

A BAYESIAN ANALYSIS OF THE COX-INGERSOLL-ROSS MODEL

2026

BANCO DE **ESPAÑA**
Eurosistema

Documentos de Trabajo
N.º 2640

Dante Amengual and Lluc Puig-Codina

A BAYESIAN ANALYSIS OF THE COX-INGERSOLL-ROSS MODEL ^(*)

Dante Amengual ^(**)

CENTRO DE ESTUDIOS MONETARIOS Y FINANCIEROS

Lluc Puig-Codina ^(***)

BANCO DE ESPAÑA AND UNIVERSITY OF ALICANTE.

(*) We are grateful to Gabriel Pérez-Quirós, Juan Mora López and Ángel León Valle for their feedback and suggestions. We thank colleagues at the Banco de España, participants at the 2026 Annual Conference on Economic Analysis at the University of Alicante and participants at the 16th RCEA Bayesian Econometrics Workshop for their helpful comments and discussions. Dante Amengual gratefully acknowledges financial support from grant PID2025-175659NB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU. The views expressed are those of the authors and do not necessarily represent the views of the Banco de España, the European Central Bank or the Eurosystem.

(**) amengual@cemfi.es

(***) lluc.puig@bde.es

Documentos de Trabajo. N.º 2640

October 2026

<https://doi.org/10.53479/44785>

The Working Paper Series seeks to disseminate original research in economics and finance. All papers have been anonymously refereed. By publishing these papers, the Banco de España aims to contribute to economic analysis and, in particular, to knowledge of the Spanish economy and its international environment.

The opinions and analyses in the Working Paper Series are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

The Banco de España disseminates its main reports and most of its publications via the Internet at the following website: <http://www.bde.es>.

Reproduction for educational and non-commercial purposes is permitted provided that the source is acknowledged.

© BANCO DE ESPAÑA, Madrid, 2026

ISSN: 1579-8666 (online edition)

Abstract

The Cox-Ingersoll-Ross model (CIR) is a diffusion process with widespread applications in economics, finance and population biology. In this paper we first obtain a full Gibbs Sampler for the exact CIR process observed at equidistant times. We do not require the imposition of the Feller condition. We show how to carry out smoothing and handling of missing data. Additionally, the introduction of a tailored single Metropolis-Hastings step allows us to handle irregular and random time intervals. Secondly, we obtain several asymptotic properties which include Local Asymptotic Normality, consistency and asymptotic normality of conditional maximum likelihood estimation and a Bernstein-von Mises theorem. Lastly, we provide an empirical illustration using U.S. Treasury yields under a variety of priors and a comparison with other models such as the Vasicek and the 3/2 models.

Keywords: Cox-Ingersoll-Ross process, square-root diffusion, auto-regressive gamma process, latent variables, data augmentation, Gibbs sampling, Bernstein-von Mises theorem.

JEL classification: C11, C15, C22, C58, G12.

Resumen

El modelo de Cox-Ingersoll-Ross (CIR) es un proceso de difusión con amplias aplicaciones en economía, finanzas y biología de poblaciones. En este documento obtenemos, en primer lugar, un muestreador de Gibbs puro para el proceso CIR exacto observado en intervalos equiespaciados. No exigimos que se cumpla la condición de Feller y mostramos cómo llevar a cabo el suavizado y la imputación de datos faltantes. Adicionalmente, la introducción de un único paso de Metropolis-Hastings diseñado específicamente para este problema nos permite manejar intervalos de tiempo irregulares y aleatorios. En segundo lugar, obtenemos varias propiedades asintóticas, entre ellas la normalidad asintótica local, la consistencia y la normalidad asintótica del estimador de máxima verosimilitud condicional, así como un teorema de Bernstein-von Mises. Por último, presentamos una aplicación empírica utilizando rendimientos de bonos del Tesoro estadounidense bajo diversas distribuciones a priori, junto con una comparación con otros modelos, tales como el modelo de Vasicek y el modelo de $3/2$.

Palabras clave: proceso de Cox-Ingersoll-Ross, difusión de raíz cuadrada, proceso autorregresivo gamma, variables latentes; aumento de datos, muestreo Gibbs, teorema de Bernstein-von Mises.

Códigos JEL: C11, C15, C22, C58, G12.

1 Introduction

Among his many contributions William Feller (1906-1970) introduced a continuous-state branching diffusion in Feller (1951a,b) known as a *square-root diffusion* process. A square-root diffusion process $\{X_t\}$ with parameters $a \in [0, \infty)$, $b \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$ is the strong solution of the stochastic differential equation

$$dX_t = (a + bX_t) dt + \sigma \sqrt{|X_t|} dW_t \quad (1)$$

with initial condition $X_0 = x_0 > 0$, where W_t is a standard one-dimensional Brownian motion. Let $\eta = (a, b, \sigma^2)$ collect the parameters and $[0, \infty) \times \mathbb{R} \times (0, \infty)$ be the parameter space.

Initially intended to model population dynamics in biology, equation 1 can be obtained as a diffusion (scaling) limit of a discrete-space branching process with immigration. As specimens reproduce and die independently, both the mean and variance of population size over a short time interval are proportional to current population, X_t , hence the name square-root diffusion.

We summarize the qualitative behavior of the process according to the range of parameter values. The process explodes exponentially at rate b whenever $b > 0$ (supercritical). For $b < 0$ (subcritical) the process converges to the long run mean $-a/b$ at speed $-b$. When $b = 0$ the process lies at the (critical) boundary between exponential growth and mean reversion. The process has a linear drift at and deviations from the trend, $X_t - at$, are a martingale. A second classification is in the behavior at zero, controlled by the *Feller index* $\frac{2a}{\sigma^2}$. If the process reaches zero the diffusion term is killed and $a > 0$ ensures the process exits zero. Furthermore, when the *Feller condition* holds, $\frac{2a}{\sigma^2} \geq 1$, the drift will be strong enough so that zero is unattainable. When $0 < \frac{2a}{\sigma^2} < 1$ the process will reach zero but it will exit instantaneously, zero is a reflecting barrier. Finally, when $a = 0$, no such pull away from 0 exists and so zero is absorbing. Given this informal discussion, X_t will never take negative values and so we can drop $|\cdot|$ from equation 1.

Despite its original intent, equation 1 was introduced to economics by the seminal paper of Cox et al. (1985) as a model of the instantaneous nominal interest rate, or state variables,

in an equilibrium model of the term structure of interest rates. From now on, we will refer to equation 1 as the *Cox-Ingersoll-Ross* (CIR) process. Financial applications typically restrict the parameter space to $b < 0$ and $a > 0$, under which the process is stationary ergodic with a Gamma stationary distribution, $\mathcal{G}\left(\frac{2a}{\sigma^2}, -\frac{\sigma^2}{2b}\right)$. As a model of the term structure of interest rates Cox et al. (1985) has faced empirical difficulties, cf. Jagannathan et al. (2003), and the financial industry has moved towards alternative, no-arbitrage, term structure models. Nevertheless the CIR process still enjoys widespread use in the modeling of instantaneous or short term rates in the macro-finance literature cf. Lin et al. (2018) and Guo et al. (2022), for the dynamics of a latent volatility process, cf. Heston (1993) and Brunnermeier et al. (2024), or as a latent hazard process, cf. Bielecki et al. (2011) and Wachter (2013).

The literature on the estimation of the CIR process can be divided into two broad strands: high- and low-frequency estimation. High frequency estimation is typically based on *continuous sampling*, under which a closed form expression for the maximum likelihood estimator can be derived, cf. Overbeck (1998) and Ben Alaya and Kebaier (2012). To make these results operational, the corresponding stochastic integrals are replaced by sums that converge under infill asymptotics, $\Delta_T \rightarrow 0$ and T fixed, cf. Section 3.3 in Overbeck and Rydén (1997), or double asymptotics, $\Delta_T \rightarrow 0$ and $\Delta_T T \rightarrow \infty$, cf. Ben Alaya and Kebaier (2013). A notable exception is Cheng et al. (2024), who study Gaussian quasi-likelihood estimation under double asymptotics. Low frequency estimation, in contrast, considers observations at fixed equidistant times, $0 < \Delta < \infty$. This setting is studied by Overbeck and Rydén (1997), who propose conditional least-squares estimators. Gouriéroux and Jasiak (2006) propose an auto-regressive gamma process, equivalent to the CIR observed at fixed equidistant times, study its properties and provide an application to duration data. Creal (2017) employs the auto-regressive gamma process as the latent variable of a stochastic volatility model, a discrete time equivalent of Heston (1993), obtains filtering and smoothing results, and estimates the parameters by maximum likelihood. León-González (2019) employs the auto-regressive gamma process as latent precision and derives an MCMC routine where sampling from the conditional posterior distribution of the parameters is done with a generic Metropolis-Hastings algorithm.

The objective of this paper is to study the CIR process from a Bayesian perspective. The posterior distribution of the parameters, $\pi(\boldsymbol{\eta} \mid \tilde{x}_T)$, after observing a single sample path of $\{X_t\}$ at discrete times, $\tilde{x}_T$, is the primary object of interest. Throughout the paper, we will devote our attention to the non-absorbing case, $a > 0$.

A general approach to Bayesian estimation of Itô diffusions, proposed by Elerian et al. (2001) and Eraker (2001), combines the Euler-Maruyama approximation with a data augmentation scheme that artificially increases the sampling frequency. As the frequency is increased, the Euler-Maruyama approximation becomes more accurate; however, the conditional posterior increasingly depends on artificially generated observations, leading to posterior degeneracy in the limit. As a result, the method exhibits an intrinsic, model-dependent limit to precision that is difficult to quantify. For the CIR process in particular, Euler- and Milstein-type discretizations may suffer from slow convergence when the Feller condition is not satisfied, cf. Hefter and Jentzen (2019). In this paper we will employ the exact transition density for the CIR process, without relying on the Euler-Maruyama or other approximations.

The structure of the paper is as follows. In Section 2 we obtain an exact full Gibbs Sampler for the case of equidistant observations at fixed and finite increments (Proposition 1), study smoothing and *data missing at random* (Section 2.2) and show how to handle irregular and random time increments by introducing a single tailored independent Metropolis-Hastings step (Section 2.3). In Section 3 we show several asymptotic results, including *Local Asymptotic Normality* (LAN, Theorem 2), consistency and asymptotic normality of conditional maximum likelihood estimation (Theorem 3) and a Bernstein-von Mises Theorem (Theorem 4). Section 4 provides an application to 3 Year U.S. Treasury yields and a comparison with the *Vasicek* and *3/2* models. Section 5 concludes. All proofs are collected in Appendix A.

2 Bayesian Analysis

2.1 Equidistant Observations

The conditional distribution of the CIR process at equidistant time intervals, $0 < \Delta < \infty$, is given by a scaled non-central χ^2 ,

$$\frac{-4b}{(1 - e^{b\Delta})\sigma^2} X_t \mid \tilde{X}_{t-1} = \tilde{x}_{t-1} \sim \chi^2_{\frac{4a}{\sigma^2}} \left(\frac{-4b}{\sigma^2} \frac{e^{b\Delta}}{1 - e^{b\Delta}} x_{t-1} \right)$$

with the case of $b = 0$ being defined by the continuity limit of the above expression. $\tilde{x}_t$ denotes the observed history of $\{X_s\}$ up to time t . We propose the following one-to-one

reparameterization,

$$\rho = e^{b\Delta}; \quad n = \frac{4a}{\sigma^2}; \quad \tau = \begin{cases} \frac{\sigma^2(e^{b\Delta}-1)}{4b} & b \neq 0 \\ \frac{\sigma^2\Delta}{4} & b = 0 \end{cases}$$

where $\rho, \tau, n \in (0, \infty)$. Let $\boldsymbol{\theta} = \{\rho, \tau, n/2\}$ collect the parameters and $\boldsymbol{\Theta} = (0, \infty)^3$ be the parameter space. With this reparameterization the transition density of the CIR process is given by,

$$\frac{1}{\tau} X_t \mid \tilde{X}_{t-1} = \tilde{x}_{t-1} \sim \chi_n^2 \left(\frac{\rho}{\tau} x_{t-1} \right)$$

Since the non-central χ^2 distribution can be expressed as a Poisson mixture of Gammas (cf. Muirhead (1982) Corollary 1.3.5.), the CIR process observed at equidistant times can be stated in the following state-space form,

$$\begin{aligned} X_t \mid \tilde{X}_{t-1} = \tilde{x}_{t-1}, \tilde{H}_t = \tilde{h}_t &\sim \mathcal{G} \left(\frac{n}{2} + h_t, 2\tau \right) \\ H_t \mid \tilde{X}_{t-1} = \tilde{x}_{t-1}, \tilde{H}_{t-1} = \tilde{h}_{t-1} &\sim \mathcal{P} \left(\frac{\rho}{2\tau} x_{t-1} \right) \end{aligned}$$

In addition to the auxiliary variables introduced by the mixture, $\{H_t\}$, we will also introduce an additional layer of artificial variables, building on the strategy of Hamura et al. (2023),

$$S_t \mid H_t = h_t \sim \mathcal{Beta} \left(\frac{n}{2}, h_t \right)$$

S_t 's are conditionally independent given $\tilde{h}_T$. The conditional distribution of S_t is stated with an abuse of notation. S_t is equal to 1 whenever $h_t = 0$ and is distributed $\mathcal{Beta} \left(\frac{n}{2}, h_t \right)$ whenever $h_t \geq 1$.

The econometrician observes not only $\tilde{x}_T$, but also Δ . To work with the conditional likelihood expressed by the above state-space model we need to make some assumptions

related to the time increment. When observations are equidistant we assume Δ to be fixed.

In order to carry out a Bayesian estimation of the CIR process we derive a data-augmentation based Gibbs sampler which is summarized in the next Proposition.

Proposition 1 *Gibbs Sampler for equidistant observations*

Let $0 < \Delta < \infty$ be fixed and priors take the form

$$\begin{aligned}\pi(\boldsymbol{\theta} \mid \Delta, x_0) &= \pi(\tau \mid \Delta, x_0) \pi(\rho \mid \Delta, x_0) \pi\left(\frac{n}{2} \mid \Delta, x_0\right), \\ \tau \mid \Delta, x_0 &\sim \mathcal{IG}(\underline{\alpha}, \underline{\beta}), \\ \rho \mid \Delta, x_0 &\sim \mathcal{G}(\underline{a}, \underline{b}), \quad \text{and} \\ \pi\left(\frac{n}{2} \mid \Delta, x_0\right) &\propto \underline{\lambda}^{\frac{\nu}{2}} \Gamma\left(\frac{n}{2}\right)^{-\nu},\end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function, $\underline{\alpha}, \underline{\beta}, \underline{a}, \underline{\lambda}, \underline{\nu} > 0$ and $\underline{b} < \infty$.

Then,

$$\begin{aligned}h_t \mid \tilde{x}_T, \Delta, \tau, \rho, \frac{n}{2} &\sim \mathcal{Bessel}\left(\frac{n}{2} - 1, \frac{\sqrt{\rho x_t x_{t-1}}}{\tau}\right), \\ s_t \mid \tilde{h}_T, \frac{n}{2} &\sim \mathcal{Beta}\left(\frac{n}{2}, h_t\right), \quad \text{and} \\ \tau \mid \tilde{x}_T, \Delta, \tilde{h}_T, \rho, \frac{n}{2} &\sim \mathcal{IG}(\bar{\alpha}, \bar{\beta}),\end{aligned}$$

$$\text{where } \bar{\alpha} = \underline{\alpha} + \frac{n}{2}T + 2 \sum_{t=1}^T h_t, \quad \bar{\beta} = \underline{\beta} + \frac{1}{2} \left(\rho x_0 + x_T + (1 + \rho) \sum_{t=1}^{T-1} x_t \right), \quad \text{and}$$

$$\rho \mid \tilde{x}_T, \Delta, \tilde{h}_T, \tau \sim \mathcal{G}(\bar{a}, \bar{b})$$

$$\text{with } \bar{a} = \underline{a} + \sum_{t=1}^T h_t \quad \text{and} \quad \bar{b} = \left(\frac{1}{\underline{b}} + \frac{1}{2\tau} \sum_{t=0}^{T-1} x_t \right)^{-1},$$

$$\frac{n}{2} \mid \tilde{x}_T, \Delta, \tilde{s}_T, \tau \propto \bar{\lambda}^{\frac{\bar{\nu}}{2}} \Gamma\left(\frac{n}{2}\right)^{-\bar{\nu}}$$

$$\text{where } \bar{\lambda} = \underline{\lambda}^{\frac{\nu}{\bar{\nu}}} \left[\left(\prod_{t=1}^T \frac{x_t s_t}{2\tau} \right)^{1/T} \right]^{T/\bar{\nu}} \quad \text{and} \quad \bar{\nu} = \underline{\nu} + T$$

The conditional posteriors of ρ and τ take the form of well-known distributions, Gamma and Inverse-Gamma. The conditional posterior of $\tilde{h}_T$ is a product of independent but not identically distributed *Bessel* random variables, defined in Yuan and Kalbfleisch (2000). The Bessel conditional distribution of the latent Poisson variable was first obtained by Creal (2017) in the context of the auto-regressive gamma process. Exact simulation schemes for the Bessel distribution are given in Devroye (2002) and Iliopoulos and Karlis (2003).

Finally, the posterior kernel of $\frac{n}{2}$ resembles that of a shifted Conway-Maxwell Poisson distribution if $\frac{n}{2} - 1$ was restricted to the non-negative even integers. If instead we allow for $n \in (0, \infty)$ the posterior kernel is that of the GamCon distribution of Type 1 ($\mathcal{GC}1$), introduced by Damsleth (1975) as the conjugate distribution for the shape of a gamma random variable with a known scale parameter. The $\mathcal{GC}1$ distribution is twice differentiable, part of the exponential family and has a unique mode at $\psi^{-1}(\ln(\lambda))$; where $\psi(\cdot)$ is the digamma function. Efficient and exact simulation can be performed with the algorithm of Devroye (2014).

The sum of J squared independent Ornstein-Uhlenbeck (OU) processes with identical parameters yields a CIR process (cf. Revuz and Yor (1999)) with n restricted to the positive integers. Imposing $n \in 2\mathbb{Z}_{\geq 1}$ such that the prior and conditional posterior of $\frac{n}{2} - 1$ are Conway-Maxwell Poisson may be employed to reflect prior knowledge that the observed CIR process originated from an even number of squared independent OU processes with identical parameters. In later examples and applications $n/2$ is assumed to be continuous.

Conditionally conjugate priors can be of considerable help in prior elicitation as they allow for an interpretation of the prior in terms of equivalent data, cf. Box and Tiao (1973).¹ Notice that all conditional posterior distributions are proper whenever $T \geq 1$ and $x_t > 0 \forall t$, in addition to the weak restriction on the prior hyperparameters, so that a wide array of priors are allowed. The fact that all observations must be positive does not imply that the Feller condition must hold. Due to fixed $\Delta > 0$, no observation will take a value of 0 with probability one. Under these conditions Theorem 1 shows that the Gibbs Sampler is Harris ergodic, so the sequence generated by the Gibbs Sampler will converge in distribution to the posterior.

¹In this case in terms of data and states.

Theorem 1 *Ergodicity of the Gibbs Sampler.*

Assume $\underline{\alpha}, \underline{\beta}, \underline{a}, \underline{\lambda}, \underline{\nu} > 0$, $\underline{b} < \infty$, $T \geq 1$ and $x_t > 0 \forall t$. Then, the Gibbs Sampler given in Proposition 1 is Harris ergodic.

Our proposed reparameterization preserves a natural interpretation for ρ , τ and n . ρ regulates the persistence of the process over a Δ interval, τ is a scale parameter as for small Δ , $\tau \approx \sigma^2 \Delta / 4$, the diffusion variance accumulated over a small increment. Lastly, n regulates the degrees of freedom of the transition density. Our reparameterization also allows us to specify truncated prior distributions such that the parameter space is restricted to regions where the process exhibits different qualitative behaviors as discussed in the introduction. Truncating $\frac{n}{2}$ to be larger than 1 imposes that the Feller condition is satisfied. Further truncating $\frac{n}{2}$ to be larger than a counting number k guarantees the existence of the k -th negative moment, $\mathbb{E} \{X_t^{-k}\}$. Truncating ρ to be below, equal to, or above 1 restricts the process to be in the subcritical, critical or supercritical region. The corresponding posteriors are the same as in Proposition 1 truncated to the aforementioned regions.

The Gibbs sampler in Proposition 1 can be summarized by the following steps. Begin with an arbitrary value $\theta^{(0)} \in \Theta$. Then,

1. Sample $(\tilde{s}_T^{(m+1)}, \tilde{h}_T^{(m+1)})$ from $\pi(\tilde{s}_T | \tilde{h}_T, (n/2)^{(m)}) \times \pi(\tilde{h}_T | \tilde{x}_T, \tau^{(m)}, \rho^{(m)}, (n/2)^{(m)})$.
2. Sample $\tau^{(m+1)}$ from $\pi(\tau | \tilde{x}_T, \tilde{h}_T^{(m+1)}, \rho^{(m)}, (n/2)^{(m)})$.
3. Sample $(\rho^{(m+1)}, (n/2)^{(m+1)})$ from $\pi(\rho | \tilde{x}_T, \tilde{h}_T^{(m+1)}, \tau^{(m+1)}) \times \pi(\frac{n}{2} | \tilde{x}_T, \tilde{s}_T^{(m+1)}, \tau^{(m+1)})$.

