\, f(x_t) \sum_{y=0}^{\infty} \frac{\{x_t x_0 \phi(t)\}^y}{y!\xi\{y,\phi(t)\}} \tag{6.5}$$

define a continuous-time Markov process. Once more, the extension preserves entirely the chain virtues. The resulting process is strongly stationary with invariant distribution f.

Let us focus on a particular case of the GIG-stationary Poisson-driven Markov process, the case of the Gamma invariant distribution. Such an invariant distribution can be utilized for the construction of other continuous models and, its appealing feature, is the availability of an explicit functional form for $\phi(t)$ assuring the Chapman-Kolmogorov equation is fulfilled.

Mena and Walker (2009) showed that, in the $\mathsf{Ga}(a,b)$ case, the Chapman-Kolmogorov equation is fulfilled if and only if

$$\phi(t) = \frac{b}{e^{ct}-1}, \tag{6.6}$$

for some $c > 0$. Even more, they proved the process satisfies

$$\lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}[\,|\,X_{t+h} - X_t|^p \,|\, X_t = x_t] = 0 \quad \text{for} \quad p > 2. \tag{6.7}$$

Hence, X cannot have jump discontinuities, and it is indeed a diffusion process characterized by its infinitesimal conditional mean and variance coefficients given by

$$\mu(x,t) = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}[X_{t+h} - X_t \mid X_0 = x] \quad \text{and} \quad \sigma^2(x,t) = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}[(X_{t+h} - X_t)^2 \mid X_0 = x].$$

Computing the limits, X can be represented as the solution of the stochastic differential equation

$$dX_t = c\left(\frac{a}{b} - X_t\right) dt + \sqrt{\frac{2\,c}{b}\,X_t}\, dW_t, \tag{6.8}$$

where $W = (W_t)_{t\geq 0}$ denotes a standard Brownian motion. Therefore, X is in fact a reparametrization of the mean reverting Cox-Ingersoll-Ross model (Cox et al., 1985), commonly used to model nominal interest rates.

Based on this construction, we can build up different continuous-time Markov processes by the means of simple transformations. In particular, let us assume that we want to construct an f-stationary Poisson-driven Markov process, where f is the density corresponding to a random variable $Y = h(Z)$ with $Z \sim \mathsf{Ga}(z|a,b)$, and h is a $\mathbb{R}$-valued function with known and differentiable inverse (notice that this implies that the model can be extended to cover marginals f with support $\mathbb{R}$).

Letting $g(x) := h^{-1}(x)$ and $\mathcal{J}(x) := |g'(x)|$, we have that $f(x) = \mathsf{Ga}\{g(x)|a,b\}\,\mathcal{J}(x)$, and using equation (2.5) it follows directly that

$$\hat{f}(x|y,\phi) = \frac{g(x)^y e^{-g(x)\phi}\mathsf{Ga}\{g(x)|a,b\}\mathcal{J}(x)}{\xi(y,\phi,g)},$$

where

$$\begin{aligned}
\xi(y,\phi,g) &= \int_{\mathbb{R}_+} g(w)^y e^{-g(w)\phi}\mathsf{Ga}\{g(x)|a,b\}\mathcal{J}(x)dw \\
&= \int_{\mathbb{R}_+} z^y e^{-z\phi}\mathsf{Ga}(z|a,b)dw \\
&= \xi^{\mathsf{Ga}}(y,\phi)
\end{aligned}$$

Therefore,

$$\hat{f}(x|y,\phi) = \hat{f}^{\mathsf{Ga}}\{g(x)|y,\phi\}\,\mathcal{J}(x) = \mathsf{Ga}\{g(x)|a+y,b+\phi\}\,\mathcal{J}(x). \tag{6.9}$$

The corresponding transition probability is given by

$$\begin{aligned} P_t(x_0, x_t) &= \sum_{y=0}^{\infty} \hat{f}\{x_t|y, \phi(t)\}\, \mathsf{Po}\{y|g(x_0)\phi(t)\} \\ &= \sum_{y=0}^{\infty} \hat{f}^{\mathsf{Ga}}\{g(x)|y, \phi\}\, \mathcal{J}(x)\, \mathsf{Po}\{y|g(x_0)\phi(t)\} \\ &= P_t^{\mathsf{Ga}}\{g(x_0), g(x_t)\}\mathcal{J}(x_t). \end{aligned} \tag{6.10}$$

where P_t^{Ga} denotes the transition of a Ga-stationary Poisson-driven Markov process $Z = (Z_t)_{t\geq 0}$. Therefore, the Gamma case leads to a large class of tractable continuous stationary Poisson-driven Markov processes $X = \{X_t = h(Z_t)\}_{t\geq 0}$.

When the transformation h is twice differentiable, applying Itô's lemma to (6.8) we can find the associated stochastic differential equation. By setting $h'(Z_t) = \partial h(Z_t)/\partial Z_t$ and $h''(Z_t) = \partial^2 h(Z_t)/\partial^2 Z_t$ it follows that

$$\begin{aligned} dX_t &= dh(Z_t) \\ &= h'(Z_t)\, dZ_t + \frac{1}{2}\, h''(Z_t)(dZ_t)^2 \\ &= h'(Z_t)\left\{ c\,(a/b - Z_t)dt + \sqrt{\frac{2c}{b} Z_t}\, dW_t \right\} + \frac{1}{2} h''(Z_t)\left\{ c(a/b - Z_t)\, dt + \sqrt{\frac{2c}{b} Z_t}\, dW_t \right\}^2 \\ &= c\left\{ h'(Z_t)(a/b - Z_t) + h''(Z_t)\frac{Z_t}{b} \right\} dt + h'(Z_t)\sqrt{\frac{2\,c}{b}\, Z_t}\, dW_t \\ &= c\left[h'\left\{g(X_t)\right\}\left\{a/b - g(X_t)\right\} + h''\left\{g(X_t)\right\}\, \frac{g(X_t)}{b} \right] dt + h'\left\{g(X_t)\right\}\, \sqrt{\frac{2\,c}{b}\, g(X_t)}\, dW_t. \end{aligned} \tag{6.11}$$

The key aspect to remark, which represents a highly attractive feature in terms of practical implementation, is that we still have the explicit representation (6.10) for the transition density, with the same function $\phi(t)$ in equation (6.6). Hence, this derivation allows us to avoid the challenging task of having to solve the Chapman-Kolmogorov equation directly for each case.

As an example, let us illustrate the case of the inverse Gamma distribution. We saw in Section 2.2 that the inverse Gamma can be obtained from the Gamma using the transformation $h(z) = 1/z$. Hence, $g(x) = 1/x$, $\mathcal{J}(x) = 1/x^2$, and substituting in (6.9) we

get that

$$\hat{f}(x|y,\phi) = \mathsf{IGa}(x|y+a,\phi+b)$$

Since h is twice differentiable, $h'(x) = -1/x^2$ and $h''(x) = 2/x^3$, the associated stochastic differential equation (6.11) is given by

$$\begin{aligned} dX_t &= c\left[-X_t^2\left\{a/b - X_t^{-1}\right\} + 2X_t^3\ \frac{X_t^{-1}}{b}\right]dt - X_t^2\sqrt{\frac{2\,c}{b}\,X_t^{-1}}\,dW_t \\ &= c\,X_t\left(1 - X_t\ \frac{a-2}{b}\right)dt - \sqrt{\frac{2c}{b}X_t^3}\,dW_t. \end{aligned}$$

The same can be done for other transformations of the Gamma, for example, for the GEV family and, consequently, for the Gumbel, Fréchet, and Weibull distributions, which are particular cases of the GEV. These are displayed in Table VI, together with the resulting conditional distributions. The stochastic differential equations corresponding to each of the resulting f-stationary Poisson driven processes are reported in Table VII.

Table VI: Transform function h for the marginal distribution f, and corresponding Poisson weighted density $\hat{f}$.

$\boldsymbol{h(z)}$	$\boldsymbol{f(x)}$	$\boldsymbol{\hat{f}(x\|y,\phi)}$
$\frac{\sigma}{\nu}\left(z^{-\nu}-1\right)+\mu$	$\mathsf{GEV}(x\|\mu,\sigma,\nu)$	$\mathsf{GEV}(x\|\mu,\sigma,\nu)\,(1+\phi)^{1+y}\left[1+\nu\frac{(x-\mu)}{\sigma}\right]^{-y/\nu}$ $\times\exp\left\{-\phi\left[1+\nu\frac{(x-\mu)}{\sigma}\right]^{-1/\nu}\right\}\frac{1}{y!}\,\mathbb{I}_{(\nu\mu-\sigma,\infty)}(x\nu)$
$-\sigma\log(z)+\mu$	$\mathsf{Gum}(x\|\mu,\sigma)$	$\mathsf{Gum}(x\|\mu,\sigma)\exp\left\{-y\frac{x-\mu}{\sigma}-\phi e^{-\frac{x-\mu}{\sigma}}\right\}(1+\phi)^{1+y}\,\frac{1}{y!}\,\mathbb{I}_{(-\infty,\infty)}(x)$
$\sigma\, z^{-1/\alpha}$	$\mathsf{Fre}(x\|\sigma,\alpha)$	$\mathsf{Fre}(x\|\sigma,\alpha)\,\frac{(1+\phi)^{1+y}}{\Gamma(1+y)}\left(\frac{x}{\sigma}\right)^{-\alpha\, y}\exp\left\{-(1+\phi)\left(\frac{x}{\sigma}\right)^{-\alpha}\right\}\mathbb{I}_{(0,\infty)}(x)$
$\sigma z^{1/\alpha}$	$\mathsf{Wei}(x\|\sigma,\alpha)$	$\mathsf{Wei}(x\|\sigma,\alpha)\,(\phi+1)^{y+1}\exp\left\{-(x/\sigma)^{\alpha}\phi\right\}(x/\sigma)^{\alpha y}\,\frac{1}{y!}\,\mathbb{I}_{[0,\infty)}(x)$

Table VII: SDEs related to diffusions with the invariant distributions of Table VI.