We illustrate the above sampler with a simulated example. We simulate a sample of size $T = 500$ from a CIR process observed at equidistant times with $\Delta = 0.4$, parameters $a = 0.5$, $b = -0.2$, $\sigma^2 = 0.2$ and initial value $x_0 = 2.5$. This implies values of $\tau_0 = 0.0192$, $\rho_0 = 0.9231$ and $\frac{n_0}{2} = 5$. Figure 1 depicts the simulated sample. Figure 2 reports the posterior distributions of both the continuous time (η) and the discrete time (θ) parametrizations. The dashed vertical lines indicate the true parameter values. We iterate MCMC simulation 500 thousand times after a burn-in period of 15 thousand MCMC draws. The prior for τ and $\frac{n}{2}$ is given by $\underline{\alpha} = \underline{\beta} = \underline{\nu} = 0.1$ and $\underline{\lambda} = 1$, which correspond to a proper prior but are close to the improper flat prior equivalent to the case of “no prior observations”. For the prior of ρ we employ $\underline{a} = 1/3$ and $\underline{b} = 99$ which correspond to a proper prior but are close to the *non-informative neutral* prior proposed by Kerman (2011).

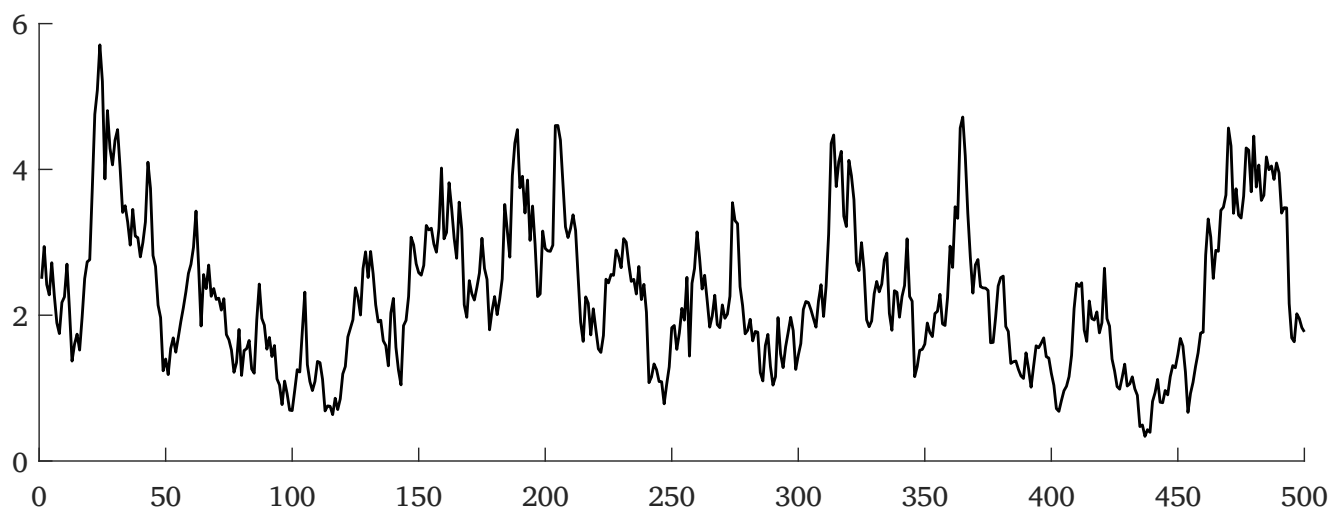


Figure 1: Simulated sample at equidistant times.

$$dX_t = (0.5 - 0.2X_t) dt + \sqrt{0.2X_t} dW_t$$

$$T = 500, \quad X_0 = 2, \quad \Delta = 0.4,$$

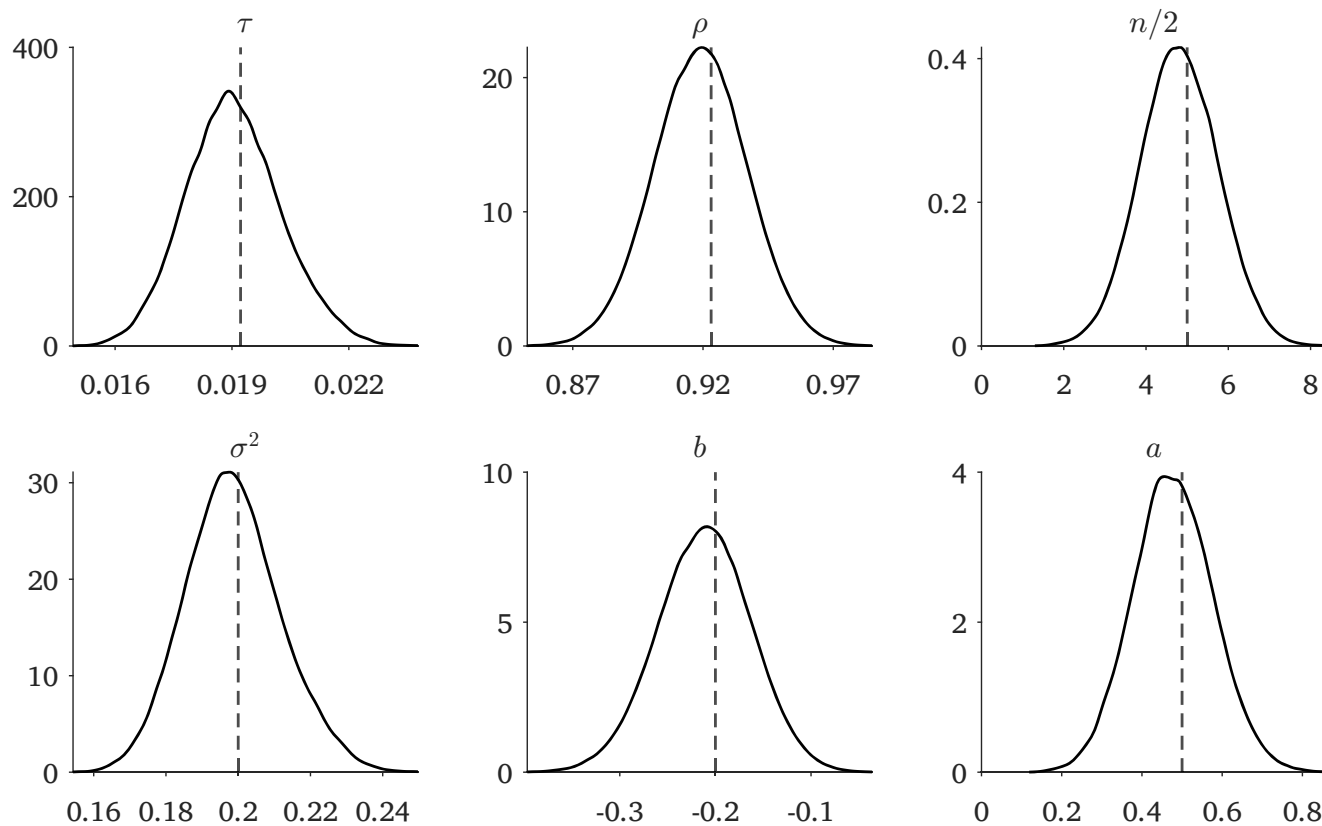


Figure 2: Marginal Posterior Distributions.

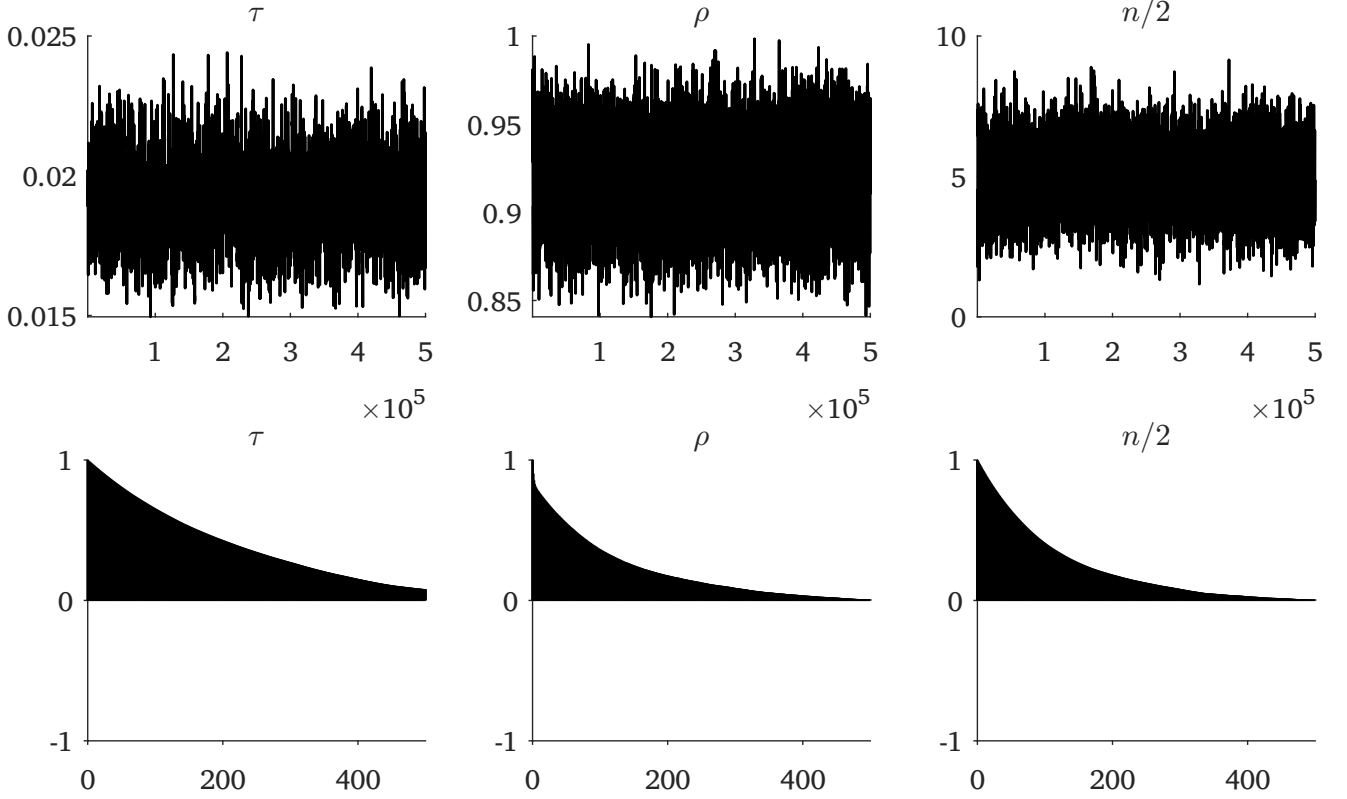


Figure 3: Trace and autocorrelation plots of θ .

Regarding the mixing properties of the sampler, Figure 3 shows the sample paths of the simulated θ and autocorrelations. The autocorrelations of ρ and $\frac{n}{2}$ decay rather fast while autocorrelations of τ decay slowly, reaching 0.12 at lag 500. Unreported simulation experiments show that the persistence of τ is sensitive to the parameter values, being more persistent when the Feller condition does not hold and when the true τ is very small. When the true τ is very small the complication is similar to that of hierarchical linear and generalized linear models with small values of the variance parameters. In our case, sampling small values of τ will result in sampling large values of $\tilde{h}_T$ which in turn will result in small sampled values of τ in the next iteration. Small values of $n/2$ also induce higher values of $\tilde{h}_T$, reinforcing the τ loop. In the case of linear hierarchical models, Gelman et al. (2008) propose the introduction of redundant parameters, also known as *parameter expanded-data augmentation* (PX-DA), which improves mixing substantially. Other techniques to speed up data augmentation algorithms are described in detail by Roy et al. (2024).

In Appendix C we explore additional techniques and compare their efficiency, which include a Collapsed Gibbs Sampler, a Random-Walk Metropolis-Hastings and a hybrid combination. Table 1 reports the effective sample size per minute using the data of Figure 1.

Table 1: Effective Sample Size per Minute

	τ	ρ	$n/2$
Gibbs (Proposition 1)	86	172	158
Collapsed Gibbs	10064	229	175
Random-walk Metropolis-Hastings	79	419	1772
Hybrid	17819	1237	1256

Notes: All four samplers are estimated on the same simulated sample of Section 2.1 ($T = 500$), from identical starting values and single-threaded on the same machine, with running times between 11.3 and 13.8 minutes. Entries report the effective sample size (ESS) per minute; larger values indicate more efficient sampling. The ESS is computed as $N/\hat{\tau}_{\text{int}}$, where $\hat{\tau}_{\text{int}}$ is the integrated autocorrelation time estimated by the initial monotone sequence estimator of Geyer (1992), based on 500 thousand MCMC draws after a burn-in of 15 thousand. The collapsed Gibbs, random-walk Metropolis-Hastings and hybrid samplers are those of Sections C.1, C.2 and C.3, respectively.

As expected, the random-walk Metropolis-Hastings is the worst performing on τ since it retains the $\tau \leftrightarrow \tilde{h}_T$ loop and adds the persistence on $\theta^{(m)}$, while it produces a substantial improvement in the mixing of ρ and $n/2$. The collapsed Gibbs shows a substantial improvement in the mixing of τ due to breaking the $\tau \leftrightarrow \tilde{h}_T$ loop, while the mixing improvements of ρ and $n/2$ are moderate. The hybrid approach gathers the strengths of the collapsed Gibbs and the random-walk Metropolis-Hastings, achieving large improvements in the mixing of all parameters. See Appendix C for details and further comparisons.

Finally we would like to offer some intuition as to the data augmentation of $\tilde{s}_T$. This intuition is provided in terms of the birth-death process with immigration and will reveal some interesting features. Given the state variable H_t , interpreted as the number of survivors, we can decompose the observation equation into the contributions of new arrivals, Λ_t , and the contribution of survivors, $\xi_t(H_t)$, both of which arrive/reproduce at some random rate.

$$\begin{aligned}
X_t &= \xi_t(H_t) + \Lambda_t \\
H_t \mid X_{t-1} = x_{t-1} &\sim \mathcal{P}\left(\frac{\rho}{2\tau}x_{t-1}\right), \\
\xi_t \mid H_t = h_t &\sim \mathcal{G}(h_t, 2\tau) \quad \text{and} \quad \Lambda_t \sim \mathcal{G}\left(\frac{n}{2}, 2\tau\right)
\end{aligned}$$

ξ_t and Λ_t are, conditionally on $\tilde{h}_T$, independent contemporaneously as well as in leads and lags. Define $s_t \equiv \Lambda_t / (\Lambda_t + \xi_t(h_t))$, the latent share of x_t attributable to immigration given survivors h_t . Because Λ_t and $\xi_t(h_t)$ are independent Gamma random variables sharing the common scale 2τ , the Beta-Gamma duality implies that s_t is independent of $\Lambda_t + \xi_t(h_t) = x_t$ given h_t and, in particular,

$$s_t \mid h_t \sim \mathcal{Beta}\left(\frac{n}{2}, h_t\right), \quad \Lambda_t \mid x_t, h_t \stackrel{d}{=} s_t x_t.$$

The multiplication of the share s_t by the observation x_t recovers the arrivals Λ_t exactly in distribution. The introduction of s_t isolates the part of x_t that loads on $\frac{n}{2}$, yielding a weighted geometric mean in the conditional posterior $\pi\left(\frac{n}{2} \mid \tilde{x}_T, \Delta, \tilde{s}_T, \tau\right)$. Furthermore, $s_t x_t$ are iid gamma with shape $n/2$ to which the GamConType1 distribution is conditionally conjugate.

When h_t is large the contribution of new arrivals tends to be small relative to survivors, so s_t takes values close to 0. On the other hand, when h_t takes small values, new arrivals contribute a large share to x_t and so s_t will be close to 1. In the special case of $h_t = 0$, no survivors exist and only immigration contributes to x_t , so s_t takes a value of 1. In particular,

$$\mathbb{E}\{s_t \mid h_t\} = \frac{n/2}{n/2 + h_t} = \frac{\mathbb{V}\text{ar}\{\Lambda_t\}}{\mathbb{V}\text{ar}\{\xi_t \mid h_t\} + \mathbb{V}\text{ar}\{\Lambda_t\}}$$

takes the same signal-over-total-variance form as the Kalman gain, cf. Chapter 13 of Hamilton (1994) or the Bühlmann credibility factor, cf. Bühlmann and Gisler (2005). Moreover, using $\mathbb{E}\{X_t \mid h_t\} = 2\tau(n/2 + h_t)$ and $\mathbb{E}\{\Lambda_t\} = n\tau$, the conditional mean of Λ_t satisfies the Kalman update,

$$\mathbb{E} \{ \Lambda_t \mid x_t, h_t \} = x_t \mathbb{E} \{ s_t \mid h_t \} = \mathbb{E} \{ \Lambda_t \} + \mathbb{E} \{ s_t \mid h_t \} (x_t - \mathbb{E} \{ X_t \mid h_t \})$$

Proposition 2 provides the posterior density of $\tilde{s}_T$ after integrating out the contribution of $\tilde{H}_T$.

Proposition 2 *Posterior density of $\tilde{s}_T$*

$$\begin{aligned} \pi \left(\tilde{s}_T \mid \tilde{x}_T, \rho, \tau, \frac{n}{2} \right) &= \prod_{t=1}^T \pi \left(s_t \mid x_t, x_{t-1}, \rho, \tau, \frac{n}{2} \right) \\ \pi \left(s_t \mid x_t, x_{t-1}, \rho, \tau, \frac{n}{2} \right) &= \gamma_t \delta_1(s_t) + f(s_t) \mathbb{I} \{ 0 < s_t < 1 \} \\ \gamma_t &= \frac{1}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)} \frac{\left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}-1}}{\Gamma \left(\frac{n}{2} \right)} \\ f(s_t) &= \frac{1}{\Gamma \left(\frac{n}{2} \right)} \frac{s_t^{\frac{n}{2}-1}}{\sqrt{1-s_t}} \left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}} \frac{\mathbf{I}_1 \left(\frac{\sqrt{\rho x_t x_{t-1}} (1-s_t)}{\tau} \right)}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)} \end{aligned}$$

where $\mathbf{I}_\nu(a)$ is the modified Bessel function of the first kind of order ν and $\delta_1(\cdot)$ is the Dirac delta at 1.

In terms of a birth-death process, γ_t is the probability that at time t there were no survivors. The posterior density of $\tilde{\xi}_T$ is given by a Bessel mixture of Gammas. After integrating out the $\tilde{H}_T$ component, the resulting posterior density exhibits the closed form expression of a type II randomized Gamma distribution. This distribution has long been studied, cf. Section 2.7 in Feller (1975), and this result has already been obtained in continuous time by Pitman and Yor (1982) in their study of *Squared Bessel bridges*, see also Chapter XI of Revuz and Yor (1999).

2.2 Smoothing and Missing Data

After estimation is carried out, an object of interest might be that of the value of the process at an unobserved time t_k , x_{t_k} . Here t_k is not a multiple of Δ , which is the time notation

we have been using, but rather a time that lies between k and $k + 1$, $t_k \in (k, k + 1)$, where k is a multiple of Δ . Hence, x_k and x_{k+1} are the nearest observed neighbors of x_{t_k} . Due to the fixed Δ we can treat x_{t_k} as missing at random. If the parameters were known, from the Markov property of the CIR process, all information about x_{t_k} is summarized in the nearest neighbors, $\pi(x_{t_k} | \tilde{x}_T, \theta) = \pi(x_{t_k} | x_k, x_{k+1}, \theta)$ which we shall refer to as the *smoothed* density of x_{t_k} . In practice, since the parameters are not known, we learn about them through the observed sample $\tilde{x}_T$ so that the posterior density of x_{t_k} is given by

$$\pi(x_{t_k} | \tilde{x}_T, \Delta) = \int_{\Theta} \pi(x_{t_k} | x_k, x_{k+1}, \theta) \pi(\theta | \tilde{x}_T, \Delta) d\theta$$

Sampling from this mixture can be achieved by first sampling θ from $\pi(\theta | \tilde{x}_T, \Delta)$ and given this drawn value of θ , sample x_{t_k} from $\pi(x_{t_k} | x_k, x_{k+1}, \theta)$. To this end we can employ the Gibbs sampler in Proposition 1. Sampling from $\pi(x_{t_k} | x_k, x_{k+1}, \theta)$ can be achieved by a one-to-one transformation between the CIR process and the Squared Bessel process, and known results for sampling from a Squared Bessel bridge.