$\boldsymbol{f(x)}$	**SDE**
$\mathsf{GEV}(x\|\mu,\sigma,\nu)$ $\nu \neq 0$	$dX_t = c\sigma\left[(\nu R(X_t)+1)+\nu\left(\nu R(X_t)+1\right)^{(\nu+1)/\nu}\right]dt + -\sigma\sqrt{2c\left(\nu R(X_t)+1\right)^{(2\nu+1)/\nu}}\,dW_t$ with $R(X_t) := (X_t-\mu)/\sigma$
$\mathsf{Gumbel}(x\|\mu,\sigma)$	$dX_t = c\sigma\,dt - \sigma\sqrt{2c\exp\left\{\frac{X_t-\mu}{\sigma}\right\}}\,dW_t$
$\mathsf{Fre}(x\|\sigma,\alpha)$	$dX_t = c\left(\frac{X_t}{\sigma}\right)^2\left[\frac{\sigma}{\alpha}+\frac{2}{\alpha^2}\left(\frac{X_t}{\sigma}\right)^{\alpha}\right]dt - \frac{\sigma\sqrt{2c}}{\alpha}\left(\frac{X_t}{\sigma}\right)^{2+\frac{\alpha}{2}}\,dW_t$
$\mathsf{Wei}(x\|\sigma,\alpha)$	$dX_t = c\alpha^{-2}X_t\left[\left(\frac{X_t}{\sigma}\right)^{-\alpha}-\alpha\right]dt + \alpha^{-1}\sqrt{2c\sigma^{\alpha}X_t^{2-\alpha}}\,dW_t$

6.3 Inference

6.3.1 Estimation method

In this section we focus on estimation in the continuous-time case, since the discrete-time case can be easily recovered from it. We let $\mathbf{x} = (x_1, \ldots, x_N)$ be a set of observations in the times $t_1 < ... < t_N$ (with $x_n = x_{t_n}$), and ϕ_{τ_n} be the homogeneous time effect ($\phi_{\tau_n} = \phi(t_n) - \phi(t_{n-1})$), with $\tau_n = t_n - t_{n-1}$.

The likelihood of $\mathbf{x}$ is given by

$$\mathcal{L}_{\mathbf{x}}(\boldsymbol{\theta}) = f(x_1|\boldsymbol{\theta}) \prod_{n=2}^{N} P(x_{n-1}, x_n|\boldsymbol{\theta}),$$

where P is as in equation (6.5), and $\boldsymbol{\theta}$ denotes, generically, the set of parameters of f and the ones inherent to the dependency function $\phi(t)$.

The choice of f may not allow to perform the summation in (6.5) analytically. Hence, we propose a Gibbs sampler using slice-technique ideas. The Gibbs sampler is based on an augmented representation of transition density (6.5), given by

$$P_t(x_0, x_t, u, y) = \frac{1}{\psi_y} \mathbb{1}_{(u<\psi_y)} \exp\{-\phi(t)(x_t + x_0)\} f(x_t|\boldsymbol{\theta}) \frac{\{x_t\, x_0\, \phi(t)\}^y}{y!\, \xi\{y, \phi(t)|\boldsymbol{\theta}\}},$$

where $y \mapsto \psi_y$ is a positive decreasing invertible function, and the latent variable u is uniformly distributed given y. With this, if $\mathbf{y} = (y_2, \ldots, y_N)$ and $\mathbf{u} = (u_2, \ldots, u_N)$,

$$\begin{aligned}\mathcal{L}_{\mathbf{x},\mathbf{u},\mathbf{y}}(\boldsymbol{\theta}) = \exp\left\{-\sum_{n=2}^{N} \phi_{\tau_n}(x_n + x_{n-1})\right\} \left\{\prod_{n=1}^{N} f(x_n|\boldsymbol{\theta})\right\} \\ \times \left\{\prod_{n=2}^{N} \frac{(x_n x_{n-1} \phi_{\tau_n})^{y_n}}{y_n!\, \xi(y_n, \phi_{\tau_n}|\boldsymbol{\theta}) \psi_{y_n}} \mathbb{1}_{(u_n<\psi_{y_n})}\right\},\end{aligned}$$

leading to a log-likelihood

$$l_{\mathbf{x},\mathbf{u},\mathbf{y}}(\boldsymbol{\theta}) = -\sum_{n=2}^{N} \phi_{\tau_n}(x_n + x_{n-1}) + \sum_{n=1}^{N} \log\{f(x_n|\boldsymbol{\theta})\} + \sum_{n=2}^{N} \left[\log\{\mathbb{1}_{(u_n<\psi_{y_n})}\} - \log(\psi_{y_n})\right] + \sum_{n=2}^{N} \left[y_n \log(x_n x_{n-1}\phi_{\tau_n}) - \log\{y_n!\,\xi(y_n, \phi_{\tau_n}|\boldsymbol{\theta})\}\right], \tag{6.12}$$

Now, denoting with π the prior distribution on $\boldsymbol{\theta}$, the corresponding full log-posterior distribution can be simplified by separating the parameters in the stationary distribution, $\boldsymbol{\theta}^{(st)}$, and the parameters referring to the transition probability $\boldsymbol{\theta}^{(tr)}$. Under the assumption of independent prior distributions for $\boldsymbol{\theta}^{(st)}$ and $\boldsymbol{\theta}^{(tr)}$, the full log-posterior distributions reduce to

$$\log \pi\{\boldsymbol{\theta}^{(st)} \mid \cdots\} \propto \log \pi\{\boldsymbol{\theta}^{(st)}\} + \sum_{n=1}^{N} \log[f\{x_n|\boldsymbol{\theta}^{(st)}\}] - \sum_{n=2}^{N} \log\{\xi(y_n, \phi_{\tau_n}|\boldsymbol{\theta})\},$$

and

$$\log \pi\{\boldsymbol{\theta}^{(tr)} \mid \cdots\} \propto \log \pi\{\boldsymbol{\theta}^{(tr)}\} + \sum_{n=2}^{N} y_n \log(\phi_{\tau_n}) - \sum_{n=2}^{N} \phi_{\tau_n}(x_n + x_{n-1}) - \sum_{n=2}^{N} \log\{\xi(y_n, \phi_{\tau_n}|\boldsymbol{\theta})\}.$$

Therefore, simulating from the full posteriors can be easily achieved, for instance, using the ARMS algorithm.

The full conditional distributions for the latent variables can be obtained componentwise via

$$\begin{aligned} \pi(u_n \mid \cdots) &= \mathsf{U}\{u_n|0, \psi(y_n)\}, \\ \pi(y_n \mid \cdots) &\propto \frac{(x_n\, x_{n-1}\, \phi_{\tau_n})^{y_n}}{y_n!\,\xi(y_n, \phi_{\tau_n}|\boldsymbol{\theta})\,\psi_{y_n}} \mathbb{1}_{(u_n<\psi_{y_n})}, \end{aligned} \tag{6.13}$$

for $n = 2, \ldots, N$. Note that the above distribution has support $y_n = 0, \ldots, \lfloor \psi^*(u_n) \rfloor$, where ψ^* denotes the inverse of ψ. This is the advantage of using the slice mechanism, we only need to sample from a finite support instead of a distribution supported on $\mathbb{N}$.

In particular, for the Gamma transformed models in continuous time, we may assign independent unitary exponential priors to $\boldsymbol{\theta}^{(st)} = (a, b, c)$. Hence, the corresponding log-

posteriors are given by

$$\log\pi(a\mid\cdots)\propto a\left\{\log(b)-1+\sum_{n=1}^{N}\log(b+\phi_{\tau_n})+\sum_{n=1}^{N}\log g(x_n)\right\}-\log\Gamma(a)-\sum_{n=2}^{N}\log\Gamma(y_n+a),$$
$$\log\pi(b\mid\cdots)\propto\log(b)\left(a+2\sum_{n=2}^{N}y_n+a\,N\right)-b\left\{\sum_{n=1}^{N}g(x_n)+1\right\}-\sum_{n=2}^{N}\left\{g(x_n)+g(x_{n-1})\right\}\phi_{\tau_n},$$
$$\log\pi(c\mid\cdots)\propto-\sum_{n=2}^{N}\left\{g(x_n)+g(x_{n-1})\right\}\phi_{\tau_n}+\log(\phi_{\tau_n})\left(2\sum_{n=2}^{N}y_n+a\,N\right)+c\left(\sum_{n=2}^{N}y_n+a\,N-1\right).$$

The log-posterior for $\boldsymbol{\theta}^{(tr)}$ is

$$\begin{aligned}\log\pi\{\boldsymbol{\theta}^{(tr)}\mid\cdots\}\propto\log\pi\{\boldsymbol{\theta}^{(tr)}\}+\sum_{n=1}^{N}\log\mathcal{J}\{x_n|\boldsymbol{\theta}^{(tr)}\}+(a-1)\sum_{n=1}^{N}\log g(x_n)-b\sum_{n=1}^{N}g(x_n)\\-\sum_{n=2}^{N}\left\{g(x_n)+g(x_{n-1})\right\}\phi_{\tau_n}+\sum_{n=2}^{N}y_n\left\{\log g(x_n)+\log g(x_{n-1})\right\},\end{aligned}\tag{6.14}$$

which can be sampled with the ARMS method and, depending on the specific transformation, may admit further simplifications.