The Squared Bessel process Y_s is the strong solution to

$$dY_s = \delta ds + 2\sqrt{Y_s} dB_s$$

where $Y_0 = y_0 > 0$, $\delta \geq 0$ and B_s is a standard one-dimensional Brownian motion. There exists a one-to-one relationship between the non-absorbing CIR process and the non-absorbing Squared Bessel process given by $Y_{s(t)} = e^{-bt} X_t$, where $s(t) = \frac{\sigma^2(1-e^{-bt})}{4b}$ when $b \neq 0$, $s(t) = \frac{\sigma^2 t}{4}$ when $b = 0$ and $\delta = \frac{4a}{\sigma^2}$, cf. Revuz and Yor (1999). The process Y_s conditional on $Y_0 = y_0$ and $Y_S = y_S$, where $0 < s < S$, is known as the Squared Bessel bridge and is studied in detail by Pitman and Yor (1982). The results presented in this section draw heavily from the sampling schemes proposed in Yuan and Kalbfleisch (2000), Devroye (2002) and Makarov and Glew (2010). The density of a Squared Bessel bridge is given by the following randomized gamma distribution of the second type,

$$\pi(Y_s = y_s \mid Y_0 = y_0, Y_S = y_S) = \frac{S}{2s(S-s)} \exp \left\{ -\frac{y_0(S-s)^2 + y_s S^2 + y_S s^2}{2sS(S-s)} \right\} \\ \times \frac{\mathbf{I}_{\delta/2-1}(\sqrt{y_0 y_s}/s) \mathbf{I}_{\delta/2-1}(\sqrt{y_s y_S}/(S-s))}{\mathbf{I}_{\delta/2-1}(\sqrt{y_0 y_S}/S)}$$

Sampling from this distribution can be achieved by first sampling

$$U_s \sim \mathcal{P} \left(\left(y_0 \frac{(S-s)}{s} + y_s \frac{s}{S-s} \right) \frac{1}{2S} \right) \quad \text{and} \quad Z_s \sim \mathcal{Bessel} \left(\delta/2 - 1, \frac{\sqrt{y_0 y_S}}{S} \right)$$

where $U_s \perp\!\!\!\perp Z_s$, and finally,

$$Y_s \mid U_s, Z_s \sim \mathcal{G} \left(U_s + 2Z_s + \delta/2, \frac{2s(S-s)}{S} \right)$$

Once Y_s is sampled we can obtain the equivalent draw from the smoothed density of x_{t_k} by applying the transformation from the Squared Bessel process back to the CIR process.

Smoothing and handling of missing data are closely related. Let $\tilde{x}_T^O$ denote the observed sample and $\tilde{x}_T^M$ the missing values. Assume w.l.o.g. that there are only two non-overlapping sequences of missing data, $\tilde{x}_{k+1:k+n-1} = (x_{k+1}, \dots, x_{k+n-1})$ and $\tilde{x}_{l+1:l+p-1} = (x_{l+1}, \dots, x_{l+p-1})$, missing at random, where $0 \leq k < l$, $n, p \geq 2$, $k+n \leq l$ and $l+p \leq T$. By the Markov property and the missing at random assumption the two sequences are independent conditional on their nearest neighbors,

$$\pi(\tilde{x}_{k+1:k+n-1}, \tilde{x}_{l+1:l+p-1} \mid \tilde{x}_T^O, \boldsymbol{\theta}) = \pi(\tilde{x}_{k+1:k+n-1} \mid x_k, x_{k+n}, \boldsymbol{\theta}) \times \pi(\tilde{x}_{l+1:l+p-1} \mid x_l, x_{l+p}, \boldsymbol{\theta})$$

The previous sampling scheme from the smoothed density of $\pi(x_{t_k} \mid x_k, x_{k+1}, \boldsymbol{\theta})$ is also valid to sample from $\pi(x_j \mid x_k, x_{k+n}, \boldsymbol{\theta})$, where $j \in \{k+1, k+2, \dots, k+n-1\}$, after an appropriate redefinition of Δ . Sampling from the joint distribution of $\tilde{x}_{k+1:k+n-1}$ can be done in many ways, decomposing the joint distribution into a product of individual bridges

by the law of total probability and the Markov property. We will employ a dyadic scheme. That is, first we sample the midpoint $k + \lfloor n/2 \rfloor$. Given the draw $x_{k+\lfloor n/2 \rfloor}$ we are left with two conditionally independent bridges from which we sample their respective midpoints and so on.

By the above procedure we can sample the missing data given the observed sample and the parameters. To deal with missing data in estimation, under the assumption of data missing at random we can augment the Gibbs sampler of Proposition 1 as,

1. Sample $\tilde{x}_T^{M(m+1)}$ from $\pi(\tilde{x}_T^M \mid \tilde{x}_T^O, \theta^{(m)})$.
2. Sample $(\tilde{h}_T^{(m+1)}, \tilde{s}_T^{(m+1)}, \theta^{(m+1)})$ from

$$\pi(\tilde{h}_T^{(m+1)}, \tilde{s}_T^{(m+1)}, \theta^{(m+1)} \mid \tilde{x}_T^{(m+1)} \equiv \{\tilde{x}_T^O, \tilde{x}_T^{M(m+1)}\}, \theta^{(m)}).$$

As an example from simulated data, see Figure 4 where the pointwise posterior mean, median and 30-68-95% credible intervals are reported for the missing data. We simulate a single trajectory of $\tilde{X}_T$ with $T = 500$, $\Delta = 0.2$, $a_0 = 0.4$, $b_0 = -0.2$, $\sigma_0^2 = 0.2$ and $X_0 = -a_0/b_0 = 2$. This implies a $\tau_0 = 0.0098$, $\rho_0 = 0.96$ and $n_0/2 = 4$. The priors for τ , ρ and $n/2$ are the same as in the simulated example of Section 2.1, where in addition the prior of ρ is truncated to be in the stationary region. Nevertheless, this approach still assumes that Δ is fixed, which in some settings might be an excessively restrictive assumption. In the following section we allow for irregular random time intervals.

2.3 Irregularly Spaced Observations

Let observations occur at irregular times $\{t_i\}_{i=1}^T$ with associated irregular increments $\Delta_{t_i} = t_i - t_{i-1}$, where $0 < \Delta_{t_i} < \infty \quad \forall i \geq 1$. The conditional distribution of the CIR process observed at irregular intervals is given by

$$\frac{-4b}{(1 - e^{b\Delta_{t_i}}) \sigma^2} X_{t_i} \mid \tilde{X}_{t_{i-1}} = \tilde{x}_{t_{i-1}} \sim \chi_{\frac{4a}{\sigma^2}}^2 \left(\frac{-4b}{\sigma^2} \frac{e^{b\Delta_{t_i}}}{1 - e^{b\Delta_{t_i}}} x_{t_{i-1}} \right).$$

with the case of $b = 0$ being defined by the continuity limit of the above expression. We propose the following one-to-one reparameterization,

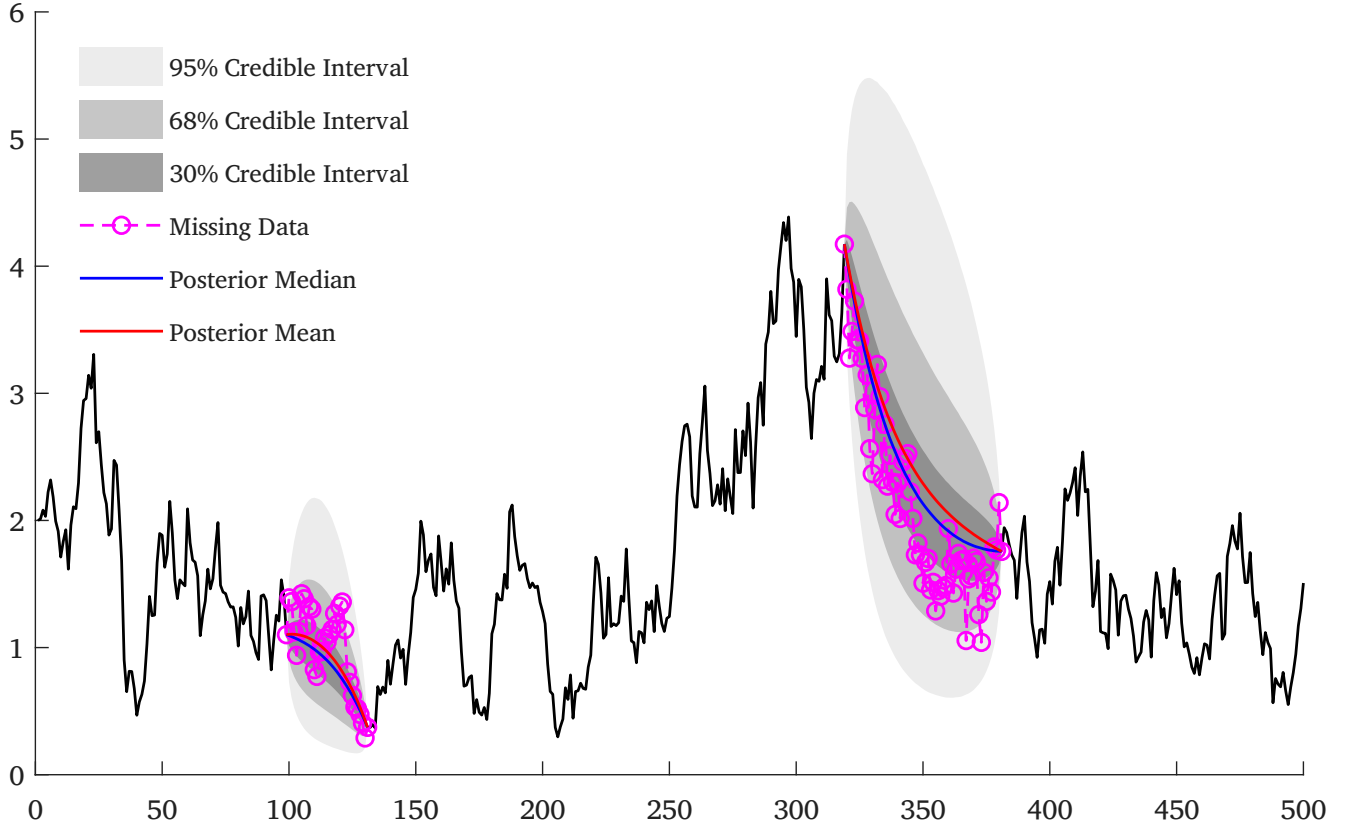


Figure 4: Bayesian estimation with data missing at random.

$$\begin{aligned}
 dX_t &= (0.4 - 0.2X_t) dt + \sqrt{0.2X_t} dW_t \\
 T &= 500, \quad X_0 = 2.5, \quad \Delta = 0.2, \\
 \underline{\alpha} = \underline{\beta} = \underline{\nu} &= 0.1, \quad \underline{\lambda} = 1, \quad \underline{a} = 1/3, \quad \underline{b} = 99 \quad \text{and} \quad \rho < 1.
 \end{aligned}$$

$$\rho = e^b; \quad n = \frac{4a}{\sigma^2}; \quad \tau = \begin{cases} \frac{\sigma^2(e^b - 1)}{4b} & b \neq 0 \\ \frac{\sigma^2}{4} & b = 0 \end{cases}$$

With this reparameterization the transition density of the CIR process observed at irregular intervals when $b \neq 0$ is

$$\frac{1 - \rho}{1 - (\rho)^{\Delta_{t_i}}} \frac{1}{\tau} x_{t_i} \mid \tilde{X}_{t_{i-1}} = \tilde{x}_{t_{i-1}} \sim \chi_n^2 \left(\frac{(\rho)^{\Delta_{t_i}}}{\tau} \frac{1 - \rho}{1 - (\rho)^{\Delta_{t_i}}} x_{t_{i-1}} \right)$$

and the case of $b = 0$ is obtained by the continuity limit of the above expression given by,

$$\frac{1}{\tau \Delta_{t_i}} X_{t_i} \mid \tilde{X}_{t_{i-1}} = \tilde{x}_{t_{i-1}} \sim \chi_n^2 \left(\frac{1}{\tau \Delta_{t_i}} x_{t_{i-1}} \right)$$

In state-space form,

$$\begin{aligned} X_{t_i} \mid \tilde{X}_{t_{i-1}} = \tilde{x}_{t_{i-1}}, \tilde{H}_{t_i} = \tilde{h}_{t_i} &\sim \mathcal{G} \left(\frac{n}{2} + h_{t_i}, 2\tau \frac{1 - (\rho)^{\Delta_{t_i}}}{1 - \rho} \right) \\ h_{t_i} \mid \tilde{X}_{t_{i-1}} = \tilde{x}_{t_{i-1}}, \tilde{H}_{t_{i-1}} = \tilde{h}_{t_{i-1}} &\sim \mathcal{P} \left(\frac{(\rho)^{\Delta_{t_i}}}{2\tau} \frac{1 - \rho}{1 - (\rho)^{\Delta_{t_i}}} x_{t_{i-1}} \right) \end{aligned}$$

Instead of observing a single Δ , the econometrician now observes a sequence of time increments $\{\Delta_{t_i}\}_{i=1}^T$. Therefore, we need to make some assumption related to this sequence. We will impose that time increments are *non-anticipative* and *non-informative*, which is formalized in Assumption 1. We also impose in Assumption 2 that the posterior of ϕ , the parameters governing the transition density of Δ_{t_i} , is proper.

Assumption 1 $\tilde{\Delta}_{t_T}$ are *non-anticipative* and *non-informative*.

$$\forall i \geq 1 : \pi \left(\Delta_{t_i} \mid \tilde{x}_{t_{i-1}}, \tilde{\Delta}_{t_{i-1}}, \theta, \phi \right) = \pi \left(\Delta_{t_i} \mid \tilde{x}_{t_{i-1}}, \tilde{\Delta}_{t_{i-1}}, \phi \right)$$

where $\theta = \{\tau, \rho, \frac{n}{2}\}$ and $\theta \cap \phi = \emptyset$

Assumption 2 The posterior density of ϕ , $\pi \left(\phi \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T} \right)$, is proper.

Non-anticipative implies that the sampling intervals are independent of present and future x 's given past x 's. The non-informative part of the assumption implies that the probability laws of time increments do not involve the parameters of interest θ . This assumption is also known as $\tilde{\Delta}_{t_T}$ being *sequentially exogenous* for θ , cf. Gouriéroux and Monfort (1995) Definition 1.11.

Assumption 1 allows for time increments depending on past x 's, reflecting phenomena such as stop-loss triggered trading. The type of phenomena this assumption rules out is a contemporaneous impact on trading times or anticipatory behaviors such as insider trading. It also rules out dynamics of trading times whose transition probability laws contain

information on θ , say a process whereby trading frequency is increasing as a function of past $|x|$ and whose parametrization involves θ . If the parametrization does not involve θ such a process would be allowed. Under Assumptions 1 and 2 and a particular set of priors we obtain the following conditional posterior distributions, the irregular equivalent to Proposition 1.

Proposition 3 *Unfeasible Gibbs Sampler for Irregularly Spaced Observations*

Assume Assumptions 1 and 2 hold and priors take the form

$$\begin{aligned}\pi(\boldsymbol{\theta}, \boldsymbol{\phi} \mid x_{t_0}) &= \pi(\boldsymbol{\theta} \mid x_{t_0}) \pi(\boldsymbol{\phi} \mid x_{t_0}) \\ \pi(\boldsymbol{\theta} \mid x_{t_0}) &= \pi(\tau \mid x_{t_0}) \pi(\rho \mid x_{t_0}) \pi\left(\frac{n}{2} \mid x_{t_0}\right) \\ \tau \mid x_{t_0} &\sim \mathcal{IG}(\underline{\alpha}, \underline{\beta}) \\ \frac{n}{2} \mid x_{t_0} &\sim \mathcal{GC1}(\underline{\lambda}, \underline{\nu})\end{aligned}$$

where $\underline{\alpha}, \underline{\beta}, \underline{\lambda}, \underline{\nu} > 0$.

Then,

$$h_{t_i} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tau, \rho, \frac{n}{2} \sim \mathcal{Bessel}\left(\frac{n}{2} - 1, \frac{1 - \rho}{1 - \rho^{\Delta_{t_i}}} \frac{\sqrt{\rho^{\Delta_{t_i}} x_{t_i} x_{t_{i-1}}}}{\tau}\right)$$

$$s_{t_i} \mid \tilde{h}_{t_T}, \frac{n}{2} \sim \mathcal{Beta}\left(\frac{n}{2}, h_{t_i}\right), \quad \text{and}$$

$$\tau \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}, \rho, \frac{n}{2} \sim \mathcal{IG}(\bar{\alpha}, \bar{\beta})$$

$$\text{where } \bar{\alpha} = \underline{\alpha} + \frac{n}{2}T + 2 \sum_{i=1}^T h_{t_i}, \quad \bar{\beta} = \underline{\beta} + \frac{1 - \rho}{2} \sum_{i=1}^T \frac{x_{t_i} + \rho^{\Delta_{t_i}} x_{t_{i-1}}}{1 - \rho^{\Delta_{t_i}}}, \quad \text{and}$$

$$\begin{aligned}\pi\left(\rho \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}, \tau, \frac{n}{2}\right) &\propto \pi(\rho \mid x_{t_0}) \times \left[\prod_{i=1}^T \left(\frac{1 - \rho}{1 - \rho^{\Delta_{t_i}}} \right)^{\frac{n}{2} + 2h_{t_i}} \right] \rho^{\sum_{i=1}^T \Delta_{t_i} h_{t_i}} \\ &\times \exp \left\{ -\frac{1 - \rho}{2\tau} \sum_{i=1}^T \frac{x_{t_i} + \rho^{\Delta_{t_i}} x_{t_{i-1}}}{1 - \rho^{\Delta_{t_i}}} \right\}\end{aligned}$$

$$\text{with } \frac{n}{2} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{s}_{t_T}, \tau, \rho \sim \mathcal{GC1}(\bar{\lambda}, \bar{\nu})$$

$$\text{where } \bar{\lambda} = \underline{\lambda}^{\nu/\bar{\nu}} \left[\left(\prod_{i=1}^T \frac{1 - \rho}{1 - \rho^{\Delta_{t_i}}} \frac{x_{t_i} s_{t_i}}{2\tau} \right)^{1/T} \right]^{T/\bar{\nu}} \quad \text{and } \bar{\nu} = \underline{\nu} + T$$

We call the Gibbs sampler in Proposition 3 unfeasible because the density of the conditional posterior of ρ is not from a known or easy to sample distribution. Therefore, this step will require the use of techniques such as a Metropolis-Hastings (MH) step.

To sample from $\pi\left(\rho \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}, \tau, \frac{n}{2}\right)$ we suggest employing a tailored independent MH step. Tailored independent MH steps have been employed successfully in the literature to deal with highly irregular densities, cf. Chib and Ergashev (2009) and Hiraki et al. (2025) where tailored Gaussian proposals are built. Since in the equidistant case the conditional posterior is Gamma, we find it appropriate to build a tailored Gamma proposal density. At the $(m+1)$ iteration for ρ we will first find a mode of the conditional posterior density, $\hat{\rho}$, and calculate the negative inverse Hessian evaluated at the mode,

$$\hat{\rho} = \arg \max_{\rho} \ln \pi\left(\rho \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}^{(m+1)}, \tau^{(m+1)}, (n/2)^{(m)}\right)$$

$$\mathbf{V}_{\rho} = \left(- \frac{\partial^2 \ln \pi\left(\rho \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}^{(m+1)}, \tau^{(m+1)}, (n/2)^{(m)}\right)}{\partial \rho^2} \right)^{-1} \Big|_{\rho=\hat{\rho}}$$

Given the current value $\rho^{(m)}$, generate a candidate $\rho^{\dagger}$ from $\mathcal{G}(1 + \hat{\rho}^2/\mathbf{V}_{\rho}, \mathbf{V}_{\rho}/\hat{\rho})$ and set $\rho^{(m+1)} = \rho^{\dagger}$ with probability

$$\alpha(\rho^{(m)}, \rho^{\dagger}) = \min \left\{ 1, \frac{\pi\left(\rho^{\dagger} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}^{(m+1)}, \tau^{(m+1)}, (n/2)^{(m)}\right) f_{\mathcal{G}}(\rho^{(m)} | 1 + \hat{\rho}^2/\mathbf{V}_{\rho}, \mathbf{V}_{\rho}/\hat{\rho})}{\pi\left(\rho^{(m)} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}^{(m+1)}, \tau^{(m+1)}, (n/2)^{(m)}\right) f_{\mathcal{G}}(\rho^{\dagger} | 1 + \hat{\rho}^2/\mathbf{V}_{\rho}, \mathbf{V}_{\rho}/\hat{\rho})} \right\}$$

where $f_{\mathcal{G}}$ is a Gamma density. Otherwise, set $\rho^{(m+1)} = \rho^{(m)}$. The gamma proposal parametrization is set to match the mode of the conditional posterior density and the local curvature at the mode. In practice, finding a mode is not strictly necessary and a high density point suffices to build a successful proposal distribution.