The remaining full conditionals to be considered in the Gibbs sampler are (6.13) and

$$\pi(y_n\mid\cdots)\propto\frac{\left\{g(x_{n-1})\,g(x_n)\,\phi_{\tau_n}\,(b+\phi_{\tau_n})\right\}^{y_n}}{y_n!\,\Gamma(a+y_n)\,\psi_{y_n}}\mathbb{1}_{\{y_n<\psi^{-1}(u_n)\}}.$$

6.3.2 Simulation study

Over this section we test the estimation procedure by simulating Poisson-driven Markov processes with GIG and GEV stationary distributions in discrete and continuous time, respectively. These choices are done given their relevance in areas such as Economics and Finance. See, e.g., Nakajima et al. (2012).

The GIG case consists in two series of 1500 simulated observations from a discrete-time model with $\mathsf{GIG}(1,2,3)$ invariant distribution. In the first series $\phi=1$, while in the second one $\phi=80$. The two series paths appear in Figure 6.1. We can notice that, as the dependence parameter ϕ gets larger, cluster structures appear in the data.

We assign independent unitary exponential priors to δ, γ, and ϕ, an independent standard normal prior for λ, and then derive the full conditional distributions. Additionally, we set the truncation function ψ_y equal to $e^{-\eta y}$ with $\eta = 0.4$, this choice allows to have a known inverse ψ^* and, therefore, to immediately identify the support of the latent variable y.

In this case, for a set of observations $\mathbf{x} = (x_1, \ldots, x_N)$ at times $(1, \ldots, N)$, the augmented log-likelihood (6.12) is given by

$$\begin{aligned} l_{\mathbf{x},\mathbf{u},\mathbf{y}}(\boldsymbol{\theta}) = & -\phi \sum_{n=2}^{N} (x_n + x_{n-1}) + (\lambda - 1) \sum_{n=1}^{N} \log(x_n) - \frac{1}{2} \sum_{n=1}^{N} (\delta^2 x_n^{-1} + \gamma^2 x_n) - \log\{A(\lambda, \delta, \gamma)\} \\ & + \sum_{n=2}^{N} \left[y_n \log(x_n x_{n-1} \phi_{\tau_n}) - \log\left\{ y_n!\, A(\lambda + y_n, \delta, \sqrt{\gamma^2 + 2\phi}) \right\} \right] \\ & + \sum_{n=2}^{N} \left[\log\{\mathbb{1}_{(u_n < \psi_{y_n})}\} - \log(\psi_{y_n}) \right], \end{aligned}$$

Hence, the corresponding full log-posterior distributions are

$$\begin{aligned}
\log \pi(\lambda \mid \cdots) \propto & -\frac{\lambda^2}{2} + \lambda \sum_{n=1}^{N} \log(x_n) + \lambda \left\{ -N \log(\delta) + \log(\gamma) + \frac{N-1}{2} \log(\gamma^2 + 2\phi) \right\} \\
& - \log\{K_\lambda(\delta\gamma)\} - \sum_{n=2}^{N} \log\left\{ K_{\lambda+y_n}\left(\delta\sqrt{\gamma^2+2\phi}\right)\right\}, \\
\log \pi(\delta \mid \cdots) \propto & -\delta - \frac{1}{2}\delta^2 \sum_{n=1}^{N} x_n^{-1} - \log\{K_\lambda(\delta\gamma)\} - \log(\delta) \left\{ \lambda N + \sum_{n=2}^{N} y_n \right\} \\
& - \sum_{n=2}^{N} \log\left\{ K_{\lambda+y_n}\left(\delta\sqrt{\gamma^2+2\phi}\right)\right\}, \\
\log \pi(\gamma \mid \cdots) \propto & -\gamma - \frac{1}{2}\gamma^2 \sum_{n=1}^{N} x_n - \log\{K_\lambda(\delta\gamma)\} + \lambda \log(\gamma) + \frac{1}{2} \log(\gamma^2 + 2\phi) \sum_{n=2}^{N} (\lambda + y_n) \\
& - \sum_{n=2}^{N} \log\left\{ K_{\lambda+y_n}\left(\delta\sqrt{\gamma^2+2\phi}\right)\right\}, \\
\log \pi(\phi \mid \cdots) \propto & -\phi + \log(\phi) \sum_{n=2}^{N} y_n - \phi \sum_{n=2}^{N} (x_n + x_{n-1}) + \frac{1}{2} \log(\gamma^2 + 2\phi) \sum_{n=2}^{N} (\lambda + y_n) \\
& - \sum_{n=2}^{N} \log\left\{ K_{\lambda+y_n}\left(\delta\sqrt{\gamma^2+2\phi}\right)\right\}.
\end{aligned}$$

Finally, for $n = 2, \ldots, N$,

$$\begin{aligned}
\log \pi(y_n \mid \cdots) \propto & \, y_n \log\left(x_n \, x_{n-1} \, \phi \, \frac{\sqrt{\gamma^2+2\phi}}{\delta} \right) - \log(y_n!) + \eta y_n + \log\{\mathbb{1}_{(u_n < e^{-\eta y})}\} \\
& - \log\left\{ K_{\lambda+y_n}\left(\delta\sqrt{\gamma^2+2\phi}\right)\right\}.
\end{aligned}$$

We perform the estimation method on the two data sets. The KL-divergence between the real invariant distribution and the estimated one is 0.0011 for the first data series, and 0.0019 for the second one (see Figure 6.2). For the ϕ parameter, the posterior distribution mode is given by 1.299 in the first series, and 66.236 in the second one. Hence, there is clear evidence of the good performance of the proposed model.

Again, an immediate procedure for obtaining a set of m trajectories in new times is available by, first, simulating m parameter values from the posterior distributions, and,

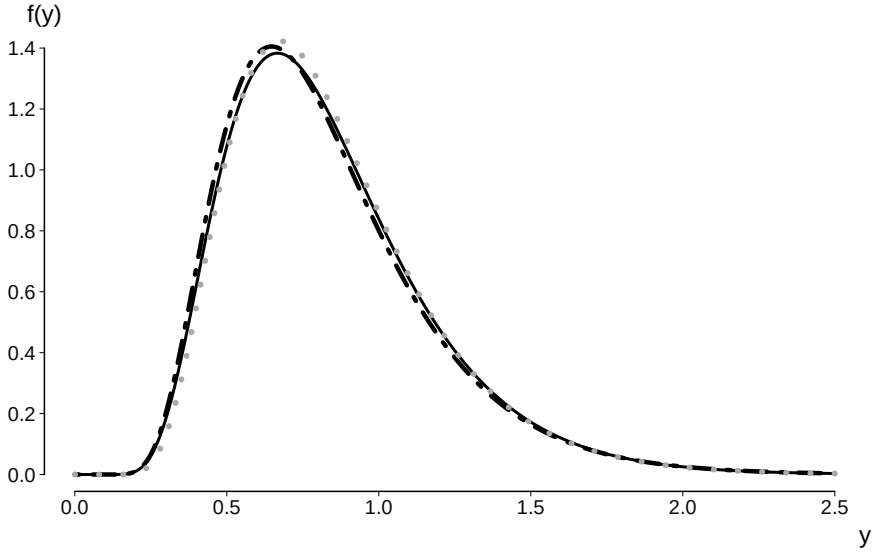


Figure 6.2: In black and solid appears the $\mathsf{GIG}(1,2,3)$ density. In black and dashed appears the estimated denisty for the $\mathsf{GIG}(1,2,3)$-stationary Poisson-driven Markov processes with $\phi = 1$. In gray and dotted appears the estimated denisty for the $\mathsf{GIG}(1,2,3)$-stationary Poisson-driven Markov processes with $\phi = 80$.

second, for each one of the m values, simulating a realization of the process starting on the last observation and parameterized by such a value. As an example, we computed highest posterior density intervals of probability 0.9 for both series. This appears in Figure 6.3.

The GEV illustration is in continuous time. We consider two simulated data sets with 1000 observations, the first one with a type I GEV (or Gumbel) distribution with density $f_1(x) = \mathsf{Gum}(x|1,4)$, and the second one with a type II GEV distribution with density $f_2(x) = \mathsf{GEV}(x|1,0.8,0.7)$. Additionally, we set the parameter $\boldsymbol{\theta}^{(tr)} = c = 1$ in both cases.

We assign independent unitary exponential priors to σ, ν, and c, and an independent standard normal to μ. For expositions' sake, we set $\tau_n = 1$ for $n = 1,2,\ldots,N$, $w_n = (x_n - \mu)/\sigma$ and $z_n = 1 + (x_n + \mu)/\sigma$. The symbols Gum and GEV will be used to distinguish the distributions corresponding to, respectively, type I and II GEV.