We now summarize the steps of the Metropolis-within-Gibbs Sampler. Begin with an arbitrary value $\theta^{(0)} \in \Theta$. Then,

1. Sample $(\tilde{s}_{t_T}^{(m+1)}, \tilde{h}_{t_T}^{(m+1)})$ from $\pi(\tilde{s}_{t_T} \mid \tilde{h}_{t_T}, (n/2)^{(m)}) \times \pi(\tilde{h}_{t_T} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tau^{(m)}, \rho^{(m)}, (n/2)^{(m)})$.

2. Sample $\tau^{(m+1)}$ from $\pi\left(\tau \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}^{(m+1)}, \rho^{(m)}, (n/2)^{(m)}\right)$.
3. Sample $\rho^{(m+1)}$ via the tailored independent MH step.
4. Sample $(n/2)^{(m+1)}$ from $\pi\left(\frac{n}{2} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{s}_{t_T}^{(m+1)}, \tau^{(m+1)}, \rho^{(m+1)}\right)$.

Finally, we illustrate the above scheme with a simulated example. To model random time increments, $\tilde{\Delta}_{t_T}$, we will employ the logarithmic Auto-regressive Conditional Duration model (log-ACD₂) of Bauwens and Giot (2000). The log-ACD₂ model is given by the following three equations,

$$\begin{aligned}\Delta_{t_i} &= \frac{\psi_i}{\Gamma(1 + 1/\gamma)} \epsilon_i \\ \ln \psi_i &= \omega + \alpha \epsilon_{i-1} + \beta \ln \psi_{i-1} \\ \epsilon_i &\sim \mathcal{W}eibull(1, \gamma)\end{aligned}$$

where $\epsilon_i \sim \mathcal{W}eibull(1, \gamma)$ has mean $\Gamma(1 + 1/\gamma)$, so that ψ_i is the conditional expected duration and the lagged excess duration entering the recursion is $\epsilon_{i-1} = \Gamma(1 + 1/\gamma) \Delta_{t_{i-1}} / \psi_{i-1}$. A simulated series for $\tilde{\Delta}_{t_T}$ and the CIR process is plotted in Figure 5. The parameters of the log-ACD₂ model are taken from the estimates for Boeing by Bauwens and Giot (2000): $\omega = -0.104$, $\alpha = 0.095$, $\beta = 0.953$ and $\gamma = 0.905$. The parameters of the CIR process are chosen such that the model is ergodic but highly persistent and the Feller condition does not hold: $a_0 = 0.38$, $b_0 = -0.1$ and $\sigma_0^2 = 0.8$. This implies values of $\tau_0 = 0.1903$, $\rho_0 = 0.9048$ and $\frac{n_0}{2} = 0.95 < 1$. When $0 < \frac{n_0}{2} < 1$ and $\rho_0 < 1$, the CIR process exhibits explosive bursts from and to the origin.

We simulate from the posterior distribution of θ employing the Metropolis-within-Gibbs sampler described above for irregularly spaced observations at samples of increasing size, $T = 100, 400$ and 1600 . We iterate MCMC simulation 1 million times after a burn-in period of 30 thousand MCMC draws. The priors for τ , ρ and $n/2$ are the same as in the simulated example of Section 2.1. The prior of ρ is not truncated to be below 1. The posterior distribution of the parameters at each sample size is shown in Figure 6, where we can observe the posterior concentrating close to the true values as the sample size increases. The acceptance rate for the tailored MH step of ρ is close to 100% across all sample sizes.

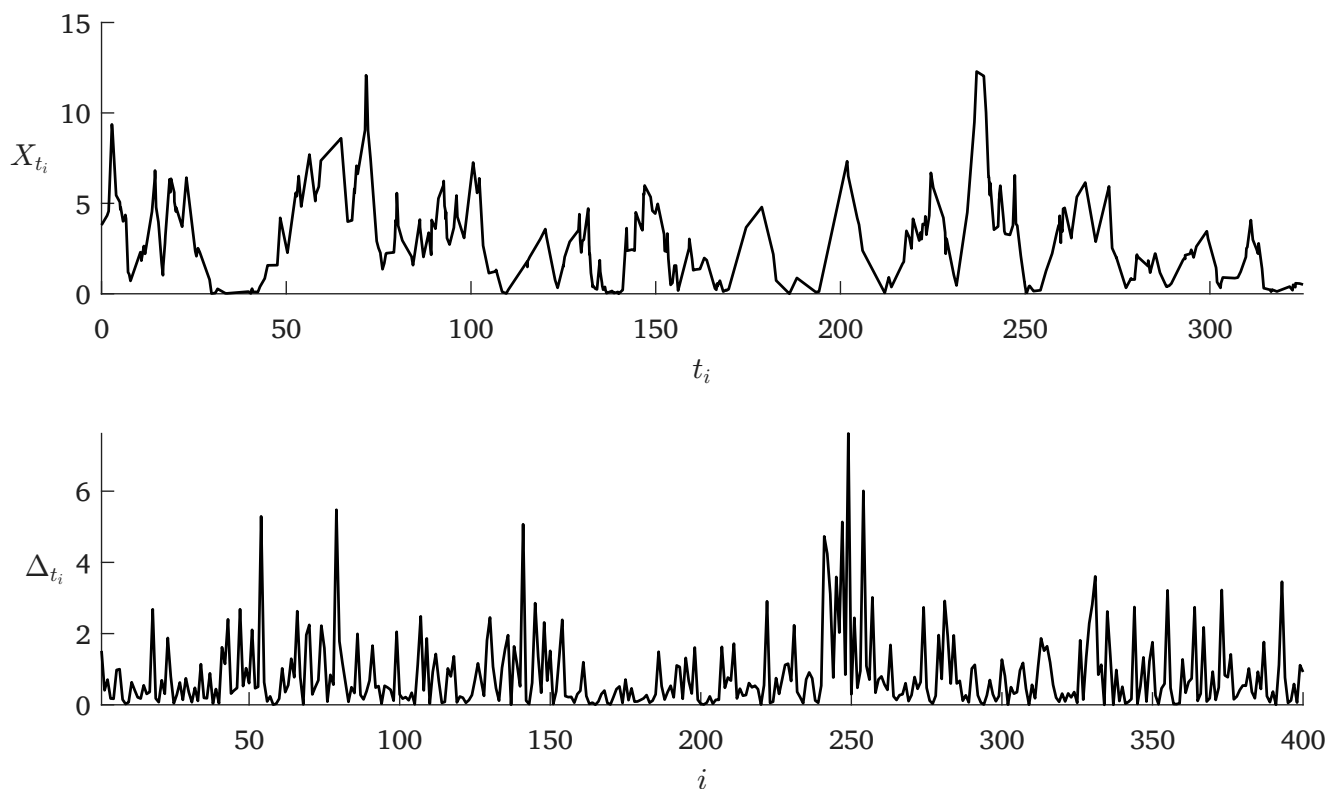


Figure 5: Simulated sample from a log-ACD₂ model and a CIR process. $T = 400$.

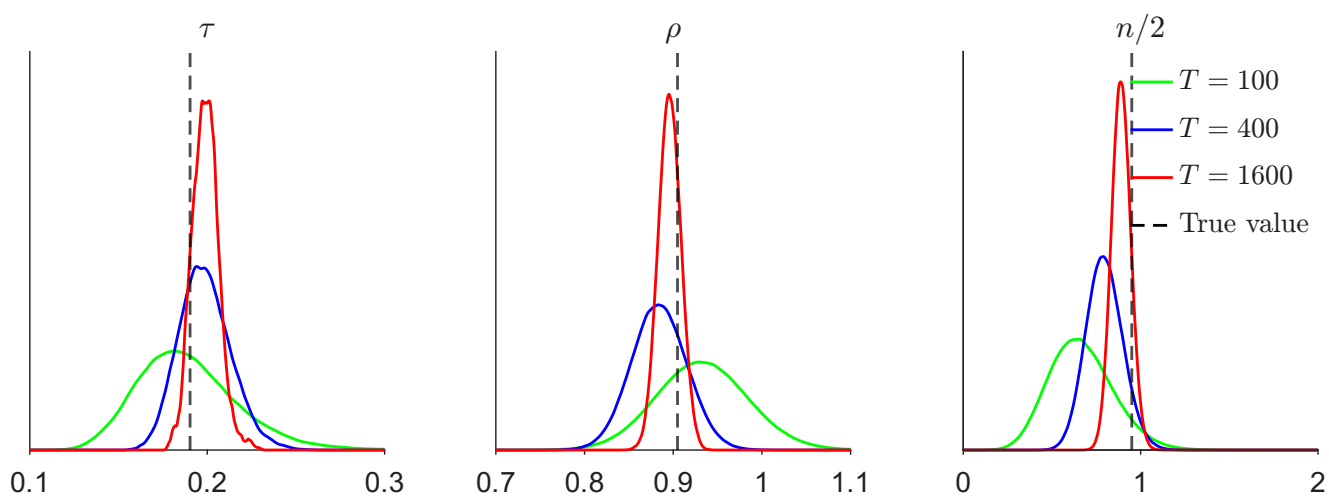


Figure 6: Marginal Posterior Distributions of θ

3 Large-Sample Theory and the Frequentist-Bayesian Bridge

This section is devoted to the analysis of asymptotic properties. These complement the Bayesian analysis in Section 2 by concluding with a Bernstein-von Mises theorem, which reconciles frequentist and Bayesian inference. We will restrict the parameter space to be that of the stationary ergodic region, $\Theta^* \subset \Theta$ and the process to be observed at equidistant times. That is, $\Theta^* = \Theta_\tau^* \times \Theta_\rho^* \times \Theta_{n/2}^* = (0, \infty) \times (0, 1) \times (0, \infty)$ and $0 < \Delta < \infty$ fixed.

The first property we will show is Local Asymptotic Normality (LAN). Let $\pi(x_1, \dots, x_T \mid \theta, x_0)$ denote the joint density of $X_1, \dots, X_T$ given $X_0 = x_0$ under θ and consider the log-likelihood ratio of local alternatives around the true parameter $\theta_0 \in \Theta^*$. That is, $\theta_0 + \delta_T \mathbf{u} \in \Theta^*$; where $\delta_T \rightarrow 0$ is the normalizing factor and $\mathbf{u} \in \mathbb{R}^3$ the direction and intensity of the deviation.

$$\Lambda(\theta_0 + \delta_T \mathbf{u}, \theta_0 \mid x_0) \equiv \ln \frac{\pi(X_1, \dots, X_T \mid \theta_0 + \delta_T \mathbf{u}, x_0)}{\pi(X_1, \dots, X_T \mid \theta_0, x_0)}$$

By the Markov property,

$$\Lambda(\theta_0 + \delta_T \mathbf{u}, \theta_0 \mid x_0) = \sum_{t=1}^T \ln \pi(X_t \mid X_{t-1}, \theta_0 + \delta_T \mathbf{u}) - \ln \pi(X_t \mid X_{t-1}, \theta_0)$$

The experiment is said to be LAN (cf. Ibragimov and Hasminskii (1981) Definition 2.1 and Van der Vaart (1998) Definition 7.14) if for some normalizing factor δ_T , and each $\mathbf{u} \in \mathbb{R}^3$,

$$\Lambda(\theta_0 + \delta_T \mathbf{u}, \theta_0 \mid x_0) = \mathbf{u}^\top \mathbf{S}_T - \frac{1}{2} \mathbf{u}^\top \mathbf{K}_T \mathbf{u} + o_{P_{\theta_0, x_0}}(1)$$

for some sequences of random vectors $\mathbf{S}_T$ and symmetric random matrices $\mathbf{K}_T$ s.t. $\mathbf{K}_T$ converges in probability to a finite, positive semi-definite, non-random matrix $\mathbf{K}_{\theta_0}$, the Fisher information matrix, and $\mathbf{S}_T$ converges P_{θ_0, x_0} -weakly to $\mathcal{N}(0, \mathbf{K}_{\theta_0})$.

Overbeck and Rydén (1997) show LAN of the CIR statistical experiment under stationarity and the Feller condition being satisfied. In Theorem 2 we are able to drop the assumption

of the Feller condition being satisfied. Moreover, we show the Fisher information matrix to be finite and non-singular.

Theorem 2 *Local Asymptotic Normality*

Assume $0 < \Delta < \infty$ and $X_0 = x_0$, fixed. The statistical experiment is LAN at $\theta_0 \in \Theta^*$ with normalizing factor $\delta_T = T^{-1/2}$ and asymptotic information matrix $\mathbf{K}_{\theta_0} = (\mathbf{K})_{ij}$ given by,

$$(\mathbf{K})_{ij} = \mathbb{E}_{\theta_0} \left\{ \frac{\partial}{\partial \theta_i} \ln \pi(X_1^* | X_0^*, \theta_0) \times \frac{\partial}{\partial \theta_j} \ln \pi(X_1^* | X_0^*, \theta_0) \right\}$$

where (X_1^*, X_0^*) corresponds to a generic consecutive pair and expectations, $\mathbb{E}_{\theta} \{\cdot\}$, are taken w.r.t the invariant density of (X_1^*, X_0^*) . Furthermore, $\mathbf{K}_{\theta_0}$ is finite and non-singular.

The main consequence of LAN and a finite, non-singular Fisher information matrix is that regular estimators cannot converge to θ_0 at a rate faster than $1/\sqrt{T}$, and among those that converge at a $1/\sqrt{T}$ rate, their asymptotic variance is bounded below by the Cramér-Rao lower bound.

Theorem 2 also allows us to answer an open question posed by Overbeck and Rydén (1997) regarding *superefficiency*. When the Feller condition does not hold, under continuous sampling one recovers the true parameter a_0 in finite time. When the process hits 0 it becomes deterministic with a linear trend $a_0 dt$ during an instant of time. Continuous sampling guarantees that we observe the process hitting 0 and its immediate deterministic increment from which a_0 can be perfectly recovered. In the case of observations recorded at discrete equidistant times, none will take a value of 0 with probability one. This reveals superefficiency to be a particular feature of continuous sampling that does not carry over to discrete samples.

Building on the LAN result, Theorem 3 shows consistency and asymptotic normality of maximum likelihood estimation.

Theorem 3 Consistency and Asymptotic Normality of Conditional Maximum Likelihood

Assume

1. $0 < \Delta < \infty$ and $X_0 = x_0$ fixed,
2. The parameter space $\Theta^\dagger$ is a compact subset of Θ^* .
3. $\theta_0 \in \text{int}(\Theta^\dagger)$

Define $\hat{\theta}_T = \arg \max_{\theta \in \Theta^\dagger} \ln \pi(\tilde{x}_T | \theta, x_0)$. Then,

- i. $\hat{\theta}_T \xrightarrow{P_{\theta_0, x_0}} \theta_0$,
- ii. $\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow \mathcal{N}(0, \mathbf{K}_{\theta_0}^{-1})$ P_{θ_0, x_0} -weakly.

Finally, we present a reconciliation between classical and Bayesian inference, a *Bernstein-von Mises* theorem, stated in terms of convergence to the conditional maximum likelihood estimator in view of the previous theorem.

Theorem 4 Bernstein-von Mises

Assume assumptions 1, 2 and 3 hold, and

4. The prior distribution of θ , $\pi(\theta | \Delta, x_0)$, is absolutely continuous in a neighborhood of θ_0 with a continuous positive density at θ_0 .

Define $\Pi(\cdot | \tilde{X}_T)$ to be the cumulative distribution function (CDF) of the posterior distribution and $F_{\mathcal{N}}(\cdot; \hat{\theta}_T, \frac{1}{T}\mathbf{K}_{\theta_0}^{-1})$ a normal CDF centered at the maximum likelihood estimate with variance equal to the Cramér-Rao lower bound. Then,

$$\left\| \Pi(\cdot | \tilde{X}_T) - F_{\mathcal{N}}\left(\cdot; \hat{\theta}_T, \frac{1}{T}\mathbf{K}_{\theta_0}^{-1}\right) \right\|_{TV} \xrightarrow{P_{\theta_0, x_0}} 0$$

where $\|\cdot\|_{TV}$ denotes the total variation distance.

The family of priors defined in Proposition 1 satisfies assumption 4 of Theorem 4.



Figure 7: 3-Year U.S. Treasuries, Wednesdays

4 An application to 3 year U.S. Treasury yields

We illustrate the methods developed in this paper with an application to 3-Year U.S. Treasury yields.² Our sample covers 20 years, the period from March 29th, 2006 to March 25th, 2026 at a weekly frequency, where we take the daily yield recorded on Wednesdays.³ The sample comprises 1044 observations; Figure 7 depicts the data in basis points.

Table 2 reports posterior summary statistics of the CIR process estimated under a non-informative prior: the posterior mean, median, standard deviation and the 95% credible interval. The non-informative prior corresponds to a proper but wide prior for τ and $n/2$ ($\underline{\alpha} = \underline{\beta} = \underline{\nu} = 0.1$ and $\underline{\lambda} = 1$) and for ρ a proper prior close to the neutral prior of Kerman (2011) ($\underline{a} = 1/3$ and $\underline{b} = 99$), which we have already employed in the numerical examples of Section 2. Results are presented for both the discrete-time θ parametrization and the continuous-time η parametrization, where the latter is recovered using a time increment of $\Delta = 1/52$, corresponding to one week expressed as a fraction of a year. The table also

²FRED Code: DGS3.

³Since the average of weekly observations from a CIR process no longer follows a CIR process, we select a single day, Wednesday, to construct the weekly series. This also reduces the sample size and avoids the treatment of missing data, alleviating the computational burden.

reports the posterior probability of ρ lying in the stationary region, the posterior probability that the Feller condition is satisfied and the marginal likelihood. Figure 8 depicts the corresponding marginal posterior distributions.

Table 2: CIR Process Posterior Summary Statistics

$$dX_t = (a + bX_t) dt + \sigma\sqrt{X_t} dW_t$$

	Mean	Median	Std. Dev.	95% C.I.
<i>Discrete-time parametrization (θ)</i>				
τ	0.182	0.182	0.008	(0.167, 0.198)
ρ	0.996	0.996	0.002	(0.991, 1.001)
$n/2$	1.974	1.944	0.861	(0.382, 3.742)
<i>Continuous-time parametrization (η)</i>				
σ^2	37.979	37.917	1.718	(34.746, 41.364)
b	-0.207	-0.206	0.126	(-0.458, 0.035)
a	37.493	36.872	16.441	(7.173, 71.222)
<i>Posterior model probabilities</i>				
$\mathbb{P}(\rho < 1 \tilde{x}_T) = \mathbb{P}(b < 0 \tilde{x}_T) = 95.20\%$				
$\mathbb{P}(n/2 \geq 1 \tilde{x}_T) = \mathbb{P}(2a/\sigma^2 \geq 1 \tilde{x}_T) = 86.70\%$				
<i>Marginal Likelihood</i>				
$\ln \pi(\tilde{X}_T) = -3915.436$				

Notes: The data are weekly 3-Year U.S. Treasury yields (Wednesdays), expressed in basis points, covering the period from March 29th, 2006 to March 25th, 2026 ($T = 1043$). Posterior distributions are based on 3.5 million MCMC draws after a burn-in of 150 thousand. The prior is non-informative: $\underline{\alpha} = \underline{\beta} = \underline{\nu} = 0.1$, $\underline{a} = 1/3$, $\underline{b} = 99$ and $\underline{\lambda} = 1$. The 95% credible interval is computed as the 2.5th and 97.5th posterior percentiles. The η parametrization is recovered via $b = \ln(\rho)/\Delta$, $\sigma^2 = 4\tau b/(\rho - 1)$ and $a = (n/2)\sigma^2/2$, with $\Delta = 1/52$.

The results in Table 2 reveal an identification hierarchy. The scale parameters τ and σ^2 are sharply identified, with narrow, nearly symmetric posteriors and a coefficient of variation of about 4.5%. The tight identification of σ^2 reflects the fact that the diffusion coefficient is determined by the local quadratic variation $d\langle X \rangle_t = \sigma^2 X_t dt$, which does not depend on the drift parameters. Under continuous sampling, σ^2 can be recovered exactly from a single path; at a fixed discrete frequency a similar mechanism operates, so that even moderate sample sizes pin down σ^2 with high precision. Since $\tau \approx \sigma^2 \Delta/4$ for small Δ , the scale parameter τ inherits the same sharp identification.

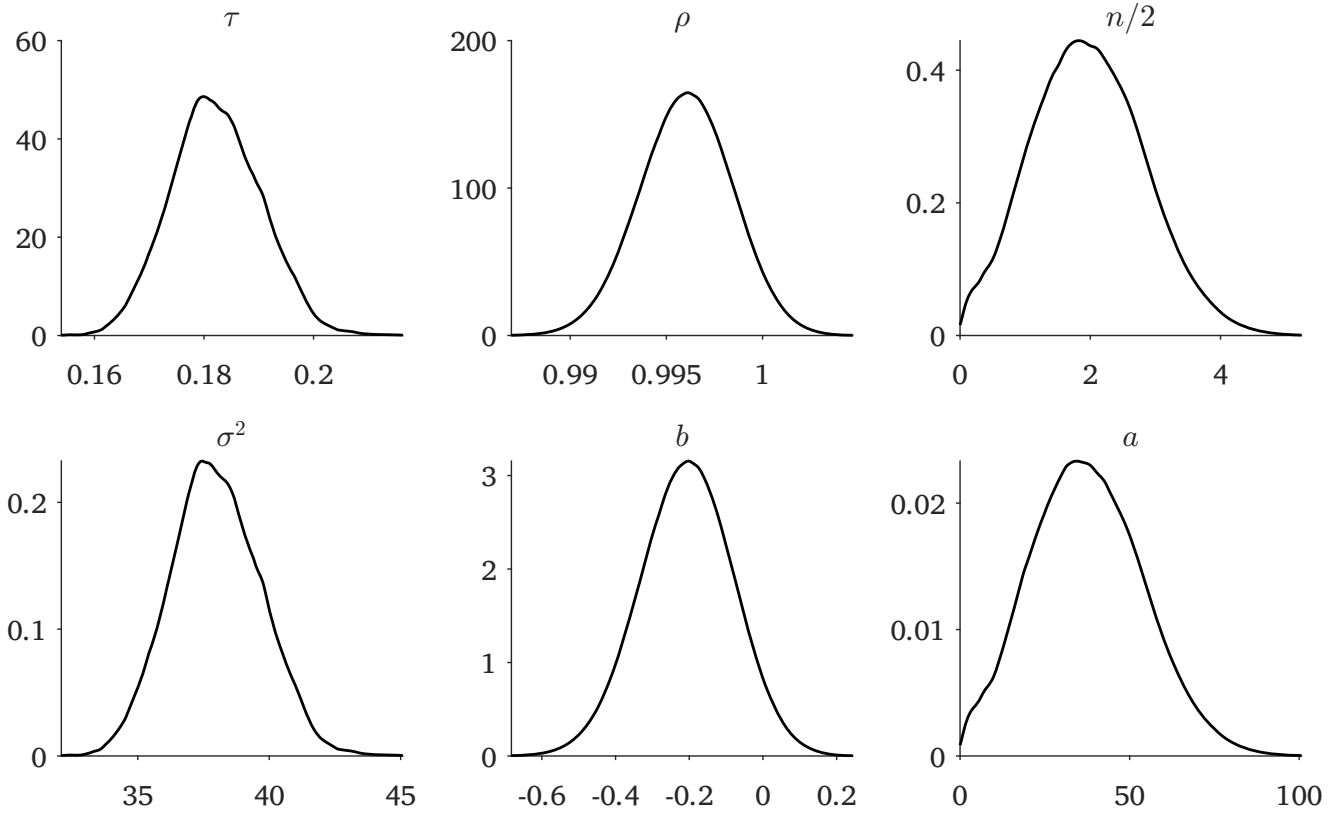


Figure 8: CIR Process Marginal Posterior Distributions

Notes: The data are weekly 3-Year U.S. Treasury yields (Wednesdays), expressed in basis points, covering the period from March 29th, 2006 to March 25th, 2026 ($T = 1043$). Posterior distributions are based on 3.5 million MCMC draws after a burn-in of 150 thousand. The prior is non-informative: $\underline{\alpha} = \underline{\beta} = \underline{\nu} = 0.1$, $\underline{a} = 1/3$, $\underline{b} = 99$ and $\underline{\lambda} = 1$. The η parametrization is recovered via $b = \log(\rho)/\Delta$, $\sigma^2 = 4\tau b/(\rho - 1)$ and $a = (n/2)\sigma^2/2$, with $\Delta = 1/52$.

The process exhibits near unit-root persistence as ρ centers on 0.9960 with a tight posterior whose 95% credible interval nonetheless contains unity; equivalently, the credible interval for the continuous-time drift b includes zero. The posterior probability of ρ lying in the stationary region is high, at 95.20%, but this result should be taken with care as we employed an almost flat prior on an unbounded space, $(0, \infty)$. In the non-stationary region, $[1, \infty)$, all values of ρ are given a similar weight, which include very large values for which the process would quickly explode. Therefore, it is only reasonable to consider a small range of values equal to or above 1 in the non-stationary region. A flat prior for ρ does precisely the opposite, and the neutral prior of Kerman (2011) shares this behavior by giving a higher mass to values in the stationary region and being almost flat over the unbounded non-stationary region.

The least well-identified parameter is the Feller index $n/2$, whose credible interval spans

nearly an order of magnitude. Because $a = (n/2) \sigma^2/2$ and σ^2 is tight, the uncertainty in a is essentially a scaled copy of the uncertainty in $n/2$. The posterior probability that the Feller condition is *not* met stands at 13.30%, assigning non-trivial weight to a reflecting-barrier scenario that most of the literature rules out by assumption.

The same issues with flat, unbounded priors affect the assessment of the Feller condition. In the context of testing for a unit root in an AR(1) model with normal innovations, Phillips (1991) documents a bias towards stationarity from employing flat and unbounded priors. Phillips (1991) and Berger and Yang (1994) proposed the use of *objective* Bayesian analysis, the former employing *Jeffreys* prior and the latter proposing a modified *reference* prior. We leave an objective Bayesian analysis, treatment of the initial condition or robust Bayes factors, all part of the Bayesian literature on unit roots, to future work. Instead, we perform a sensitivity analysis by employing different and easily implementable priors for θ .

We consider two alternative priors. Both are calibrated from a pre-sample period, April 9th, 1986 to March 29th, 2006 ($T = 1042$), in which we estimate η by conditional least squares following Overbeck and Rydén (1997). The pre-sample estimates are $\hat{\sigma}^2 = 18.3549$, $\hat{b} = -0.159$ and $\hat{a} = 80.2179$, implying $\hat{\tau} = 0.0881$, $\hat{\rho} = 0.9969$ and $\widehat{n/2} = 8.7407$.

The first alternative is a *weakly informative* prior within the family of priors in Proposition 1, with prior modes centered at the pre-sample estimates. For τ and ρ , we set the prior standard deviations equal to five times their respective modes, yielding priors whose information content is small relative to the nominal sample size T . For $n/2$, however, the naive criterion $\underline{\nu} \ll T$ is inadequate. The parameter $n/2$ enters the likelihood only through the immigration component of the birth-death-immigration decomposition, whose share of the observed level is $1 - \rho$ in stationarity. The rate at which the data inform the posterior of $n/2$ is therefore governed by $\approx (1 - \rho) \times T$ rather than T itself. Substituting the pre-sample estimate $\hat{\rho} = 0.9969$ and $T = 1042$ gives a data effective sample size of $(1 - \hat{\rho}) \times T = 0.0031 \times 1042 \approx 3.23$ observations. We set $\underline{\nu} = 1.5$, approximately half of this data effective sample size, to deliver regularization genuinely weaker than the data. The resulting hyperparameters are $\underline{\alpha} = 2.2662$, $\underline{\beta} = 0.2878$, $\underline{a} = 1.2210$, $\underline{b} = 4.5111$, $\underline{\lambda} = 8.2458$ and $\underline{\nu} = 1.5$.

The second alternative is a *spike-and-slab* prior that places a point mass value at the unit root. We keep the weakly informative prior for τ and $n/2$ but replace the prior of ρ with a two-component mixture: with probability $\underline{p}$, ρ takes a point mass value at 1, and with

probability $1 - \underline{p}$, ρ is drawn from a uniform distribution on $(0, 1)$. We set $\underline{p} = 0.5$ so that *a priori* both the unit root and stationarity are equally likely. Appendix B details the corresponding Gibbs sampling step. In the context of an AR(1) with normal innovations, Sims (1988) and Schotman and van Dijk (1991) employed priors placing a point mass at $\rho = 1$ and a uniform prior over a stationary region with the objective of testing for unit-roots.

For each of the three priors considered, including the non-informative prior, Table 3 reports the posterior mean and 95% credible intervals, along with the posterior probabilities of ρ lying in the stationary region and $n/2$ satisfying the Feller condition.

Parameters ρ and τ exhibit minor changes across prior specifications. The posterior probability of ρ belonging to the stationary region is robust to employing the proper weakly informative prior, but falls apart once we assign a positive mass to the unit-root hypothesis as in the spike-and-slab prior. $n/2$ is also sensitive to minor prior changes: the posterior mean increases substantially and the probability that the Feller condition is not satisfied becomes negligible.

We conjecture that the sensitivity of $n/2$ to the prior is related to the near unit-root behavior. Ben Alaya and Kebaier (2013) Theorems 6 and 9 show that when $b = 0$ and the Feller condition is satisfied, the speed of convergence of the maximum likelihood estimator of a is $\sqrt{\log T}$ under continuous sampling and $\sqrt{\log t_n}$ under double asymptotics, in comparison with the standard $\sqrt{T}$ results obtained in Section 3 and equivalent $\sqrt{T}$ convergence in continuous sampling for the stationary case. Coupled with a prior centered at $\widehat{n/2} = 8.7407$, even a small prior weight of $\underline{\nu} = 1.5$ can have a sizable effect on the posterior.

Table 3: CIR Posterior Summary Statistics under Alternative Priors

	Non-informative	Weakly informative	Spike and slab
<i>Discrete-time parametrization (θ)</i>			
τ	0.182 (0.167, 0.198)	0.181 (0.167, 0.198)	0.181 (0.166, 0.198)
ρ	0.996 (0.991, 1.001)	0.994 (0.989, 0.998)	0.998 (0.991, 1.000)
$n/2$	1.974 (0.382, 3.742)	3.397 (1.932, 4.929)	2.458 (1.120, 4.361)
<i>Continuous-time parametrization (η)</i>			
σ^2	37.979 (34.746, 41.364)	37.844 (34.753, 41.234)	37.750 (34.621, 41.153)
b	-0.207 (-0.458, 0.035)	-0.339 (-0.576, -0.105)	-0.081 (-0.489, 0.000)
a	37.493 (7.173, 71.222)	64.230 (36.346, 93.441)	46.380 (21.115, 82.571)
<i>Posterior model probabilities</i>			
$\mathbb{P}(\rho < 1 \mid \tilde{x}_T)$	95.20%	99.78%	24.03%
$\mathbb{P}(n/2 \geq 1 \mid \tilde{x}_T)$	86.70%	99.97%	98.52%

Notes: The data are weekly 3-Year U.S. Treasury yields (Wednesdays), expressed in basis points, covering the period from March 29th, 2006 to March 25th, 2026 ($T = 1043$). Posterior distributions are based on 3.5 million MCMC draws after a burn-in of 150 thousand. Entries report the posterior mean; 95% credible intervals (2.5th and 97.5th percentiles) are in parentheses. The continuous-time parametrization is recovered via $b = \ln(\rho)/\Delta$, $\sigma^2 = 4\tau b/(\rho - 1)$ and $a = (n/2)\sigma^2/2$, with $\Delta = 1/52$. The non-informative prior sets $\underline{\alpha} = \underline{\beta} = \underline{\nu} = 0.1$, $\underline{\lambda} = 1$, $\underline{a} = 1/3$ and $\underline{b} = 99$. The weakly informative prior is calibrated from a pre-sample (1986–2006); see text for hyperparameters. The spike-and-slab prior places mass $\underline{p} = 0.5$ on $\rho = 1$ and mass $1 - \underline{p}$ on $\rho \sim \text{U}(0, 1)$; posterior statistics for θ and η use all MCMC draws, with the continuous-time map evaluated at $\rho = 1$ via its continuity limit ($b \rightarrow 0$, $\sigma^2 \rightarrow 4\tau/\Delta$).

Next, we compare the CIR process with two alternative models: the Ornstein–Uhlenbeck model of Vasicek (1977) and the 3/2 model of Ahn and Gao (1999). In the specific case of Vasicek (1977) the short rate is assumed to follow an Ornstein–Uhlenbeck process, characterized by the strong solution of the following stochastic differential equation,

$$dX_t = (a + bX_t)dt + \sigma dW_t \quad (2)$$

where $X_0 = x_0 \in \mathbb{R}$, $a, b \in \mathbb{R}$ and $\sigma > 0$. For specific details and properties of this process we refer the reader to Revuz and Yor (1999).

In the parametric term structure model of Ahn and Gao (1999), the short rate follows the 3/2 process, characterized by the strong solution to the following stochastic differential equation,

$$dX_t = - [b + (a - \sigma^2) X_t] X_t dt + \sigma X_t^{3/2} dW_t \quad (3)$$

where $X_0 = x_0 > 0$, $a > 0$, $b \in \mathbb{R}$ and $\sigma > 0$. The introduction of this short rate process is motivated by evidence that the heteroskedastic form of the conditional volatility is close to $X_t^{3/2}$, much larger than the 1/2 value of the CIR process, and a non-linear drift whereby mean-reversion is stronger at higher rates, cf. Ait-Sahalia (1996) and Stanton (1997).

The interpretation of the parameters in the 3/2 process is best understood by its close connection to the CIR process. If X_t follows a 3/2 process as in equation 3, $Y_t = 1/X_t$ will follow a CIR process as in equation 1. Therefore, the full Gibbs sampler of Proposition 1 and the Metropolis-within-Gibbs of Proposition 3 can be employed to estimate the 3/2 model by inverting the data before estimation. The parameter-free Jacobian, $\prod_{t=1}^T 1/x_t^2$, cancels in the acceptance ratio of any Metropolis-Hastings algorithm.

If $b < 0$ and $\frac{a}{\sigma^2} > 1$, the case considered by Ahn and Gao (1999), the process is stationary ergodic with the invariant distribution being $\mathcal{JG}(\frac{2a}{\sigma^2}, -\frac{2b}{\sigma^2})$. We can restrict the parameter space to this case by imposing $\rho < 1$ and $n/2 > 2$ in the prior distribution. The drift function, $-[b + (a - \sigma^2)x]x$ has two roots: 0 and $x^* = -\frac{b}{a - \sigma^2}$. $x^* > 0$ is the threshold level of the interest rate at which the drift is zero. Above x^* the drift is negative and below x^* the drift

is positive. From the drift polynomial we see that $a - \sigma^2$ determines the curvature of the drift.

The Vasicek model has a closed-form Gaussian transition density, $X_t | X_{t-1} = x_{t-1} \sim \mathcal{N}(\beta_0 + \beta_1 x_{t-1}, c^2)$ where $\beta_1 = e^{b\Delta}$, $\beta_0 = (a/b)(\beta_1 - 1)$ and $c^2 = \sigma^2(\beta_1^2 - 1)/(2b)$. Employing independent normal inverse-gamma priors for $\beta = (\beta_0, \beta_1)'$ and c^2 one obtains the standard conjugate two-block Gibbs sampler. We adopt a proper but diffuse normal prior for β centered at the unit root, $\underline{m} = (0, 0.99)'$ with covariance $\underline{V} = \text{diag}(100, 5)$, and a proper Inverse-Gamma on c^2 with scale $\underline{b}_0 = 0.4\bar{X}$ and shape 0.1. The choice of prior scale for c^2 stems from the short term behavior of the conditional variance. Over small Δ the conditional variance of the CIR process is given by $\sigma^2 \Delta x_{t-1} \approx 4\tau x_{t-1}$ while for the Vasicek model it is given by c^2 . Taking the same shape and scale prior hyperparameters as τ we scale the prior for c^2 to match the conditional variance of the CIR evaluated at the sample mean.⁴ As for the 3/2 model prior we employ the same non-informative prior as in the CIR process applied to the transformation $Y_t = 1/X_t$.

Table 4 reports posterior summary statistics for the Vasicek and 3/2 models as well as their respective marginal likelihoods. All reported marginal likelihoods, Tables 2 and 4, are computed employing the method of Chib (1995).⁵ Figure 9 depicts the corresponding marginal posterior distributions for the continuous time parametrization.

⁴Additional results setting $\underline{b}_0 = 0.1$ provide an identical marginal likelihood.

⁵The method of Chib (1995) evaluates the posterior full-conditional ordinate of $n/2$ at a single high-density point and therefore requires the normalizing constant $C(\lambda, \nu) = \int_0^\infty [\lambda^z / \Gamma(z)]^\nu dz$, which has no known closed form. We evaluate $C(\lambda, \nu)$ by Laplace's method, $\ln C(\lambda, \nu) \approx \nu[m \ln \lambda - \ln \Gamma(m)] + \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \nu - \frac{1}{2} \ln \psi'(m)$, where $m = \psi^{-1}(\ln \lambda)$ is the mode and ψ' the trigamma function.

Table 4: Alternative Models Posterior Summary Statistics

	Mean	Median	Std. Dev.	95% C.I.
<i>Vasicek model</i> $dX_t = (a + bX_t) dt + \sigma dW_t$				
$\sigma^2 (\times 10^3)$	6.881	6.873	0.301	(6.310, 7.501)
b	-0.197	-0.197	0.122	(-0.435, 0.042)
a	35.318	35.228	30.954	(-25.972, 95.323)
<i>3/2 model</i> $dX_t = -[b + (a - \sigma^2)X_t] X_t dt + \sigma X_t^{3/2} dW_t$				
σ^2	0.049	0.049	0.002	(0.045, 0.054)
b	-2.341	-2.336	0.651	(-3.630, -1.076)
a	0.036	0.036	0.005	(0.026, 0.045)
<i>Marginal Likelihood</i>				
Vasicek: $\ln \pi(\tilde{X}_T) = -4038.401$		3/2: $\ln \pi(\tilde{X}_T) = -5591.406$		

Notes: The data are weekly 3-Year U.S. Treasury yields (Wednesdays), expressed in basis points, covering the period from March 29th, 2006 to March 25th, 2026 ($T = 1043$). The 95% credible interval is computed as the 2.5th and 97.5th posterior percentiles. Vasicek: $X_t = \beta_0 + \beta_1 X_{t-1} + c \varepsilon_t$ with proper independent priors $\beta = (\beta_0, \beta_1)' \sim N(\underline{m}, \underline{V})$ and $c^2 \sim \text{IG}(\underline{a}, \underline{b})$, $\underline{m} = (0, 0.99)'$, $\underline{V} = \text{diag}(100, 5)$, $\underline{a} = 0.1$ and $\underline{b} = 4\bar{X}(0.1) = 81.363$; Posteriors are based on 50 thousand MCMC draws after a burn-in of 3 thousand. Continuous-time parameters are recovered via $b = \ln(\beta_1)/\Delta$, $\sigma^2 = 2b c^2 / (\beta_1^2 - 1)$ and $a = \beta_0 b / (\beta_1 - 1)$. 3/2 model: estimated on the inverted data $Y_t = 1/X_t$, which follows a CIR process, with the same non-informative prior used for the CIR ($\alpha = \beta = \underline{\nu} = 0.1$, $\underline{a} = 1/3$, $\underline{b} = 99$ and $\underline{\lambda} = 1$); posteriors are based on 3.5 million MCMC draws after a burn-in of 150 thousand. Parameters are recovered via $b = \ln(\rho)/\Delta$, $\sigma^2 = 4\tau b / (\rho - 1)$ and $a = (n/2) \sigma^2 / 2$. In both cases $\Delta = 1/52$.

For the Vasicek model, the drift parameters, a and b , are very close to those of the CIR model with a non-informative prior, see Table 2, with comparable credible intervals and unit-root behavior. Results for the 3/2 model are very different, with strong evidence in favor of mean reversion and the first moment being finite, $\mathbb{P}(n/2 > 1 | \tilde{x}_T) = 0.9930$. Yet, the variance of the process is infinite with high probability, $\mathbb{P}(n/2 < 2 | \tilde{x}_T) = \mathbb{P}(a - \sigma^2 < 0) = 0.9985$, meaning also that mean reversion is not stronger at higher levels, rather the opposite. When comparing models to one another, the standard Bayesian approach is to employ Bayes Factors (BF). BF have the interpretation of fair betting odds in the case of equal prior odds as well as the difference in logarithmic predictive scores between models. The BF is given by the ratio of marginal likelihoods, whose logarithmic values are reported in Tables 2 and 4. The evidence decisively favors the CIR process over the Vasicek and 3/2 models.