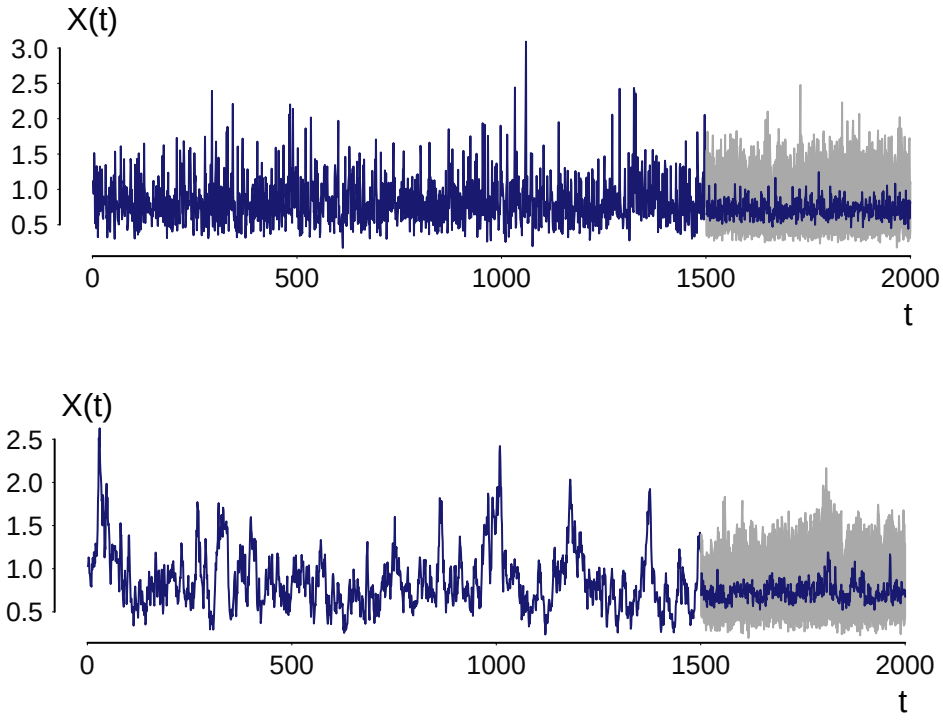


Figure 6.3: The first 1500 observations are simulations of the $\mathsf{GIG}(1,2,3)$-stationary Poisson-driven Markov processes for $\phi = 1$ (top panel), and $\phi = 80$ (bottom panel). The remaining 500 observations are the mode of the predictive distribution computed with 1000 simulated paths together with prediction intervals of 0.9 probability.

The corresponding full log-posterior distributions for the Gumbel are

$$\begin{aligned}
\log \pi^{\mathsf{Gum}}(c_1 \mid \cdots) \propto & -\frac{1}{e^{c_1}-1}\sum_{n=2}^{N}\left(e^{-w_n}+e^{-w_{n-1}}\right)-\log(e^{c_1}-1)\left(2\sum_{n=2}^{N}y_n+N\right) \\
& +c_1\left(\sum_{n=2}^{N}y_n+N-1\right), \\
\log \pi^{\mathsf{Gum}}(\mu \mid \cdots) \propto & -\frac{\mu^2}{2}-\sum_{n=2}^{N}\left\{(1+\phi_1)e^{-w_n}+\phi_1 e^{-w_{n-1}}\right\}+\frac{\mu}{\sigma}\left(N+2\sum_{n=2}^{N}y_n\right)-e^{-w_1}, \\
\log \pi^{\mathsf{Gum}}(\sigma \mid \cdots) \propto & -\sigma-\sum_{n=2}^{N}\left\{(1+\phi_1)e^{-w_n}+\phi_1 e^{-w_{n-1}}\right\}-\sum_{n=2}^{N}\left\{(1+y_n)w_n+y_n w_{n-1}\right\} \\
& -e^{-w_1}-w_1-N\log(\sigma).
\end{aligned}$$

The full log-posterior distributions for the type II GEV are

$$\begin{aligned}
\log \pi^{\mathsf{GEV}}(c_2 \mid \cdots) \propto & -\frac{1}{e^{c_2}-1}\sum_{n=2}^{N}\left(z_n^{-\frac{1}{\nu}}+z_{n-1}^{-\frac{1}{\nu}}\right)-\log(e^{c_2}-1)\left(2\sum_{n=2}^{N}y_n+N-1\right) \\
& +c_2\left(\sum_{n=2}^{N}y_n+N-1\right), \\
\log \pi^{\mathsf{GEV}}(\mu \mid \cdots) \propto & -\frac{\mu^2}{2}-\frac{\nu+1}{\nu}\log(z_1)-z_1^{-\frac{1}{\nu}}-\sum_{n=2}^{N}\left\{(1+\phi_1)\,z_n^{-\frac{1}{\nu}}+\phi_1\,z_{n-1}^{-\frac{1}{\nu}}\right\} \\
& -\frac{1}{\nu}\sum_{n=2}^{N}\left\{(y_n+\nu+1)\log(z_n)+y_n\log(z_{n-1})\right\}, \\
\log \pi^{\mathsf{GEV}}(\sigma \mid \cdots) \propto & -\sigma-\frac{\nu+1}{\nu}\log(z_1)-z_1^{-\frac{1}{\nu}}-\sum_{n=2}^{N}\left\{(1+\phi_1)\,z_n^{-\frac{1}{\nu}}+\phi_1\,z_{n-1}^{-\frac{1}{\nu}}\right\} \\
& -\frac{1}{\nu}\sum_{n=2}^{N}\left\{(y_n+\nu+1)\log(z_n)+y_n\log(z_{n-1})\right\}-N\log(\sigma), \\
\log \pi^{\mathsf{GEV}}(\nu \mid \cdots) \propto & -\nu-\frac{\nu+1}{\nu}\log(z_1)-z_1^{-\frac{1}{\nu}}-\sum_{n=2}^{N}\left\{(1+\phi_1)\,z_n^{-\frac{1}{\nu}}+\phi_1\,z_{n-1}^{-\frac{1}{\nu}}\right\} \\
& -\frac{1}{\nu}\sum_{n=2}^{N}\left\{(y_n+\nu+1)\log(z_n)+y_n\log(z_{n-1})\right\}.
\end{aligned}$$

which again can be easily sampled by the ARMS method.

Last, the following full conditional distributions complete the description of the Gibbs

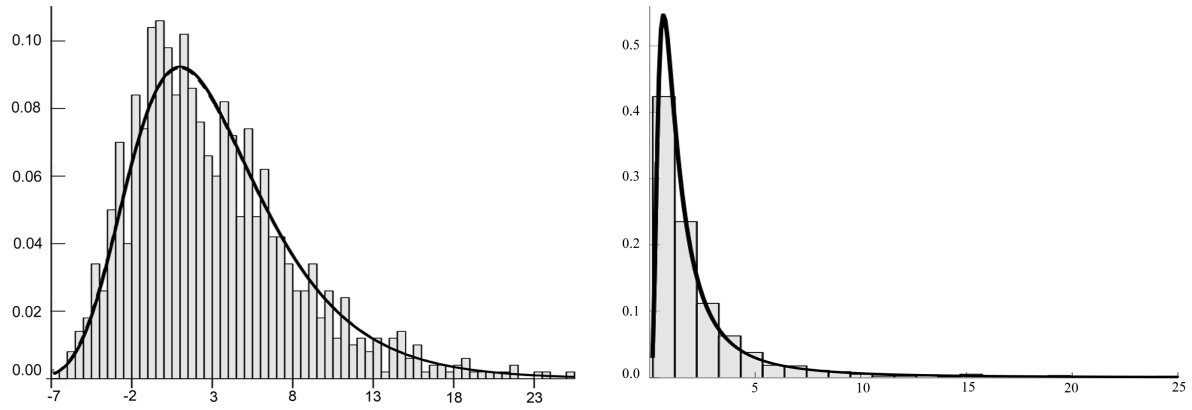


Figure 6.4: The histograms correspond to 1000 simulated data from type I (left) and type II (right) GEV-Poisson driven Markov processes, together with their corresponding estimated stationary distributions.

sampler

$$\pi^{\mathsf{Gum}}(y_n \mid \cdots) \propto \frac{\left[\exp\left\{-(w_{n-1}+w_n)\right\}\phi_1\left(1+\phi_1\right)\right]^{y_n}}{y_n!\,\Gamma(1+y_n),\psi_{y_n}} \mathbb{1}_{\{y_n<\psi^{-1}(u_n)\}},$$

$$\pi^{\mathsf{GEV}}(y_n \mid \cdots) \propto \frac{\left\{\left(z_{n-1}\,z_n\right)^{-\frac{1}{\nu}}\,\phi_1\left(1+\phi_1\right)\right\}^{y_n}}{y_n!\,\Gamma(1+y_n),\psi_{y_n}} \mathbb{1}_{\{y_n<\psi^{-1}(u_n)\}}.$$

The KL-divergence between the real invariant distribution and the estimated one is 0.00003 in the I GEV case, and 0.00037 in the II GEV case (see Figure 6.4). For the c parameter, the posterior distribution mode is given by 1.044 in the I GEV case, and 0.998 in the II GEV case. Therefore, the evidence suggests the estimation method works correctly for type I and II GEV distributions, and we can proceed by applying it to real data.

6.3.3 Empirical analysis

The f-stationary Poisson-driven Markov process can be an appealing alternative for the analysis of financial series. Indeed, a number of common stylized features typically observed in these type of data, such as heavy tail distributions and volatility clustering, can be well captured with the appropriate choice of stationary density f and through the

non-linear dependence driven by ϕ. We illustrate this by means of three datasets.

The first dataset consists of 937 daily estimations of the realized volatility of the FTSE 100 equity index, from October 31, 2003 to May 31, 2007. The estimations are provided in Heber et al. (2009). The open price, with its corresponding log-returns and realized volatility are displayed in Figure 6.5. Given the positive support, cluster pattern, and heavy tail behavior of the realized volatility, it seems plausible to adopt the discrete-time model with GIG stationary distribution described in Section 6.3.2.

The data is first cleaned, one outlier is replaced by a missing value and, with it, around four percent of the data is missing seemingly completely at random. The missing data are imputed using the predictive mean matching method provided in Buuren and Groothuis-Oudshoorn (2011) mice R package. After the cleaning, the estimation method is implemented on the realized volatility multiplied by 300. The posterior estimate for the stationary density is displayed in Figure 6.6. The posterior modes of the model parameters $(\lambda, \delta, \gamma, \phi)$ are equal to $(-2.837, 0.173, 0.316, 37.247)$.

To test the modeling, we break the sample into an estimation period (from October 31, 2003 to December 31, 2006), and a subsequent forecasting period (from January 1, 2007 to May 31, 2007). Next, we predict probability intervals for the forecasting period with 1000 simulated trajectories. The intervals appear in Figure 6.7. Indeed, 92 percent of the sample falls within the prediction intervals of probability 0.95. Therefore, we may conclude the method is working correctly for this data.

The other two data sets are the minimum daily stock returns occurring during a month of the S&P 500 and the Tokyo Stock Price Index (TOPIX). The S&P 500 is one of the most representative market indexes and rests upon the common stock prices of 500 top publicly traded American companies. TOPIX measures the market value changes of the common stocks in the Tokyo Stock Exchange. For our analysis the S&P 500 series has a coverage period of almost 12 years, from January 3, 2000 to July 9, 2012, whereas the TOPIX data are based on a 22 years period, from January 1, 1990 to July 31, 2012.

In both cases we compute daily returns taking log-differences multiplied by 100 and then compute the monthly minima. Hence, it seems plausible to adopt the models with extreme value distributions described in Section 6.3.2. Two series consisting of 151 and 271 observations are obtained for S&P 500 and TOPIX, respectively. The extracted series are displayed in Figure 6.8.

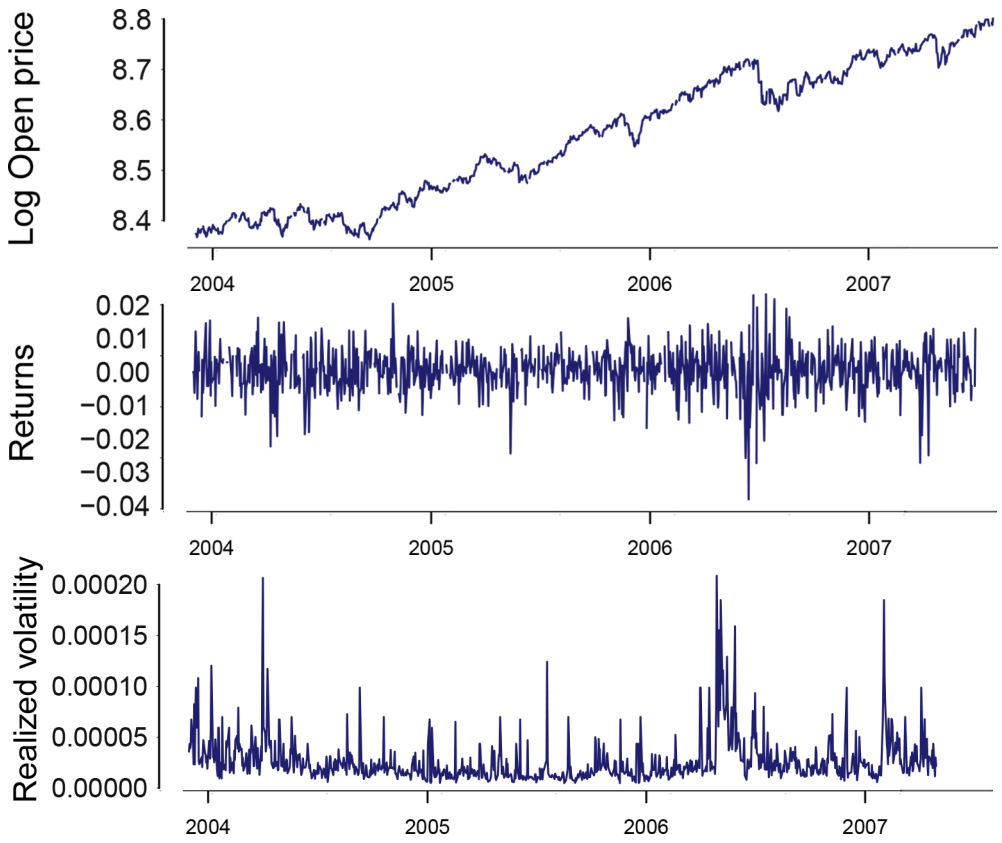


Figure 6.5: Log open prices, returns, and realized volatility series for FTSE 100 from October 31, 2003 to May 31, 2007

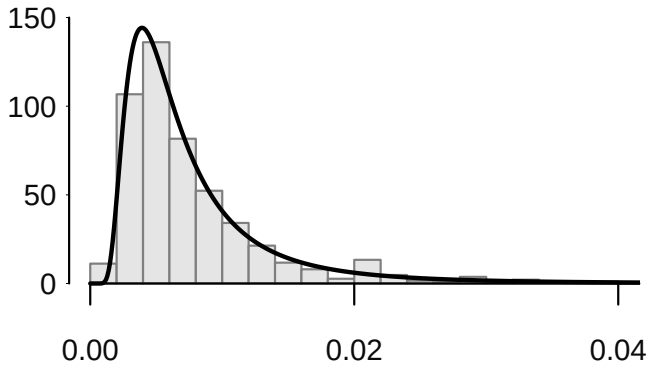


Figure 6.6: Histogram and estimate of the stationary density based on a GIG–Stationary Poisson driven model for FTSE 100 series.

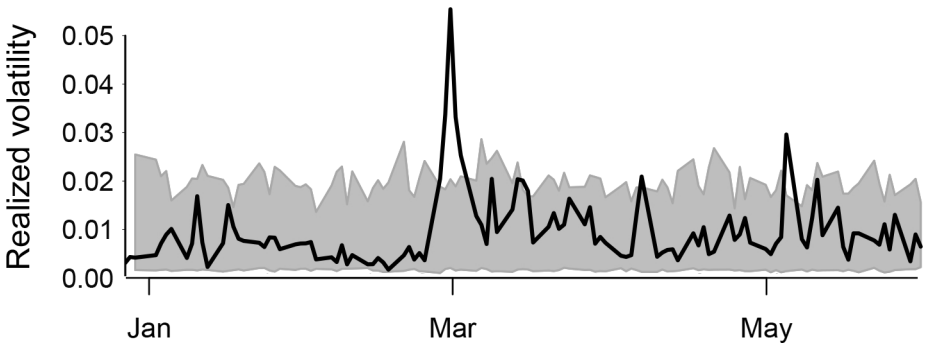


Figure 6.7: FTSE 100 realized volatility series from January 1, 2007 to May 31, 2007, along with highest posterior density intevals of probability 0.95. The intervals are computed using the estimation period from October 31, 2003 to December 31, 2006, with 1000 simulated trayectories.

We perform the estimation with both type I and type II GEV models. In the type I GEV case the posterior modes of (μ, σ, c) are given by (0.721,0.384,0.828) for S&P 500, and by (0.808, 0.412,0.725) for the TOPIX dataset. In the type II GEV case the posterior modes of (μ, σ, ν, c) are (0.618, 0.372, 0.041, 0.399) for the S&P 500 case, and (0.779, 0.385, 0.152, 0.702) for the TOPIX dataset. The posterior estimates of the stationary densities are displayed in Figure 6.9. We may conclude that the heavy tails, clearly observable in both datasets, can be satisfactorily captured by GEV-stationary Poisson-driven Markov processes.

6.4 Comparison to the SF-Harris process

In the following, we adjust the GIG-Harris SV model (see Definition 4.1) to the FTSE equity index. First, the cleaning and imputation of the data is conducted exactly as it was described in the previous section. Then, we proceed by estimating the periodicity function, appearing in Table VIII . Notice that we have daily observations, so the function only has one value per day. In fact, the values on weekdays are similar, indicating that there is no strong periodic component in these data. Next, we compute the periodically adjusted volatility, to which we adjusted the SF-Harris process using the method of Section 4.4.1. The posterior estimate for the stationary density is in Figure 6.11. The posterior modes of the model parameters $(\alpha, \lambda, \kappa, \eta)$ are equal to $(0.70, -3.14, 0.32, 0.68)$. In terms of the GIG parametrization used in the GIG-stationary Poisson-driven Markov process the pos-

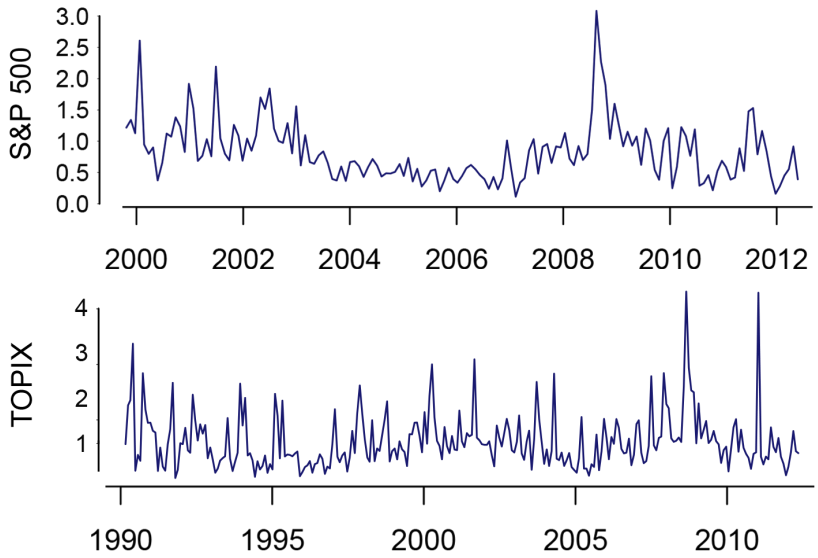


Figure 6.8: S&P 500 and TOPIX series of log monthly minima from January 3, 2000 to July 9, 2012, and from January 1, 1990 to July 31, 2012, respectively. The values are multiplied by -1 to accommodate the GEV stationary distribution.

terior modes $(\lambda, \delta, \gamma, \alpha)$ equal $(-3.14, 0.21, 0.46, 0.70)$. Hence, the stationary distributions adjusted using the GIG-Harris SV model and the GIG-stationary Poisson driven Markov model are very similar (see Figure 6.10), indeed the empirical KL-divergence between both of them is 0.021.