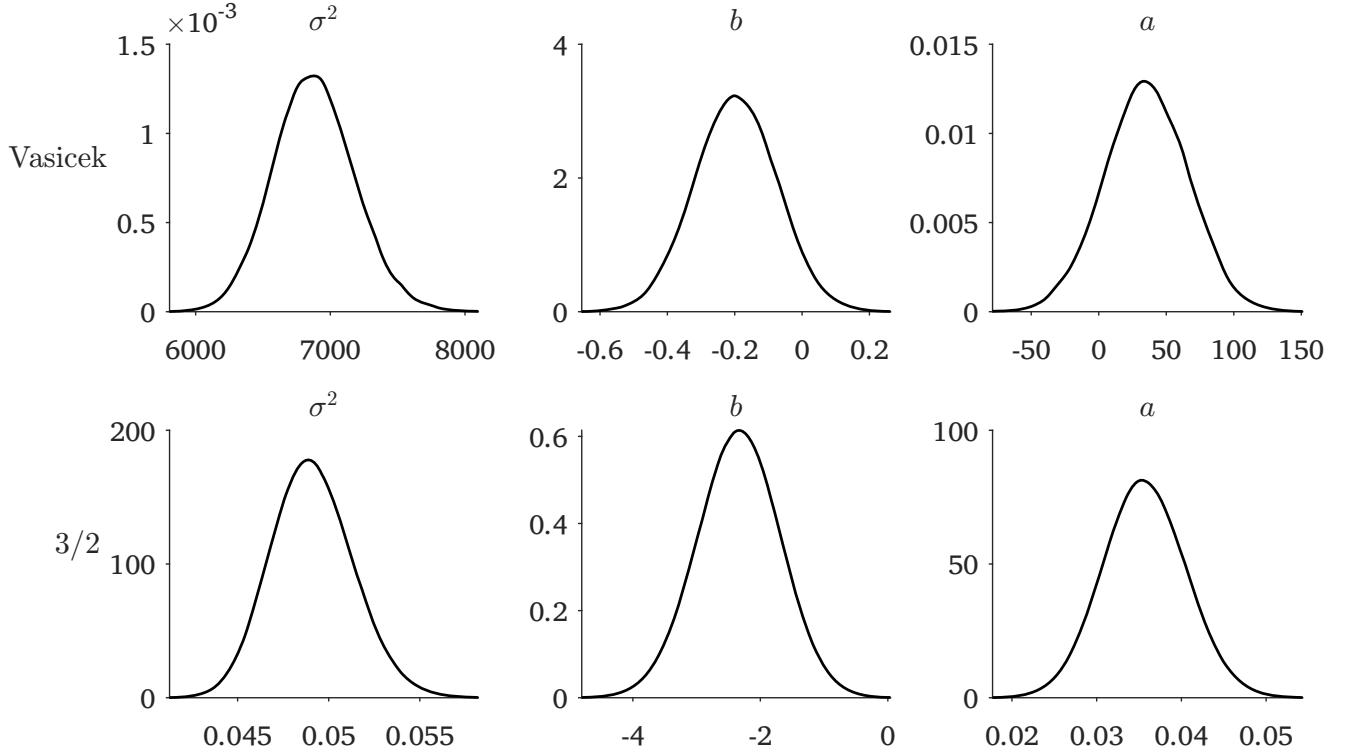


Figure 9: Alternative Models Marginal Posterior Distributions

Notes: The data are weekly 3-Year U.S. Treasury yields (Wednesdays), expressed in basis points, covering the period from March 29th, 2006 to March 25th, 2026 ($T = 1043$). Vasicek: $X_t = \beta_0 + \beta_1 X_{t-1} + c \varepsilon_t$ with proper independent priors $\beta = (\beta_0, \beta_1)' \sim N(\underline{m}, \underline{V})$ and $c^2 \sim \text{IG}(\underline{a}, \underline{b})$ where $\underline{m} = (0, 0.99)'$, $\underline{V} = \text{diag}(100, 5)$, $\underline{a} = 0.1$ and $\underline{b} = 4\bar{X}(0.1) = 81.3632$; Posteriors are based on 50 thousand MCMC draws after a burn-in of 3 thousand. Continuous-time parameters are recovered via $b = \log(\beta_1)/\Delta$, $\sigma^2 = 2b c^2/(\beta_1^2 - 1)$ and $a = \beta_0 b/(\beta_1 - 1)$. 3/2 model: estimated on the inverted data $Y_t = 1/X_t$, which follows a CIR process, with the same non-informative prior used for the CIR ($\underline{\alpha} = \underline{\beta} = \underline{\nu} = 0.1$, $\underline{a} = 1/3$, $\underline{b} = 99$ and $\underline{\lambda} = 1$); posteriors are based on 3.5 million MCMC draws after a burn-in of 150 thousand. Parameters are recovered via $b = \log(\rho)/\Delta$, $\sigma^2 = 4\tau b/(\rho - 1)$ and $a = (n/2)\sigma^2/2$. In both cases $\Delta = 1/52$.

5 Conclusion

In this paper we have obtained a full Gibbs Sampler for the exact CIR process observed at equidistant times. We do not require the imposition of the Feller condition. We show how to carry out smoothing and handling of missing data. Additionally, the introduction of a single tailored independent Metropolis-Hastings step allows us to handle irregular and random time intervals. In Appendix C we further propose a hybrid sampler, employing a collapsed Metropolis-Hastings update, which improves mixing by more than an order of magnitude. For the stationary ergodic case we obtain several asymptotic properties which include Local Asymptotic Normality, consistency and asymptotic normality of conditional maximum likelihood estimation and a Bernstein-von Mises theorem. Lastly, we provide

an empirical illustration using 3-Year U.S. Treasury yields under a variety of priors and a comparison with other processes such as the Vasicek and the 3/2 process. On the basis of marginal likelihoods, for a sample starting on March 29th, 2006 and ending on March 25th, 2026, the CIR model is strongly preferred to the Vasicek and 3/2 processes.

Two extensions building on the current results seem to us particularly interesting. The first is a multivariate extension. In multi-factor term structure models, cf. Jagannathan et al. (2003), the short rate is a sum of independent square-root factors. Conditional on the latent paths, the augmented likelihood factorizes, so that our results apply to each factor and only the sampling of the factor paths remains to be developed. The second is the case in which the CIR process is itself unobserved, as for the latent volatility of Heston (1993) or the latent hazard of Bielecki et al. (2011). In both cases the process must be imputed, and the exact conditional posteriors obtained above would replace the generic Metropolis-Hastings steps of León-González (2019), with the squared Bessel bridge results of Section 2.2 providing a natural starting point. We are currently exploring some of these research avenues.

A Proofs

This appendix contains the proofs for propositions and theorems reported in Section 2 and Section 3.

Proof of Proposition 1.

The augmented likelihood is given by

$$\mathcal{L} \left(\tilde{x}_{1:T}, \tilde{s}_T, \tilde{h}_T \mid \tau, \rho, \frac{n}{2}, x_0 \right) = \prod_{t=1}^T \pi \left(x_t \mid h_t, \tau, \frac{n}{2} \right) \pi \left(s_t \mid h_t, \frac{n}{2} \right) \pi \left(h_t \mid x_{t-1}, \tau, \rho \right),$$

where

$$\begin{aligned} \pi \left(x_t \mid h_t, \tau, \frac{n}{2} \right) &= \frac{x_t^{\frac{n}{2}+h_t-1}}{\Gamma \left(\frac{n}{2} + h_t \right) (2\tau)^{\frac{n}{2}+h_t}} \exp \left\{ -\frac{1}{2\tau} x_t \right\} \\ \pi \left(s_t \mid h_t, \frac{n}{2} \right) &= \frac{\Gamma \left(\frac{n}{2} + h_t \right)}{\Gamma \left(\frac{n}{2} \right) \Gamma \left(h_t \right)} s_t^{\frac{n}{2}-1} (1-s_t)^{h_t-1} \\ \pi \left(h_t \mid x_{t-1}, \tau, \rho \right) &= \exp \left\{ -\frac{\rho}{2\tau} x_{t-1} \right\} \left(\frac{\rho}{2\tau} x_{t-1} \right)^{h_t} \frac{1}{h_t!}. \end{aligned}$$

By Bayes' rule, the posterior is proportional to the likelihood times prior.

$$\pi \left(\tilde{h}_T, \tilde{s}_T, \tau, \rho, \frac{n}{2} \mid \tilde{x}_T \right) \propto \mathcal{L} \left(\tilde{x}_{1:T}, \tilde{h}_T, \tilde{s}_T \mid \tau, \rho, \frac{n}{2}, x_0 \right) \times \pi \left(\tau, \rho, \frac{n}{2} \mid x_0 \right).$$

By treating all unobservables except $\tilde{h}_T$ as a constant, dropping all constants of proportionality and reordering terms we are left with the following kernel, a product of the Bessel distribution introduced by Yuan and Kalbfleisch (2000):

$$\pi \left(\tilde{h}_T \mid \tilde{x}_T, \tau, \rho, \frac{n}{2} \right) \propto \prod_{t=1}^T \left(\frac{1}{h_t!} \frac{1}{\Gamma \left(\frac{n}{2} + h_t \right)} \left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{2h_t} \right).$$

Next, to derive the conditional posterior of τ , we go back to the likelihood times prior, treating all variables except τ as constant, integrating out $\tilde{s}_T$, dropping all constants of

proportionality and rearranging:

$$\begin{aligned}\pi\left(\tau \mid \tilde{x}_T, \tilde{h}_T, \rho, \frac{n}{2}\right) &\propto \pi(\tau \mid \Delta, x_0) \times (\tau)^{-\left(\frac{n}{2}T+2\sum_{t=1}^T h_t\right)} \\ &\times \exp\left\{-\frac{1}{2}\left(\rho x_0 + x_T + (1+\rho)\sum_{t=1}^{T-1} x_t\right)\frac{1}{\tau}\right\} \\ &\propto (\tau)^{-\left(\underline{a}+\frac{n}{2}T+2\sum_{t=1}^T h_t+1\right)} \exp\left\{-\left(\underline{\beta}+\frac{1}{2}\left(\rho x_0 + x_T + (1+\rho)\sum_{t=1}^{T-1} x_t\right)\right)\frac{1}{\tau}\right\}.\end{aligned}$$

Therefore the conditional posterior for τ is $\mathcal{JG}(\bar{\alpha}, \bar{\beta})$.

To derive the conditional posterior of ρ , we go back to the likelihood times prior, treating all variables except ρ as constant, dropping all constants of proportionality and rearranging:

$$\begin{aligned}\pi\left(\rho \mid \tilde{x}_T, \tilde{h}_T, \tilde{s}_T, \tau, \frac{n}{2}\right) &\propto \pi(\rho \mid \Delta, x_0) \times (\rho)^{\sum_{t=1}^T h_t} \exp\left\{-\left(\frac{1}{2\tau}\sum_{t=0}^{T-1} x_t\right)\rho\right\} \\ &\propto (\rho)^{\underline{a}+\sum_{t=1}^T h_t-1} \exp\left\{-\left(\frac{1}{\underline{b}}+\frac{1}{2\tau}\sum_{t=0}^{T-1} x_t\right)\rho\right\}.\end{aligned}$$

Therefore the conditional posterior for ρ is $\mathcal{G}(\bar{a}, \bar{b})$.

Finally, to derive the conditional posterior of $\frac{n}{2}$, we go back to the likelihood times prior, treating all variables except $\frac{n}{2}$ as constant, dropping all constants of proportionality and rearranging:

$$\begin{aligned}\pi\left(\frac{n}{2} \mid \tilde{s}_T, \tilde{h}_T, \tilde{x}_T, \tau\right) &\propto \pi\left(\frac{n}{2} \mid \Delta, x_0\right) \times \left(\prod_{t=1}^T \frac{x_t s_t}{2\tau}\right)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)^{-T} \\ &\propto \left(\underline{\lambda}^{\underline{\nu}/(\underline{\nu}+T)} \left[\left(\prod_{t=1}^T \frac{x_t s_t}{2\tau}\right)^{1/T}\right]^{T/(\underline{\nu}+T)}\right)^{(\underline{\nu}+T)\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)^{-(\underline{\nu}+T)}.\end{aligned}$$

Therefore the conditional posterior of $\frac{n}{2}$ is $\mathcal{GC1}(\bar{\lambda}, \bar{\nu})$. □

Proof of Theorem 1.

The state space of the Markov chain defined by the Gibbs sampler in Proposition 1 is $E = \mathcal{S}^T \times (0, \infty)^3$, where $\mathcal{S} = \{(h, s) : h \in \mathbb{N}_0, s \in (0, 1] \text{ and } s = 1 \text{ iff } h = 0\}$. Let ν denote the product measure consisting of the counting measure on $\mathbb{N}_0^T$, the spike-and-slab measure on each s_t and Lebesgue measure on $(0, \infty)^3$. Let $\mathcal{E}$ denote the corresponding σ -algebra, and $P : E \times \mathcal{E} \rightarrow [0, 1]$ the corresponding transition function.

Under $\underline{\alpha}, \underline{\beta}, \underline{a}, \underline{\lambda}, \underline{\nu} > 0$, $\underline{b} < \infty$, $T \geq 1$ and $x_t > 0 \forall t$, the joint posterior distribution is proper and admits density $\pi \left(\tilde{h}_T, \tilde{s}_T, \tau, \rho, \frac{n}{2} \mid \tilde{x}_T, \Delta \right)$ w.r.t. ν .

The Markov transition density for the Gibbs sampler in Proposition 1; that is, the probability density function of

$$\left(\tilde{h}_T^{(m+1)}, \tilde{s}_T^{(m+1)}, \tau^{(m+1)}, \rho^{(m+1)}, (n/2)^{(m+1)} \right) \quad \text{given} \quad \left(\tilde{h}_T^{(m)}, \tilde{s}_T^{(m)}, \tau^{(m)}, \rho^{(m)}, (n/2)^{(m)} \right)$$

is

$$\begin{aligned} & \pi \left(\tilde{s}_T^{(m+1)}, \tilde{h}_T^{(m+1)} \mid \tau^{(m)}, \rho^{(m)}, (n/2)^{(m)} \right) \pi \left(\tau^{(m+1)} \mid \tilde{h}_T^{(m+1)}, \rho^{(m)}, (n/2)^{(m)} \right) \\ & \times \pi \left(\rho^{(m+1)} \mid \tilde{h}_T^{(m+1)}, \tau^{(m+1)} \right) \pi \left((n/2)^{(m+1)} \mid \tilde{s}_T^{(m+1)}, \tau^{(m+1)} \right). \end{aligned}$$

This transition density is proper and everywhere strictly positive on E since each of the conditional distributions is everywhere strictly positive on its support:

$$\pi \left(\tilde{s}_T^{(m+1)}, \tilde{h}_T^{(m+1)} \mid \tau^{(m)}, \rho^{(m)}, (n/2)^{(m)} \right)$$

assigns a positive mass to every $\tilde{h}_T \in \mathbb{N}_0^T$ and given $\tilde{h}_T^{(m+1)}$, each $s_t^{(m+1)}$ is equal to 1 a.s. when $h_t^{(m+1)} = 0$ and has a Beta density, strictly positive in $(0, 1)$ when $h_t \geq 1$;

$$\pi \left(\tau^{(m+1)} \mid \tilde{h}_T^{(m+1)}, \rho^{(m)}, (n/2)^{(m)} \right)$$

is an inverse-gamma density, strictly positive on $(0, \infty)$;

$$\pi \left(\rho^{(m+1)} \mid \tilde{h}_T^{(m+1)}, \tau^{(m+1)} \right)$$

is a gamma density, strictly positive on $(0, \infty)$ and

$$\pi \left((n/2)^{(m+1)} \mid \tilde{s}_T^{(m+1)}, \tau^{(m+1)} \right)$$

is also strictly positive on $(0, \infty)$. Thus, for any $(\tilde{h}_T, \tilde{s}_T, \boldsymbol{\theta})$ and every ν -measurable $A \subseteq E$,

$$\int_A \pi \left(\tilde{h}_T, \tilde{s}_T, \boldsymbol{\theta} \mid \tilde{x}_T, \Delta \right) d\nu(\boldsymbol{\theta}) > 0 \Rightarrow \int_A \pi \left(\tilde{h}'_T, \tilde{s}'_T, \boldsymbol{\theta}' \mid \tilde{h}_T, \tilde{s}_T, \boldsymbol{\theta} \right) \nu \left(\tilde{h}'_T, \tilde{s}'_T, \boldsymbol{\theta}' \right) > 0,$$

which is exactly the sufficient condition in Corollary 4.5.1 in Geweke (2005) for the Gibbs sampler to be Harris ergodic. □

Proof of Proposition 2.

The event $s_t = 1$ occurs only when $h_t = 0$, whose probability is given by

$$\gamma_t = \mathbb{P} \left(h_t = 0 \mid \tilde{x}_T, \Delta, \tau, \rho, \frac{n}{2} \right) = \frac{\left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}-1}}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right) \Gamma \left(\frac{n}{2} \right)}.$$

The variable s_t takes values in $(0, 1)$ whenever $h_t \geq 1$, which is given by the following mixture,

$$\begin{aligned} f(s_t) &= \sum_{h_t=1}^{\infty} \pi \left(s_t \mid h_t, \frac{n}{2} \right) \pi \left(h_t \mid \tilde{x}_T, \Delta, \rho, \tau, \frac{n}{2} \right) \\ &= \sum_{h_t=1}^{\infty} \frac{\Gamma \left(\frac{n}{2} + h_t \right)}{\Gamma \left(\frac{n}{2} \right) \Gamma(h_t)} s_t^{\frac{n}{2}-1} (1-s_t)^{h_t-1} \frac{\left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{2h_t + \frac{n}{2}-1}}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right) h_t! \Gamma \left(\frac{n}{2} + h_t \right)} \\ &= \frac{1}{\Gamma \left(\frac{n}{2} \right)} \frac{s_t^{\frac{n}{2}-1}}{1-s_t} \frac{1}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)} \left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}-1} \sum_{h_t=1}^{\infty} \frac{\left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}(1-s_t)}}{\tau} \right)^{2h_t}}{h_t! \Gamma(h_t)} \\ &= \frac{1}{\Gamma \left(\frac{n}{2} \right)} \frac{s_t^{\frac{n}{2}-1}}{1-s_t} \frac{1}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)} \left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}-1} \sum_{h_t=0}^{\infty} \frac{\left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}(1-s_t)}}{\tau} \right)^{2(h_t+1)}}{(h_t+1)! \Gamma(h_t+1)} \\ &= \frac{1}{\Gamma \left(\frac{n}{2} \right)} \frac{s_t^{\frac{n}{2}-1}}{\sqrt{1-s_t}} \left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}} \frac{1}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)} \sum_{h_t=0}^{\infty} \frac{\left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}(1-s_t)}}{\tau} \right)^{2h_t+1}}{h_t! \Gamma(h_t+1+1)} \\ &= \frac{1}{\Gamma \left(\frac{n}{2} \right)} \frac{s_t^{\frac{n}{2}-1}}{\sqrt{1-s_t}} \left(\frac{1}{2} \frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)^{\frac{n}{2}} \frac{\mathbf{I}_1 \left(\frac{\sqrt{\rho x_t x_{t-1}(1-s_t)}}{\tau} \right)}{\mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau} \right)}. \end{aligned}$$

□

Proof of Proposition 3.