Table VIII: Approximation of the periodicity function of FTSE stock prices. Average of the spot volatility standardized by daily volatility

Day	$\hat{f}$
Monday	0.13
Tuesday	0.14
Wednesday	0.16
Thursday	0.16
Friday	0.16
Saturday	0.24

Once more, we break the sample into an estimation period (from October 31, 2004 to December 31, 2006) and a subsequent forecasting period (from January 1, 2007 to May 31, 2007), and we predict 0.95 probability intervals for the forecasting period with 1000 simulated trajectories. The intervals appear in Figure 6.12. Comparing this figure to Figure 6.7 we can observe that the data sets falling out of the intervals are exactly the

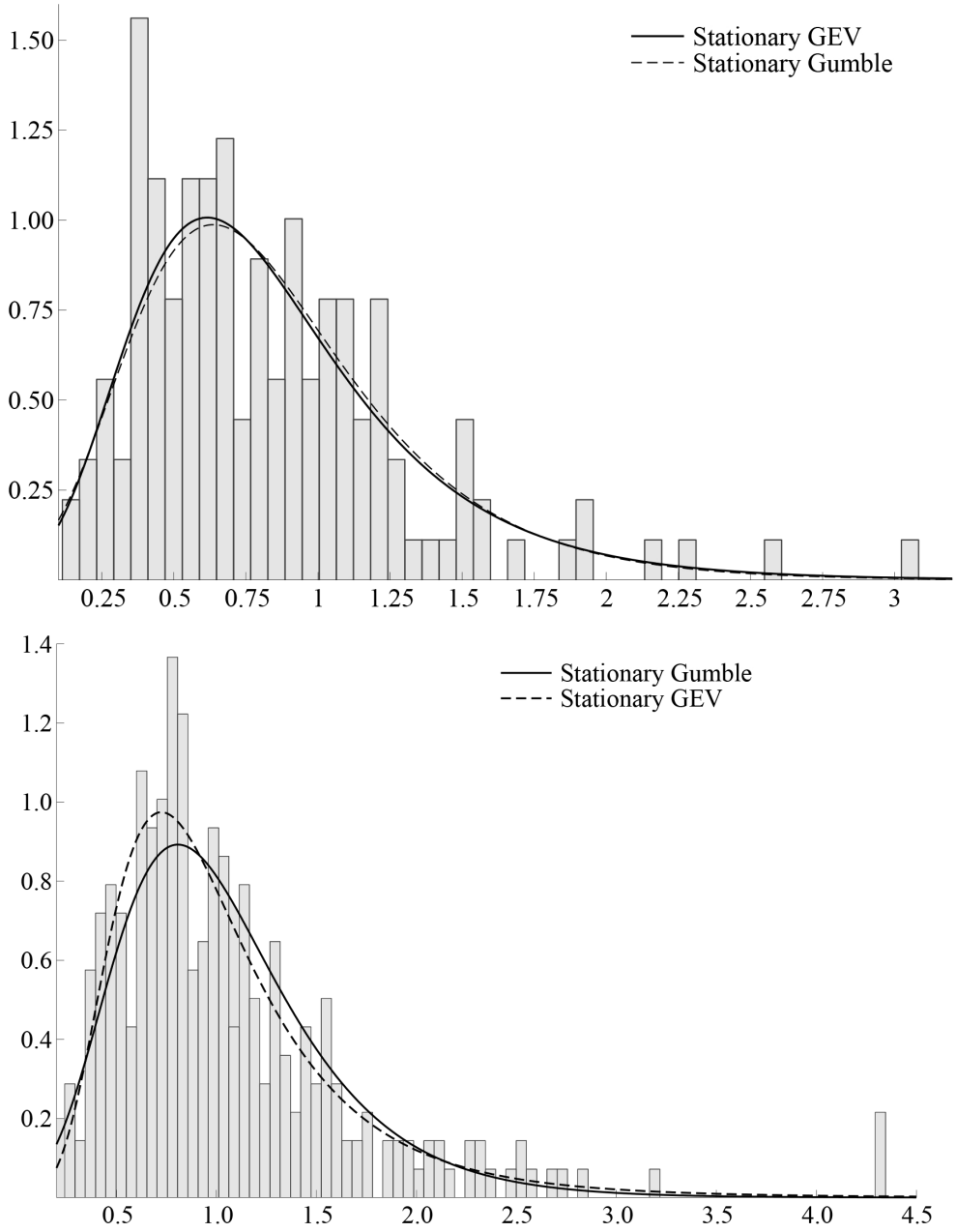


Figure 6.9: Histograms of the S&P 500 (upper panel) and TOPIX (lower panel) series together with the corresponding estimated stationary densities in the type I (solid line) and type II (dashed line) GEV cases.

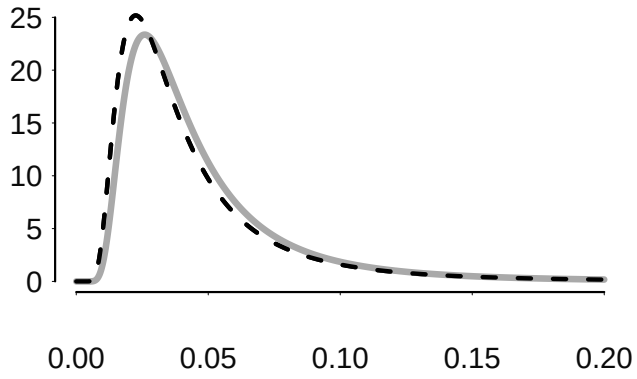


Figure 6.10: In black and dashed appears the estimated denisty for the GIG-Harris SV model. In gray and solid appears the estimated denisty for the GIG-stationary Poisson driven Markov model.

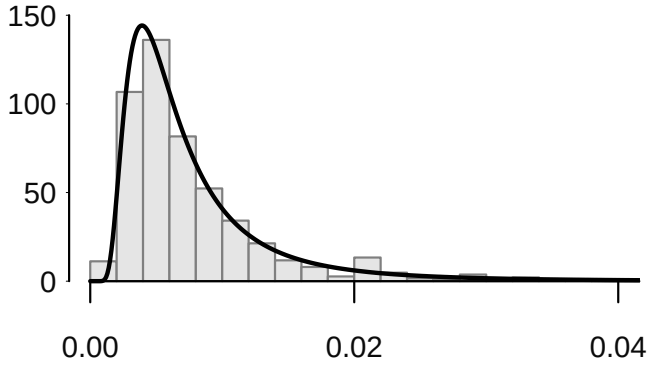


Figure 6.11: Histogram and estimate of the stationary density based on a GIG-Harris SV model for FTSE 100 series. The data are periodically adjusted.

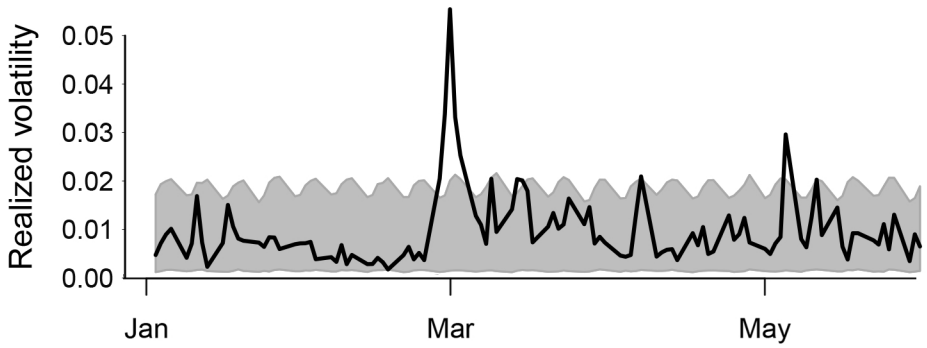


Figure 6.12: FTSE 100 realized volatility series from January 1, 2007 to May 31, 2007, along with highest posterior density intevals of probability 0.95 based on a GIG-Harris SV model. The intervals are computed using the estimation period from October 31, 2003 to December 31, 2006, with 1000 simulated trayectories.

same. As a matter of fact, the same 92 percent of the sample falls within the prediction intervals. This is because a short-memory stationary process modeling the volatility has a strong dependence in the stationary distribution adjusted. In fact, the same has occurred when comparing the SF-Harris process to the OU type process, the prediction results are equally good. However, let us stress two important issues. First, the SF-Harris process has a direct way of estimating the periodicity that, although it was not the case of this data, can improve the adjustment notably in scenarios with systematic variations. Second, the mathematical structure of the transition probabilities of the SF-Harris process is much simpler than the one of OU type processes or f-stationary Poisson driven Markov processes. This is the strength of the GIG-Harris SV model, it is easily tractable and yet equally flexible.

6.5 Comparison to Taylor's SV model

In this section we adjust Taylor (1986) benchmark model to the FTSE equity index. Here, the returns R_t from an asset at time t follow

$$R_t = \exp(\tau_t/2)\epsilon_t,$$
$$\tau_t = \gamma + \phi\tau_{t-1} + \eta_t,$$

where $\{\epsilon_t\}_{t\geq 0}$ and $\{\eta_t\}_{t\geq 0}$ are i.i.d. Gaussian distributed random variables with zero mean and variances 1 and σ_η^2 respectively. The volatility process $\{\tau_t\}_{t\geq 0}$ is a standard Gaussian AR(1) process and, if $|\phi| < 1$, $\{\tau_t\}_{t\geq 0}$ is strictly stationary. In order to estimate this model we use the `stochvol` R package, that provides an efficient estimation algorithm via MCMC methods. Algorithmic details can be found in Kastner and Frühwirth-Schnatter (2014).

Once again, we break the sample into the same estimation period (from October 31, 2003 to December 31, 2006), and a subsequent forecasting period (from January 1, 2007 to May 31, 2007), and we predict 0.95 probability intervals for the forecasting period with 1000 simulated trajectories. As noted in Hansen and Lunde (2005), the fact that the volatility is unobserved has made difficult to evaluate and compare different models. It is not clear which criteria one should use to compare. In their analysis, they compare GARCH-type models in terms of loss functions, some of which are more robust to outliers. Retaining the same idea, we compare the GIG-Harris SV model, the GIG-stationary Poisson-driven Markov process and Taylor's SV model, according to the following loss functions:

$$
\begin{aligned}
MSE_2 &= n^{-1}\sum_{i=1}^{n}\left(\hat{\tau}_i^2 - m_i^2\right)^2, \quad MSE_1 = n^{-1}\sum_{i=1}^{n}\left(\hat{\tau}_i - m_i\right)^2, \\
MAD_2 &= n^{-1}\sum_{i=1}^{n}\left|\hat{\tau}_i^2 - m_i^2\right|, \quad MAD_1 = n^{-1}\sum_{i=1}^{n}\left|\hat{\tau}_i - m_i\right|, \\
R2LOG &= n^{-1}\sum_{i=1}^{n}\left\{\log(\hat{\tau}_i^2 m_i^{-2}\right\}^2,
\end{aligned}
$$

where $\hat{\tau}_i$ denotes the approximated volatility and m_i is the mode of the predictive distribution.

The results appear in Table IX. With two loss functions the GIG-Harris SV model performs slightly better than the GIG-Po-driven SV model, and they both outperform Taylor's model. However, errors are very small, some of the order of 1e−18. Even more, the criteria MAD_2 and MAD_1 are more robust to outliers and are also the ones showing less difference between the three methods. Therefore, we believe that the three methods have a very good predictive performance. However, the GIG-Harris SV model and the the GIG-Po-driven SV model also work for continuous data.

Table IX: Five criteria for SV models comparison

Criteria	GIG-Harris	GIG-Po-driven	Taylor	order
$\mathbf{MSE_2}$	17	17	48	1e−18
$\mathbf{MSE_1}$	9	10	17	1e−10
R2LOG	28	30	39	1e−1
$\mathbf{MAD_2}$	15	15	33	1e−10
$\mathbf{MAD_1}$	19	19	29	1e−6

6.6 Concluding remarks

In this chapter we introduced the f-stationary Poisson-driven Markov processes, another class of strongly stationary Markov models with arbitrary but given invariant distributions. Given a choice of invariant density f, the dependence in the processes is introduced via a Poisson weighed density, which in turn leads to a well defined Markov process. Unlike other existing approaches in the literature, the proposed construction has a useful representation of the underlying transition probability. This leads to an effective MCMC-based estimation procedure. We also showed how the construction can be extended to build new stationary models, without compromising the transition density representation. Particular emphasis was placed on the general classes of GIG and GEV stationary distributions, which themselves constitute interesting choices of models for econometric or financial applications. However, the construction can be applied to any other distribution supported on $\mathbb{R}_+$, leading to alternative approaches in other areas where model stability is a requirement.

Given that the specific dependence of f-stationary Poisson-driven Markov processes is induced by the choice of invariant density f, one is naturally inclined to choose f as general as possible. Two future research directions we plan to pursue to achieve more generality is to adopt phase-type distributions or nonparametric hierarchical mixtures for f (Lo, 1984).

7 Conclusion

Throughout this thesis, we exposed the research advances we have achieved on the class of non-independent-increment, stationary, Markov processes. In particular, our work focuses on methods to construct processes in this class with tractable transition probabilities and given invariant distributions. Such processes are an interesting alternative when empirical data, or theoretical considerations, suggest a phenomenon have a state-space dependent structure. Indeed, they offer distributional flexibility, since the user can choose the marginal distribution, and they have efficient estimation methods, which is a rare quality for continuous-time processes with non-independent increments.

We first introduced SF-Harris processes, which arise extending Harris recurrent Markov chains to continuous time. We proved several stability properties, developed the inference, and applied the processes in two contexts, stochastic volatility modeling and time-dependent density models. In the context of stochastic volatility, we used SF-Harris processes to model spot volatility. The resulting model has a simple transition mechanism driving its dependence, and proves not to compromise its generality when confronting it with other popular models available in the literature. In the context of time-dependent density estimation, we proposed to induce the time dependence of a Bayesian mixture model through both the weights and particles, using independent collections of SF-Harris processes. This is a more elaborated way of modeling the dependence than the existing approaches and it eliminates some of the disadvantages typically found.

After SF-Harris processes, we introduced the f-stationary Poisson-driven Markov processes, another class of strongly stationary Markov models with arbitrary but fixed invariant distributions. Given an invariant density f, we induced the dependence in the processes with a Poisson weighted density. Once again, the proposed construction has a useful representation of the underlying transition probability, leading to an effective MCMC-based estimation procedure. We studied the case of a Gamma invariant distribution, and showed how the construction can be used to build new stationary models, without compromising the transition density representation. We also made a numerical comparison between the SF-Harris process and the GIG-stationary Poisson driven Markov

process. The results showed that the estimation of the stationary density is very similar in both cases and, as a consequence, the prediction results are equally good.

We mentioned possible extensions for the GIG-Harris SV model, such as considering the microstructure noise, inducing long memory, and exploring jump robust estimation methods for the periodic component (see Section 4.5). We also mentioned two future research directions to achieve more generality for the f-stationary Poisson-driven Markov processes. One is to adopt phase-type distributions and the other is to use nonparametric hierarchical mixtures for f (see Section 6.6). Additionally, the full development of the time-dependent density model is still pending (see Section 5.3).

Apart from that, there is still much to do with the processes class that was studied. To begin with, other stability properties allowing statistical treatability could be chosen. To fully understand how restrictive we are in choosing one or another property, a deep mathematical study of the contentions between the stability properties in continuous time is still missing in the literature. In particular, the question about the contention between Harris recurrence and wide-sense regeneration seems interesting. Emphasizing another direction of research, let us remember that Feller processes estimation in general has been scarcely developed and more efficient simulation techniques would allow us to apply them in a wide range of contexts, gaining more flexibility. Finally, the study of the prediction of semi-Markovian processes would immediately generate a wide class of long-memory processes that could be used with financial data.

We hope that this short work in the broad class of stochastic processes, serves to demonstrate its great power when applied in real data, which computational methods and devices have put at our disposal in the last few years.

Bibliography

Andersen, T. G. and T. Bollerslev (1997a). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance 52*(3), 975–1005.

Andersen, T. G. and T. Bollerslev (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance 4*(2), 115–158.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics 89*(4), 701–720.

Antoniano-Villalobos, I. and S. G. Walker (2016). A nonparametric model for stationary time series. *Journal of Time Series Analysis 37*(1), 126–142.

Anzarut, M. and R. H. Mena (Submitted 2016). A Harris process to model stochastic volatility. *Econometrics and Statistics*.

Anzarut, M., R. H. Mena, C. R. Nava, and I. Prünster (2017). Poisson-driven stationary Markov models. *Journal of Business & Economic Statistics*, 1–11.

Asmussen, S. (2003). *Applied probability and queues*, Volume 51. Springer-Verlag New York.

Asmussen, S. and J. Rosiński (2001). Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability 38*(2), 482–493.

Azéma, J., M. Duflo, and D. Revuz (1969). Mesure invariante des processus de Markov récurrents. *Séminaire de Probabilités III Université de Strasbourg*, 24–33.

Bandi, F. M. and R. Renò (2009). Nonparametric stochastic volatility. Global COE Hi-Stat Discussion Paper Series, Institute of Economic Research, Hitotsubashi University.

Barndorff-Nielsen, O., B. Nielsen, N. Shephard, and C. Ysusi (2002). Measuring and forecasting financial variability using realised variance with and without a model. Economics papers, Economics Group, Nuffield College, University of Oxford.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica 76*(6), 1481–1536.