Exploiting the fact that the process is conditionally Markov and Assumption 1, we can factorize the joint likelihood of $\tilde{x}_{t_1:t_T}$ and $\tilde{\Delta}_{t_T}$ as

$$\pi\left(\tilde{x}_{t_1:t_T}, \tilde{\Delta}_{t_T} \mid \boldsymbol{\theta}, \boldsymbol{\phi}, x_{t_0}\right) = \prod_{i=1}^T \pi\left(x_{t_i} \mid x_{t_{i-1}}, \Delta_{t_i}, \boldsymbol{\theta}\right) \times \prod_{i=1}^T \pi\left(\Delta_{t_i} \mid \tilde{x}_{t_{i-1}}, \tilde{\Delta}_{t_{i-1}}, \boldsymbol{\phi}\right).$$

By Bayes' rule, the posterior is proportional to the likelihood times prior.

$$\begin{aligned} \pi\left(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}\right) &\propto \prod_{i=1}^T \pi\left(x_{t_i} \mid x_{t_{i-1}}, \Delta_{t_i}, \boldsymbol{\theta}\right) \times \prod_{i=1}^T \pi\left(\Delta_{t_i} \mid \tilde{x}_{t_{i-1}}, \tilde{\Delta}_{t_{i-1}}, \boldsymbol{\phi}\right) \\ &\quad \times \pi\left(\boldsymbol{\theta} \mid x_{t_0}\right) \times \pi\left(\boldsymbol{\phi} \mid x_{t_0}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \pi\left(\boldsymbol{\theta} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}\right) &= \int \pi\left(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}\right) d\boldsymbol{\phi} \\ &\propto \prod_{i=1}^T \pi\left(x_{t_i} \mid x_{t_{i-1}}, \Delta_{t_i}, \boldsymbol{\theta}\right) \times \pi\left(\boldsymbol{\theta} \mid x_{t_0}\right) \times \int \prod_{i=1}^T \pi\left(\Delta_{t_i} \mid \tilde{x}_{t_{i-1}}, \tilde{\Delta}_{t_{i-1}}, \boldsymbol{\phi}\right) \pi\left(\boldsymbol{\phi} \mid x_{t_0}\right) d\boldsymbol{\phi} \\ &\propto \prod_{i=1}^T \pi\left(x_{t_i} \mid x_{t_{i-1}}, \Delta_{t_i}, \boldsymbol{\theta}\right) \times \pi\left(\boldsymbol{\theta} \mid x_{t_0}\right) \times \int \pi\left(\boldsymbol{\phi} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}\right) d\boldsymbol{\phi} \\ &\propto \prod_{i=1}^T \pi\left(x_{t_i} \mid x_{t_{i-1}}, \Delta_{t_i}, \boldsymbol{\theta}\right) \times \pi\left(\boldsymbol{\theta} \mid x_{t_0}\right), \end{aligned}$$

where Assumption 2 is applied in the last step. Since $(\tilde{s}_{t_T}, \tilde{h}_{t_T})$ only depend on $(\tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \boldsymbol{\theta})$,

we obtain the following augmented posterior,

$$\begin{aligned}\pi\left(\tilde{s}_{t_T}, \tilde{h}_{t_T}, \tau, \rho, \frac{n}{2} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \boldsymbol{\theta}\right) &= \prod_{i=1}^T \pi\left(s_{t_i}, h_{t_i} \mid x_{t_i}, x_{t_{i-1}}, \Delta_{t_i}, \tau, \rho, \frac{n}{2}\right) \times \pi\left(\tau, \rho, \frac{n}{2} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}\right) \\ &\propto \mathcal{L}\left(\tilde{x}_{t_1:t_T}, \tilde{s}_{t_T}, \tilde{h}_{t_T} \mid \tau, \rho, \frac{n}{2}, x_{t_0}, \tilde{\Delta}_{t_T}\right) \times \pi\left(\tau, \rho, \frac{n}{2} \mid x_{t_0}\right),\end{aligned}$$

where the augmented likelihood is given by

$$\mathcal{L}\left(\tilde{x}_{t_1:t_T}, \tilde{s}_{t_T}, \tilde{h}_{t_T} \mid \tau, \rho, \frac{n}{2}, x_{t_0}, \tilde{\Delta}_{t_T}\right) = \prod_{i=1}^T \pi\left(x_{t_i} \mid h_{t_i}, \tau, \rho, \frac{n}{2}\right) \pi\left(s_{t_i} \mid h_{t_i}, \frac{n}{2}\right) \pi\left(h_{t_i} \mid x_{t_{i-1}}, \tau, \rho\right)$$

and

$$\begin{aligned}\pi\left(x_{t_i} \mid h_{t_i}, \tau, \rho, \frac{n}{2}\right) &= \frac{x_{t_i}^{\frac{n}{2}+h_{t_i}-1}}{\Gamma\left(\frac{n}{2}+h_{t_i}\right) \left(\frac{2\tau(1-\rho^{\Delta_{t_i}})}{1-\rho}\right)^{\frac{n}{2}+h_{t_i}}} \exp\left\{-\frac{1-\rho}{2\tau(1-\rho^{\Delta_{t_i}})} x_{t_i}\right\} \\ \pi\left(s_{t_i} \mid h_{t_i}, \frac{n}{2}\right) &= \frac{\Gamma\left(\frac{n}{2}+h_{t_i}\right)}{\Gamma\left(\frac{n}{2}\right) \Gamma(h_{t_i})} s_{t_i}^{\frac{n}{2}-1} (1-s_{t_i})^{h_{t_i}-1} \\ \pi\left(h_{t_i} \mid x_{t_{i-1}}, \tau, \rho\right) &= \exp\left\{-\frac{\rho^{\Delta_{t_i}}}{2\tau} \frac{1-\rho}{1-\rho^{\Delta_{t_i}}} x_{t_{i-1}}\right\} \left(\frac{\rho^{\Delta_{t_i}}}{2\tau} \frac{1-\rho}{1-\rho^{\Delta_{t_i}}} x_{t_{i-1}}\right)^{h_{t_i}} \frac{1}{h_{t_i}!}.\end{aligned}$$

By treating all unobservables except $\tilde{h}_{t_T}$ as a constant, dropping all constants of proportionality and reordering terms we are left with the following kernel, a product of the Bessel distribution introduced by Yuan and Kalbfleisch (2000)

$$\pi\left(\tilde{h}_{t_T} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tau, \rho, \frac{n}{2}\right) \propto \prod_{i=1}^T \frac{1}{h_{t_i}!} \frac{1}{\Gamma\left(\frac{n}{2}+h_{t_i}\right)} \left(\frac{1}{2} \frac{1-\rho}{1-\rho^{\Delta_{t_i}}} \frac{\sqrt{\rho^{\Delta_{t_i}} x_{t_i} x_{t_{i-1}}}}{\tau}\right)^{2h_{t_i}}.$$

To derive the conditional posterior of τ , we go back to the likelihood times prior, treating all variables except τ as constant, dropping all constants of proportionality and rearranging:

$$\begin{aligned}\pi\left(\tau \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{h}_{t_T}, \rho, \frac{n}{2}\right) &\propto \pi(\tau \mid x_{t_0}) \times (\tau)^{-\frac{n}{2}T-2\sum_{i=1}^T h_{t_i}} \exp\left\{-\frac{1-\rho}{2\tau} \sum_{i=1}^T \frac{x_{t_i} + \rho^{\Delta_{t_i}} x_{t_{i-1}}}{1-\rho^{\Delta_{t_i}}}\right\} \\ &\propto (\tau)^{-\left(\frac{n}{2}T+2\sum_{i=1}^T h_{t_i}+1\right)} \exp\left\{-\left(\frac{\beta}{\tau} + \frac{1-\rho}{2} \sum_{i=1}^T \frac{x_{t_i} + \rho^{\Delta_{t_i}} x_{t_{i-1}}}{1-\rho^{\Delta_{t_i}}}\right) \frac{1}{\tau}\right\}.\end{aligned}$$

Therefore the conditional posterior for τ is $\mathcal{JG}(\bar{\alpha}, \bar{\beta})$.

The conditional posterior kernel of ρ is obtained by going back to the likelihood times prior term, treating all variables except ρ as constant, dropping constants of proportionality and rearranging. The resulting kernel is no longer of the exponential family and so has no conditionally conjugate prior.

Finally, to derive the conditional posterior of $\frac{n}{2}$, we go back to the likelihood times prior, treating all variables except $\frac{n}{2}$ as constant, dropping all constants of proportionality and rearranging:

$$\begin{aligned}\pi\left(\frac{n}{2} \mid \tilde{x}_{t_T}, \tilde{\Delta}_{t_T}, \tilde{s}_{t_T}, \tilde{h}_{t_T}, \tau, \rho\right) &\propto \lambda^{\frac{\nu}{2}} \Gamma\left(\frac{n}{2}\right)^{-\nu} \left[\left(\prod_{i=1}^T \frac{1-\rho}{1-\rho^{\Delta_{t_i}}} \frac{x_{t_i} s_{t_i}}{2\tau}\right)\right]^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)^{-T} \\ &\propto \left(\lambda^{\nu/(\nu+T)} \left[\left(\prod_{i=1}^T \frac{1-\rho}{1-\rho^{\Delta_{t_i}}} \frac{x_{t_i} s_{t_i}}{2\tau}\right)^{1/T}\right]^{T/(\nu+T)}\right)^{(\nu+T)\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)^{-(\nu+T)}.\end{aligned}$$

Therefore the conditional posterior of $\frac{n}{2}$ is $\mathcal{GC1}(\bar{\lambda}, \bar{\nu})$. This completes the proof. $\square$

Lemma A.1 *Bessel Moment Bounds.*

Let $X \sim \mathcal{Bessel}(\nu, a)$ with $\nu > -1$ and $a > 0$. Then

$$\begin{aligned}\mathbb{E}\{X\} &\leq \frac{\sqrt{a^2 + \nu^2} - \nu}{2}, \\ \mathbb{Var}\{X\} &\leq \mathbb{E}\{X\}, \\ \mathbb{E}\{|X - \mathbb{E}\{X\}|^3\} &\leq \mathbb{E}\{X\}(\mathbb{E}\{X\} + \max\{1, \nu\}).\end{aligned}$$

In particular, $\mathbb{E}\{X\}$ and $\mathbb{Var}\{X\}$ are at most $(1 + |\nu|)(1 + a)$, and $\mathbb{E}\{|X - \mathbb{E}\{X\}|^3\} \leq 2(1 + |\nu|)^2(1 + a)^2$.

Proof of Lemma A.1.

Lemma 2 of Devroye (2002) provides the first and second moments of X in terms of the Bessel quotient, $\mathbf{R}_\nu(a)$. Lemma 1 of Devroye (2002) provides bounds for the Bessel quotient that, once applied, deliver the bound for $\mathbb{E}\{X\}$. Since the Bessel quotient is non-increasing in ν for a fixed $a > 0$, Lemma 2 of Devroye (2002) delivers the bound for the variance.

For the third moment, let $\mu = \mathbb{E}\{X\}$. The proof of Lemma 3 in Devroye (2002) provides $\mathbb{P}(X = k) / \mathbb{P}(X = k - 1) = a^2 / (4k(k + \nu))$ for $k \geq 1$. Multiplying $k(k + \nu)\mathbb{P}(X = k) = \frac{a^2}{4}\mathbb{P}(X = k - 1)$ by $k - \mu - 1$ and summing over $k \geq 1$,

$$\mathbb{E}\{X(X + \nu)(X - \mu - 1)\} = \frac{a^2}{4}\mathbb{E}\{X - \mu\} = 0.$$

Since $X, \mu \geq 0$, we have $|X - \mu| \leq X + \mu$. Moreover, $(X - \mu)^2(X + \mu) - X(X + \nu)(X - \mu - 1) = (1 - \nu)(X - \mu)^2 + (2\mu + \nu - \mu(\mu + \nu))(X - \mu) + \mu(\mu + \nu)$. Hence,

$$\begin{aligned}\mathbb{E}\{|X - \mu|^3\} &\leq \mathbb{E}\{(X - \mu)^2(X + \mu) - X(X + \nu)(X - \mu - 1)\} \\ &= \mu(\mu + \nu) + (1 - \nu)\mathbb{Var}\{X\} \leq \mu(\mu + \max\{1, \nu\}),\end{aligned}$$

where the last inequality uses $(1 - \nu)\mathbb{Var}\{X\} \leq \max\{1 - \nu, 0\}\mu$. The last statement of the lemma follows from $\sqrt{a^2 + \nu^2} \leq a + |\nu|$ and $\max\{1, \nu\} \leq 1 + |\nu|$. $\square$

Proof of Theorem 2.

The proof of Local Asymptotic Normality presented here follows closely that of Overbeck and Rydén (1997) with a different approach to bounding derivatives in order to avoid employing the Feller condition. For completeness we reproduce the entire proof.

The transition density is given by

$$\begin{aligned}\pi(y \mid x, \boldsymbol{\theta}) &\equiv \frac{\partial}{\partial y} P_{\boldsymbol{\theta}}(X_t \leq y \mid X_{t-1} = x) \\ &= \frac{1}{2\tau} \cdot \exp \left\{ -\frac{1}{2\tau} (y + \rho x) \right\} \left(\frac{y}{\rho x} \right)^{(n-2)/4} \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right).\end{aligned}$$

A Taylor expansion up to order 2 of $\ln \pi(y \mid x, \cdot)$ around $\boldsymbol{\theta}_0$ with the normalizing factor specialized to $\delta_T = 1/\sqrt{T}$ yields,

$$\begin{aligned}\ln \pi(y \mid x, \boldsymbol{\theta}_0 + \delta_T \mathbf{u}) - \ln \pi(y \mid x, \boldsymbol{\theta}_0) &= T^{-1/2} \sum_{i=1}^3 u_i \frac{\partial}{\partial \theta_i} \ln \pi(y \mid x, \boldsymbol{\theta}_0) \\ &+ \frac{1}{2} T^{-1} \sum_{i,j=1}^3 u_i u_j \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi(y \mid x, \boldsymbol{\theta}_0) \\ &+ \frac{1}{6} T^{-3/2} \sum_{i,j,k=1}^3 u_i u_j u_k \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi \left(y \mid x, \boldsymbol{\theta}_0 + h T^{-\frac{1}{2}} \mathbf{u} \right),\end{aligned}$$

for some $h \in [0, 1]$. Hence, it will suffice to prove that

$$T^{-3/2} \sum_{t=1}^T \sum_{i,j,k=1}^3 u_i u_j u_k \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi \left(X_t \mid X_{t-1}, \boldsymbol{\theta}_0 + h_t T^{-\frac{1}{2}} \mathbf{u} \right) = o_{P_{\boldsymbol{\theta}_0, x_0}}(1), \quad (\text{D1})$$

for each sequence $\{h_t\}$ of random variables in $[0, 1]$, that

$$\lim_{T \rightarrow \infty} -\frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^3 u_i u_j \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi(X_t \mid X_{t-1}, \boldsymbol{\theta}_0) = \mathbf{u}^\top \mathbf{K}_{\boldsymbol{\theta}_0} \mathbf{u} \quad P_{\boldsymbol{\theta}_0, x_0}\text{-a.s.}, \quad (\text{D2})$$

and, by the Cramér-Wold device, that

$$T^{-\frac{1}{2}} \sum_{t=1}^T \sum_{i=1}^3 u_i \frac{\partial}{\partial \theta_i} \ln \pi (X_t | X_{t-1}, \boldsymbol{\theta}_0) \rightarrow \mathcal{N} (0, \mathbf{u}^\top \mathbf{K}_{\boldsymbol{\theta}_0} \mathbf{u}) \quad P_{\boldsymbol{\theta}_0, x_0}\text{-weakly.} \quad (\text{D3})$$

For the first condition to hold, that the remainder of the Taylor expansion is $o_{P_{\boldsymbol{\theta}_0, x_0}}(1)$, it is sufficient to show

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sup_{\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \epsilon)} \sum_{i,j,k=1}^3 \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi (X_t | X_{t-1}, \boldsymbol{\theta}) \right| < \infty \quad P_{\boldsymbol{\theta}_0, x_0}\text{-a.s.,}$$

for any sufficiently small $\epsilon > 0$, where $U(\boldsymbol{\theta}_0, \epsilon)$ is the open ball with center $\boldsymbol{\theta}_0$ and radius ϵ . The previous condition will in turn hold by ergodicity if

$$\mathbb{E}_{\boldsymbol{\theta}_0} \left\{ \sup_{\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \epsilon)} \sum_{i,j,k=1}^3 \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi (X_1^* | X_0^*, \boldsymbol{\theta}) \right| \right\} < \infty.$$

Expectations are taken w.r.t. the invariant density of a generic consecutive pair (X_1^*, X_0^*) , given by $\pi_{\boldsymbol{\theta}_0}(x_1^* | x_0^*) \times \pi_{\boldsymbol{\theta}_0}(x_0^*)$, where $\pi_{\boldsymbol{\theta}_0}(x_0^*)$ is the stationary density, $\mathcal{G} \left(\frac{n_0}{2}, \frac{2\tau_0}{1-\rho_0} \right)$. This will be achieved by bounding the third-order derivatives uniformly over $U(\boldsymbol{\theta}_0, \epsilon)$ and showing the resulting envelope to be integrable under the invariant density of (X_1^*, X_0^*) . The log density of the transition function is given by,

$$\ln \pi (y | x, \boldsymbol{\theta}) = -\ln(2\tau) - \frac{1}{2\tau}(y + \rho x) + \frac{n-2}{4} (\ln(y) - \ln(\rho) - \ln(x)) + \ln \mathbf{I}_{n/2-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right).$$

We will now bound the third derivative of each of the first three terms.

$$\begin{aligned} \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln(2\tau) \right| &= 2/\tau^3 \\ \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \frac{1}{2\tau} (y + \rho x) \right| &\leq \frac{3}{\tau^4} (y + \rho x) + \frac{1}{\tau^3} x \leq C_1(\boldsymbol{\theta})(y + x) \\ \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \frac{n-2}{4} (\ln(y) - \ln(\rho) - \ln(x)) \right| &\leq \frac{\rho + |n-2|}{2\rho^3}, \end{aligned}$$

where $C_m(\boldsymbol{\theta})$ is some continuous function of the parameters $\boldsymbol{\theta}$.

In order to obtain a bound for the third derivatives of the last term, $\ln \mathbf{I}_\nu(a)$ with $\nu = \frac{n}{2} - 1$ and $a = \sqrt{yx\rho}/\tau$, w.r.t. $(\rho, \tau, n/2)$, we first write them in terms of derivatives w.r.t. (ν, a) .

$$\begin{aligned} \frac{\partial^3}{\partial \rho^3} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{1}{8\rho^3} \left(a^3 \frac{\partial^3}{\partial a^3} \ln \mathbf{I}_\nu(a) - 3a^2 \frac{\partial^2}{\partial a^2} \ln \mathbf{I}_\nu(a) + 3a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \tau^3} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= -\frac{1}{\tau^3} \left(a^3 \frac{\partial^3}{\partial a^3} \ln \mathbf{I}_\nu(a) + 6a^2 \frac{\partial^2}{\partial a^2} \ln \mathbf{I}_\nu(a) + 6a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \rho^2 \partial \tau} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{1}{4\rho^2 \tau} \left(-a^3 \frac{\partial^3}{\partial a^3} \ln \mathbf{I}_\nu(a) - a^2 \frac{\partial^2}{\partial a^2} \ln \mathbf{I}_\nu(a) + a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \rho \partial \tau^2} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{1}{2\rho \tau^2} \left(a^3 \frac{\partial^3}{\partial a^3} \ln \mathbf{I}_\nu(a) + 4a^2 \frac{\partial^2}{\partial a^2} \ln \mathbf{I}_\nu(a) + 2a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \rho^2 \partial \frac{n}{2}} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{1}{4\rho^2} \left(a^2 \frac{\partial^3}{\partial a^2 \partial \nu} \ln \mathbf{I}_\nu(a) - a \frac{\partial^2}{\partial a \partial \nu} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \tau^2 \partial \frac{n}{2}} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{1}{\tau^2} \left(a^2 \frac{\partial^3}{\partial a^2 \partial \nu} \ln \mathbf{I}_\nu(a) + 2a \frac{\partial^2}{\partial a \partial \nu} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \tau \partial \rho \partial \frac{n}{2}} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= -\frac{1}{2\rho \tau} \left(a^2 \frac{\partial^3}{\partial a^2 \partial \nu} \ln \mathbf{I}_\nu(a) + a \frac{\partial^2}{\partial a \partial \nu} \ln \mathbf{I}_\nu(a) \right) \\ \frac{\partial^3}{\partial \rho \partial \left(\frac{n}{2}\right)^2} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{1}{2\rho} a \frac{\partial^3}{\partial \nu^2 \partial a} \ln \mathbf{I}_\nu(a) \\ \frac{\partial^3}{\partial \tau \partial \left(\frac{n}{2}\right)^2} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= -\frac{1}{\tau} a \frac{\partial^3}{\partial \nu^2 \partial a} \ln \mathbf{I}_\nu(a) \\ \frac{\partial^3}{\partial \left(\frac{n}{2}\right)^3} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) &= \frac{\partial^3}{\partial \nu^3} \ln \mathbf{I}_\nu(a). \end{aligned}$$

This leaves us with seven different quantities involving partial derivatives of $\ln \mathbf{I}_\nu(a)$. Each of these quantities, by term by term differentiation of $\ln \mathbf{I}_\nu(a)$ and simple but tedious algebraic manipulations, can be written in terms of expectations of a Bessel random variable functional. Let $Z \sim \mathcal{Bessel}(\nu, a)$ and $\mu = \mathbb{E}\{Z\}$. Then,

$$\begin{aligned}
a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) &= \nu + 2\mu \\
a^2 \frac{\partial^2}{\partial a^2} \ln \mathbf{I}_\nu(a) &= 4\mathbb{V}\text{ar}\{Z\} - \nu - 2\mu \\
a^3 \frac{\partial^3}{\partial a^3} \ln \mathbf{I}_\nu(a) &= 8\mathbb{E}\{(Z - \mu)^3\} - 12\mathbb{V}\text{ar}\{Z\} + 2\nu + 4\mu \\
a \frac{\partial^2}{\partial a \partial \nu} \ln \mathbf{I}_\nu(a) &= 1 - 2\mathbb{E}\{(Z - \mu)\psi(Z + \nu + 1)\} \\
a^2 \frac{\partial^3}{\partial a^2 \partial \nu} \ln \mathbf{I}_\nu(a) &= 2\mathbb{E}\{(Z - \mu)\psi(Z + \nu + 1)\} - 1 \\
&\quad - 4\mathbb{E}\{(Z - \mu)^2(\psi(Z + \nu + 1) - \mathbb{E}\{\psi(Z + \nu + 1)\})\} \\
a \frac{\partial^3}{\partial a \partial \nu^2} \ln \mathbf{I}_\nu(a) &= 2\mathbb{E}\{(Z - \mu)(\psi(Z + \nu + 1) - \mathbb{E}\{\psi(Z + \nu + 1)\})^2\} \\
&\quad - 2\mathbb{E}\{(Z - \mu)\psi^{(1)}(Z + \nu + 1)\} \\
\frac{\partial^3}{\partial \nu^3} \ln \mathbf{I}_\nu(a) &= 3\mathbb{E}\{(\psi(Z + \nu + 1) - \mathbb{E}\{\psi(Z + \nu + 1)\})\psi^{(1)}(Z + \nu + 1)\} \\
&\quad - \mathbb{E}\{(\psi(Z + \nu + 1) - \mathbb{E}\{\psi(Z + \nu + 1)\})^3\} - \mathbb{E}\{\psi^{(2)}(Z + \nu + 1)\}.
\end{aligned}$$