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

Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(2), 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2003). Integrated OU processes and non-Gaussian OU-based stochastic volatility models. *Scandinavian Journal of statistics 30*(2), 277–295.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics 2*(1), 1–37.

Bass, R. F. (1988). Uniqueness in law for pure jump Markov processes. *Probability Theory and related fields 79*(2), 271–287.

Bertoin, J. (1998). *Lévy processes*, Volume 121. Cambridge university press.

Bibby, B. M., I. M. Skovgaard, M. Sørensen, et al. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli 11*(2), 191–220.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. *Handbook of econometrics 4*, 2959–3038.

Böttcher, B. (2010). Feller processes: The next generation in modeling. Brownian motion, Lévy processes and beyond. *PloS one 5*(12), e15102.

Böttcher, B., R. Schilling, and J. Wang (2013). Lévy matters III. *Lévy-type Processes: Construction, Approximation and Sample Path Properties*.

Böttcher, B. and A. Schnurr (2011). The Euler scheme for Feller processes. *Stochastic Analysis and Applications 29*(6), 1045–1056.

Boudt, K., C. Croux, and S. Laurent (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance 18*(2), 353–367.

Buuren, S. and K. Groothuis-Oudshoorn (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software 45*(3).

Caron, F., M. Davy, A. Doucet, E. Duflos, and P. Vanheeghe (2008). Bayesian inference for linear dynamic models with Dirichlet process mixtures. *IEEE Transactions on Signal Processing 56*(1), 71–84.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society 41*(1), 135–155.

Comte, F., V. Genon-Catalot, et al. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. H. Poincaré Probab. Statist 46*(3), 595–617.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*(4), 291–323.

Contreras-Cristan, A., R. H. Mena, and S. G. Walker (2009). On the construction of stationary AR(1) models via random distributions. *Statistics 43*(3), 227–240.

Cox, J. C., J. E. Ingersoll Jr, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 385–407.

Down, D., S. P. Meyn, and R. L. Tweedie (1995, 10). Exponential and uniform ergodicity of Markov processes. *Ann. Probab. 23*(4), 1671–1691.

Dunson, D. B. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics 7*(4), 551–568.

Feller, W. (1966). *An Introduction to probability theory and its applications, vol. II.* New York, London, Sydney: John Wiley and Sons.

Figueroa-López, J. E. and C. Houdré (2006). Risk bounds for the non-parametric estimation of Lévy processes. *Lecture Notes-Monograph Series 51*, 96–116.

Foster, D. P. and D. B. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*(1), 139–74.

Gander, M. P. and D. A. Stephens (2007). Stochastic volatility modelling in continuous time with general marginal distributions: inference, prediction and model selection. *Journal of Statistical Planning and Inference 137*(10), 3068–3081.

Getoor, R. K. (1980). Transience and recurrence of Markov processes. *Séminaire de probabilités de Strasbourg 14*, 397–409.

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. *Handbook of Statistics 14*, 119–191.

Gilks, W. R., N. G. Best, and K. K. C. Tan (1995). Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Applied Statistics 44*(4), 455–472.

Glynn, P. W. (2011). Wide-sense regeneration for Harris recurrent Markov processes: an open problem. *Queueing Systems 68*(3-4), 305–311.

Griffin, J. and M. Steel (2006). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *Journal of Econometrics 134*(2), 605–644.

Griffin, J. and M. Steel (2010). Bayesian inference with stochastic volatility models using continuous superpositions of non-Gaussian Ornstein-Uhlenbeck processes. *Computational Statistics & Data Analysis 54*(11), 2594–2608.

Griffin, J. E. (2010). Inference in infinite superpositions of non-Gaussian Ornstein–Uhlenbeck processes using Bayesian nonparametic methods. *Journal of Financial Econometrics*.

Griffin, J. E. and M. F. Steel (2011). Stick-breaking autoregressive processes. *Journal of econometrics 162*(2), 383–396.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of applied econometrics 20*(7), 873–889.

Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009). Oxford-Man Institute's realized library, version 0.1.

Hoh, W. (1998). Pseudo differential operators generating Markov processes. *Habilitationsschrift, Universität Bielefeld*.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The Journal of finance 42*(2), 281–300.

Jacob, N. (2005). *Pseudo differential operators and Markov processes*, Volume 3. Imperial College Press.

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, 664–677.

Jørgensen, B. and P. X.-K. Song (1998). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability*, 78–92.

Kallenberg, O. (2002). *Foundations of modern probability*. Probability and its applications. New York, Berlin, Paris: Springer.

Kalli, M., J. E. Griffin, and S. G. Walker (2011). Slice sampling mixture models. *Statistics and computing 21*(1), 93–105.

Kanaya, S. and D. Kristensen (2016). Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory*, 1–56.

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

Kibot (2015, January). Historical intraday data. `http://www.kibot.com`.

Kunita, H. (1997). *Stochastic flows and stochastic differential equations*, Volume 24. Cambridge university press.

Kyprianou, A. E. (2006). *Introductory lectures on fluctuations of Lévy processes with applications*. Springer.

Lévy, P. (1954). Processus semi-Markoviens. In *Proceedings of the International Congress of Mathematicians*, Volume 3, pp. 416–426.

Liggett, T. M. (2010). *Continuous time Markov processes: an introduction*, Volume 113. American Mathematical Soc.

Lo, A. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The annals of statistics 12*(1), 351–357.

MacEachern, S. N. (2000). Dependent Dirichlet processes. *Technical report, Department of Statistics, The Ohio State University*, 1–40.

McKenzie, E. (2003). Ch. 16. Discrete variate time series. *Handbook of statistics 21*, 573–606.

Mena, R. H. and M. Ruggiero (2016). Dynamic density estimation with diffusive Dirichlet mixtures. *Bernoulli 22*(2), 901–926.

Mena, R. H. and S. G. Walker (2005). Stationary autoregressive models via a bayesian nonparametric approach. *Journal of Time Series Analysis 26*(6), 789–805.

Mena, R. H. and S. G. Walker (2009). On a construction of Markov models in continuous time. *Metron 67*(3), 303–323.

Meyn, S. P. and R. L. Tweedie (1993). Stability of Markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability 25*(3), 487–517.

Meyn, S. P. and R. L. Tweedie (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Nakajima, J., T. Kunihama, Y. Omori, and S. Frühwirth-Schnatter (2012). Generalized extreme value distribution with time-dependence using the AR and MA models in state space form. *Computational Statistics & Data Analysis 56*(11), 3241–3259.

Nava, C. R. (2013). *Some advances in Bayesian modeling and computation*. Ph. D. thesis, Università di Torino.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer Journal 7*(4), 308–313.

Neumann, M. H., M. Reiß, et al. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli 15*(1), 223–248.

Nocedal, J. and S. J. Wright (2006). *Numerical optimization* (2nd ed.). New York: Springer.

Nolan, J. (2003). *Stable distributions: models for heavy-tailed data*. Birkhauser.

Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 43*(4), 309–318.

Oodaira, H. and K.-i. Yoshihara (1972). Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. *Kodai Mathematical Seminar Reports 24*(3), 259–269.

Pitt, M. K., C. Chatfield, and S. G. Walker (2002). Constructing first order stationary autoregressive models via latent processes. *Scandinavian Journal of Statistics 29*(4), 657–663.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhyā: The Indian Journal of Statistics, Series A*, 311–324.

Rodriguez, A. and D. B. Dunson (2011, 03). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian analysis 6*(1), 145–177.

Rodriguez, A. and E. Ter Horst (2008). Bayesian dynamic density estimation. *Bayesian Analysis 3*(2), 339–365.

Ross, S. M. (2014). *Introduction to Probability Models (11th Edition)*. Academic press.

Samorodnitsky, G. (2007). Long range dependence. *Foundations and Trends in Stochastic Systems 1*(3), 163–257.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. Cox, D. V. Hinkley, and O. E. Barndorff-Neilsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.

Shephard, N. and T. G. Andersen (2009). Stochastic volatility: Origins and overview. In T. Mikosch, J.-P. Kreib, R. A. Davis, and T. G. Andersen (Eds.), *Handbook of Financial Time Series*, pp. 233–254. Springer Berlin Heidelberg.

Sigman, K. and R. W. Wolff (1993). A review of regenerative processes. *SIAM review 35*(2), 269–288.

Smith, W. L. (1955). Regenerative stochastic processes. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Volume 232, pp. 6–31. The Royal Society.

Sørensen, M. (2000). Prediction-based estimating functions. *Econometrics Journal 3*(2), 123–147.

Stroock, D. W. (2003). *Markov processes from K. Itô's perspective*. Number 155 in Annals of Mathematics Studies. Princeton University Press.

Tankov, P. (2003). *Financial modelling with jump processes*. CRC Financial Mathematics Series.

Urbanik, K. (1967). Some prediction problems for strictly stationary processes. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1*, Berkeley, Calif., pp. 235–258. University of California Press.

Vega-Amaya, Ó. (2012). On the regularity property of semi-Markov processes with Borel state spaces. In *Optimization, Control, and Applications of Stochastic Systems*, pp. 301–309. Springer.

Volnỳ, D. (1993). Approximating martingales and the central limit theorem for strictly stationary processes. *Stochastic processes and their applications 44*(1), 41–74.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics Simulation and Computation 36*(1), 45–54.

Wiggins, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of financial economics 19*(2), 351–372.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli 12*(6), 1019–1043.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100*(472), 1394–1411.