By applying the Bessel random variable moment bounds of Lemma A.1 and the mean value theorem, one can bound the previous quantities. Let $\psi'(\cdot)$ and $\psi''(\cdot)$ denote the trigamma and tetragamma functions. With $C_2(\boldsymbol{\theta}) \equiv \psi'(\nu + 1)$ and $C_3(\boldsymbol{\theta}) \equiv |\psi''(\nu + 1)|$, which are continuous on Θ^* since $\nu + 1 = \frac{n}{2} > 0$, for all $\nu > -1$ and $a > 0$,

$$\begin{aligned}
\left| a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) \right| &\leq \max\{|\nu|, \nu + 2\mu\} \leq \sqrt{a^2 + \nu^2} \leq (1 + |\nu|)(1 + a) \\
\left| a^2 \frac{\partial^2}{\partial a^2} \ln \mathbf{I}_\nu(a) \right| &\leq 2\mu + |\nu| \leq \sqrt{a^2 + \nu^2} + |\nu| - \nu \leq (1 + 3|\nu|)(1 + a) \\
\left| a^3 \frac{\partial^3}{\partial a^3} \ln \mathbf{I}_\nu(a) \right| &\leq 8\mathbb{E}\{|Z - \mu|^3\} + 12\mathbb{V}\text{ar}\{Z\} + 4\mu + 2|\nu| \leq 34(1 + |\nu|)^2(1 + a)^2 \\
\left| a \frac{\partial^2}{\partial a \partial \nu} \ln \mathbf{I}_\nu(a) \right| &\leq 1 + 2C_2(\boldsymbol{\theta})\mathbb{V}\text{ar}\{Z\} \leq (1 + 2C_2(\boldsymbol{\theta})(1 + |\nu|))(1 + a) \\
\left| a^2 \frac{\partial^3}{\partial a^2 \partial \nu} \ln \mathbf{I}_\nu(a) \right| &\leq 1 + 2C_2(\boldsymbol{\theta})\mathbb{V}\text{ar}\{Z\} + 8C_2(\boldsymbol{\theta})\mathbb{E}\{|Z - \mu|^3\} \\
&\leq (1 + 18C_2(\boldsymbol{\theta}))(1 + |\nu|)^2(1 + a)^2 \\
\left| a \frac{\partial^3}{\partial a \partial \nu^2} \ln \mathbf{I}_\nu(a) \right| &\leq 2C_3(\boldsymbol{\theta})\mathbb{V}\text{ar}\{Z\} + 6C_2(\boldsymbol{\theta})^2\mathbb{E}\{|Z - \mu|^3\} \\
&\leq (2C_3(\boldsymbol{\theta}) + 12C_2(\boldsymbol{\theta})^2)(1 + |\nu|)^2(1 + a)^2 \\
\left| \frac{\partial^3}{\partial \nu^3} \ln \mathbf{I}_\nu(a) \right| &\leq C_3(\boldsymbol{\theta}) + 6C_2(\boldsymbol{\theta})C_3(\boldsymbol{\theta})\mathbb{V}\text{ar}\{Z\} + 8C_2(\boldsymbol{\theta})^3\mathbb{E}\{|Z - \mu|^3\} \\
&\leq (C_3(\boldsymbol{\theta}) + 6C_2(\boldsymbol{\theta})C_3(\boldsymbol{\theta}) + 16C_2(\boldsymbol{\theta})^3)(1 + |\nu|)^2(1 + a)^2,
\end{aligned}$$

where, after the first inequality, each line uses $\mu \leq (\sqrt{a^2 + \nu^2} - \nu)/2 \leq (1 + |\nu|)(1 + a)$, $\mathbb{V}\text{ar}\{Z\} \leq \mu$, $\mathbb{E}\{|Z - \mu|^3\} \leq \mu(\mu + \max\{1, \nu\}) \leq 2(1 + |\nu|)^2(1 + a)^2$ and $(1 + |\nu|)(1 + a) \geq 1$. Hence each of the seven quantities is bounded in absolute value by $C_4(\boldsymbol{\theta})(1 + a)^2$, where

$$C_4(\boldsymbol{\theta}) = (1 + |\nu|)^2 (34 + 18C_2(\boldsymbol{\theta}) + 12C_2(\boldsymbol{\theta})^2 + 16C_2(\boldsymbol{\theta})^3 + 2C_3(\boldsymbol{\theta}) + 6C_2(\boldsymbol{\theta})C_3(\boldsymbol{\theta}))$$

is continuous on Θ^* , the non-Feller range $0 < \frac{n}{2} < 1$ included.

The chain-rule composites above write each third derivative of $\ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right)$ as a combination of the seven quantities whose coefficients are a constant times products of powers of $1/\tau$ and $1/\rho$. Among the 27 ordered triples (i, j, k) , the three pure third derivatives appear once, the six derivatives with exactly two equal indices appear three times and $\frac{\partial^3}{\partial \tau \partial \rho \partial \frac{n}{2}}$ appears six times. Therefore,

$$\sum_{i,j,k=1}^3 \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \mathbf{I}_{\frac{n}{2}-1} \left(\frac{\sqrt{yx\rho}}{\tau} \right) \right| \leq C_5(\boldsymbol{\theta}) C_4(\boldsymbol{\theta}) (1+a)^2,$$

where

$$C_5(\boldsymbol{\theta}) = \frac{13}{\tau^3} + \frac{7}{8\rho^3} + 1 + \frac{9}{4\rho^2\tau} + \frac{21}{2\rho\tau^2} + \frac{3}{2\rho^2} + \frac{9}{\tau^2} + \frac{6}{\rho\tau} + \frac{3}{\tau} + \frac{3}{2\rho}$$

collects the absolute values of the coefficients of the ten composites weighted by their multiplicities. Since $2a \leq 1 + a^2$ and $a^2 = \rho xy / \tau^2 \leq xy / \tau^2$ for $\rho < 1$, $(1+a)^2 \leq 2(1+a^2) \leq 2(1 + \frac{1}{\tau^2})(1 + yx)$. For the first three terms of the log density only four of the ten third derivatives are non-zero, $\frac{\partial^3}{\partial \tau^3}$, $\frac{\partial^3}{\partial \rho^3}$, $\frac{\partial^3}{\partial \tau^2 \partial \rho}$ and $\frac{\partial^3}{\partial \rho^2 \partial \frac{n}{2}}$, and with the same multiplicities their absolute values sum to at most $\frac{2}{\tau^3} + \frac{|n-2|}{2\rho^3} + \frac{3}{2\rho^2} + \left(\frac{3}{\tau^4} + \frac{3}{\tau^3} \right) (x+y)$. Therefore, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$ and $x, y > 0$,

$$\sum_{i,j,k=1}^3 \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi(y | x, \boldsymbol{\theta}) \right| \leq C_6(\boldsymbol{\theta}) (1 + y + x + yx),$$

with

$$C_6(\boldsymbol{\theta}) = \frac{5}{\tau^3} + \frac{3}{\tau^4} + \frac{|n-2|}{2\rho^3} + \frac{3}{2\rho^2} + 2 \left(1 + \frac{1}{\tau^2} \right) C_5(\boldsymbol{\theta}) C_4(\boldsymbol{\theta}).$$

The function $C_6(\boldsymbol{\theta})$ is continuous on $\boldsymbol{\Theta}^*$, since $C_4(\boldsymbol{\theta})$ and $C_5(\boldsymbol{\theta})$ are continuous on $\boldsymbol{\Theta}^*$ and so are the remaining terms for $\tau, \rho > 0$.

Since $C_6(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, we can upgrade the pointwise bound into a uniform bound over $U(\boldsymbol{\theta}_0, \epsilon)$. Fix $\epsilon > 0$ sufficiently small such that $\overline{U}(\boldsymbol{\theta}_0, \epsilon) \subset \boldsymbol{\Theta}^*$. Since $C_6(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*$, it is bounded on the compact set $\overline{U}(\boldsymbol{\theta}_0, \epsilon)$. Hence, there exists $C_6(\boldsymbol{\theta}_0, \epsilon) \equiv \max_{\boldsymbol{\theta} \in \overline{U}(\boldsymbol{\theta}_0, \epsilon)} C_6(\boldsymbol{\theta}) < \infty$ such that for each $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}^*$ and $\forall x, y > 0$,

$$\sup_{\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \epsilon)} \sum_{i,j,k=1}^3 \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi(y | x, \boldsymbol{\theta}) \right| \leq C_6(\boldsymbol{\theta}_0, \epsilon) (1 + y + x + yx).$$

Condition D1 is satisfied,

$$\begin{aligned}
& \mathbb{E}_{\theta_0} \left\{ \sup_{\theta \in U(\theta_0, \epsilon)} \sum_{i,j,k=1}^3 \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln \pi (X_1^* | X_0^*, \theta) \right| \right\} \\
& \leq C_6(\theta_0, \epsilon) (1 + \mathbb{E}_{\theta_0} \{X_0^*\} + \mathbb{E}_{\theta_0} \{X_1^*\} + \mathbb{E}_{\theta_0} \{ \mathbb{E}_{\theta_0} \{X_1^* | X_0^*\} X_0^* \}) \\
& \leq C_6(\theta_0, \epsilon) (1 + \mathbb{E}_{\theta_0} \{X_0^*\} + \mathbb{E}_{\theta_0} \{X_1^*\} + \rho_0 \mathbb{E}_{\theta_0} \{(X_0^*)^2\} + (1 - \rho_0) \mathbb{E}_{\theta_0} \{X_1^*\} \mathbb{E}_{\theta_0} \{X_0^*\}) \\
& < \infty.
\end{aligned}$$

Following a similar approach we can show the existence of

$$\mathbb{E}_{\theta_0} \left\{ \sum_{i,j=1}^3 u_i u_j \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi (X_1^* | X_0^*, \theta_0) \right\}$$

for each $\theta_0 \in \Theta^*$. Therefore, by ergodicity,

$$\begin{aligned}
& \lim_{T \rightarrow \infty} -\frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^3 u_i u_j \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi (X_t | X_{t-1}, \theta_0) \\
& = \mathbb{E}_{\theta_0} \left\{ -\sum_{i,j=1}^3 u_i u_j \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi (X_1 | X_0, \theta_0) \right\} \quad P_{\theta_0, x_0}\text{-a.s.}
\end{aligned}$$

Rewritten in matrix notation, we obtain that $\mathbf{K}_T$ converges almost surely to the matrix with $\mathbb{E}_{\theta_0} \left\{ -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi (X_1^* | X_0^*, \theta_0) \right\}$ as its ij entry. Hence $\mathbf{K}_T \rightarrow \mathbf{K}_{\theta_0}$ P_{θ_0, x_0} -a.s. and condition D2 is proven.

By a similar strategy one can show that

$$\sup_{\theta \in U(\theta_0, \epsilon)} \sum_{i,j=1}^3 \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi (X_1^* | X_0^*, \theta) \right| \quad \text{and} \quad \sup_{\theta \in U(\theta_0, \epsilon)} \sum_{i=1}^3 \left| \frac{\partial}{\partial \theta_i} \ln \pi (X_1^* | X_0^*, \theta) \right|$$

are integrable under the invariant density of (X_1^*, X_0^*) for each $\theta_0 \in \Theta^*$, since the bounds

will involve the first, second and (absolute) logarithmic moments of X_1^* and X_0^* . We can therefore exchange differentiation with integration. This, jointly with the fact that $\pi(y | x, \boldsymbol{\theta}_0)$ is a density w.r.t. Lebesgue measure, implies,

$$\mathbb{E}_{\boldsymbol{\theta}_0} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \pi(X_1^* | X_0^*, \boldsymbol{\theta}_0) \right\} / \pi(X_1^* | X_0^*, \boldsymbol{\theta}_0) = 0,$$

and so

$$-\mathbb{E}_{\boldsymbol{\theta}_0} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \pi(X_1^* | X_0^*, \boldsymbol{\theta}_0) \right\} = \mathbb{E}_{\boldsymbol{\theta}_0} \left\{ \frac{\partial}{\partial \theta_i} \ln \pi(X_1^* | X_0^*, \boldsymbol{\theta}_0) \times \frac{\partial}{\partial \theta_j} \ln \pi(X_1^* | X_0^*, \boldsymbol{\theta}_0) \right\}.$$

To prove condition D3, we will show that higher-order moments of $\partial \ln \pi(y | x, \boldsymbol{\theta}) / \partial \theta_i$ exist, so we can employ the martingale central limit theorem (see, e.g. Durrett (1991), Ch. VII, Theorem 7.5). Following the same approach employed to bound the third-order derivatives and applying Hölder's inequality we can obtain the following bound,

$$\left| \frac{\partial \ln \pi(y | x, \boldsymbol{\theta})}{\partial \theta_i} \right|^p \leq C_7(\boldsymbol{\theta}) (1 + y^p + x^p + |\ln y|^p + |\ln x|^p),$$

where every term is integrable under the invariant density of (X_1^*, X_0^*) . In particular, with $p = 2$, $(\mathbf{K}_{\boldsymbol{\theta}_0})_{ii} < \infty$ and, by the Cauchy–Schwarz inequality, $|(\mathbf{K}_{\boldsymbol{\theta}_0})_{ij}| \leq ((\mathbf{K}_{\boldsymbol{\theta}_0})_{ii}(\mathbf{K}_{\boldsymbol{\theta}_0})_{jj})^{1/2} < \infty$, so $\mathbf{K}_{\boldsymbol{\theta}_0}$ is finite.

Lastly, Lemma A.4 shows that $\mathbf{K}_{\boldsymbol{\theta}_0}$ is non-singular. □

Lemma A.2 Identification.

The CIR process is identified. That is, for $\theta, \theta' \in \Theta^*$, $\pi(y | x, \theta) = \pi(y | x, \theta')$ almost everywhere in (y, x) implies $\theta = \theta'$.

Proof of Lemma A.2.

The transition density of the CIR process is that of a scaled non-central χ^2 , $\frac{1}{\tau}X_t | X_{t-1} = x \sim \chi_n^2\left(\frac{\rho}{\tau}x\right)$. Therefore, the first two conditional moments are,

$$\begin{aligned}\mathbb{E}\{X_t | X_{t-1} = x\} &= \rho x + n\tau \\ \mathbb{V}\text{ar}\{X_t | X_{t-1} = x\} &= 4\tau\rho x + 2n\tau^2.\end{aligned}$$

Assume $\pi(y | x, \theta) = \pi(y | x, \theta')$ for a set of x with positive probability under the stationary distribution. Since the stationary distribution is continuous in $(0, \infty)$, this set includes more than one value of x . Equality in transition distributions implies equality of the first two conditional moments, so the above affine functions of x must have matching slopes and intercepts under θ and θ' .

From the slope of the conditional expectation we obtain $\rho = \rho'$. From the slope of the conditional variance we obtain $4\tau\rho x = 4\tau'\rho'x$ and, since $\rho = \rho'$, it implies $\tau = \tau'$. From the intercept of the conditional expectation we obtain $n\tau = n'\tau'$ and, since $\tau = \tau'$, it implies $n = n'$. $\square$

Lemma A.3 *Uniform Convergence over $\Theta^\dagger$.*

For every $\theta_0 \in \Theta^\dagger$,

$$\sup_{\theta \in \Theta^\dagger} \left| \frac{1}{T} \sum_{t=1}^T \ln \pi(X_t | X_{t-1}, \theta) - \mathbb{E}_{\theta_0} \{ \ln \pi(X_1^* | X_0^*, \theta) \} \right| \xrightarrow{P_{\theta_0, x_0}} 0.$$

Proof of Lemma A.3.

From the series representation of $\mathbf{I}_\nu(a)$,

$$\begin{aligned} \mathbf{I}_\nu(a) &= \sum_{k=0}^{\infty} \frac{(a/2)^{2k+\nu}}{k! \Gamma(k + \nu + 1)} \leq \frac{(a/2)^\nu}{\Gamma(\nu + 1)} \left(1 + \frac{1}{\nu + 1} \sum_{k=1}^{\infty} \frac{(a/2)^{2k}}{k! (k-1)!} \right) \\ &\leq \frac{(a/2)^\nu}{\Gamma(\nu + 1)} \left(1 + \frac{a}{2(\nu + 1)} \sum_{k=1}^{\infty} \frac{a^{2k-1}}{(2k-1)!} \right) \\ &= \frac{(a/2)^\nu}{\Gamma(\nu + 1)} \left(1 + \frac{a \sinh a}{2(\nu + 1)} \right) \leq \frac{(a/2)^\nu}{\Gamma(\nu + 1)} \left(1 + \frac{a}{2(\nu + 1)} \right) e^a, \end{aligned}$$

for all $\nu > -1$ and $a > 0$. The first inequality keeps the $k = 0$ term, equal to $(a/2)^\nu / \Gamma(\nu + 1)$, and uses $\Gamma(k + \nu + 1) = \Gamma(\nu + 1)(\nu + 1)(\nu + 2) \cdots (\nu + k) \geq \Gamma(\nu + 1)(\nu + 1)(k-1)!$ for $k \geq 1$, since $\nu + j > j - 1$. The second uses $k! (k-1)! \geq (2k-1)! / 2^{2k-1}$, since the binomial coefficient $\binom{2k-1}{k}$ is at most 2^{2k-1} . The last uses $1 \leq e^a$ and $\sinh a \leq e^a$. Then,

$$\sup_{\theta \in \Theta^\dagger} \left| \ln \pi(y | x, \theta) \right| \leq C_{\Theta^\dagger} (1 + y + x + |\ln y| + |\ln x|),$$

where every term is integrable under the invariant density of (X_1^*, X_0^*) . The result follows from the application of Newey and McFadden (1994) Lemma 2.4, where $\Theta^\dagger$ is assumed compact, the transition density is continuous at all $\theta \in \Theta^\dagger$, the integrable bound is given above and the i.i.d. assumption is replaced with the process being stationary ergodic. $\square$

Lemma A.4 *Non-singular Fisher Information.*

For every $\theta_0 \in \Theta^*$, $\mathbf{K}_{\theta_0}$ is positive definite and, in particular, non-singular.

Proof of Lemma A.4.

As in the proof of Theorem 2, let $\nu = \frac{n}{2} - 1$, $a = \sqrt{yx\rho}/\tau$, $Z \sim \mathcal{Bessel}(\nu, a)$ and $\mu = \mathbb{E}\{Z\}$. Differentiating the series representation of $\mathbf{I}_\nu(a)$ term by term gives $\frac{\partial}{\partial \nu} \ln \mathbf{I}_\nu(a) = \ln(a/2) - \mathbb{E}\{\psi(Z + \nu + 1)\}$. This, $a \frac{\partial}{\partial a} \ln \mathbf{I}_\nu(a) = \nu + 2\mu$, $\frac{\partial a}{\partial \tau} = -\frac{a}{\tau}$ and $\frac{\partial a}{\partial \rho} = \frac{a}{2\rho}$ applied to the log density of the transition function give the score in closed form,

$$\begin{aligned}\frac{\partial}{\partial \tau} \ln \pi(y | x, \theta) &= -\frac{n}{2\tau} + \frac{y + \rho x}{2\tau^2} - \frac{2\mu}{\tau} \\ \frac{\partial}{\partial \rho} \ln \pi(y | x, \theta) &= -\frac{x}{2\tau} + \frac{\mu}{\rho} \\ \frac{\partial}{\partial \frac{n}{2}} \ln \pi(y | x, \theta) &= \ln y - \ln(2\tau) - \mathbb{E}\{\psi(Z + \nu + 1)\}.\end{aligned}$$

From here on, the score is evaluated at θ_0 and, to lighten notation, $(\tau, \rho, \frac{n}{2})$ stands for $(\tau_0, \rho_0, \frac{n_0}{2})$.

$$\mathbf{u}^\top \nabla_\theta \ln \pi(y | x, \theta) = u_\tau \frac{\partial}{\partial \tau} \ln \pi(y | x, \theta) + u_\rho \frac{\partial}{\partial \rho} \ln \pi(y | x, \theta) + u_{n/2} \frac{\partial}{\partial \frac{n}{2}} \ln \pi(y | x, \theta).$$

Since $\mathbf{K}_{\theta_0}$ is finite by the proof of Theorem 2,

$$\mathbf{u}^\top \mathbf{K}_{\theta_0} \mathbf{u} = \mathbb{E}_{\theta_0} \left\{ \left(\mathbf{u}^\top \nabla_\theta \ln \pi(X_1^* | X_0^*, \theta_0) \right)^2 \right\} \geq 0.$$

Assume $\mathbf{u}^\top \mathbf{K}_{\theta_0} \mathbf{u} = 0$. Then $\mathbf{u}^\top \nabla_\theta \ln \pi(y | x, \theta_0) = 0$ almost everywhere under the invariant density $\pi_{\theta_0}(y | x) \times \pi_{\theta_0}(x)$, which is positive on $(0, \infty)^2$. The score is continuous in (y, x) on $(0, \infty)^2$, since μ and $\mathbb{E}\{\psi(Z + \nu + 1)\}$ are continuous in a . Therefore the set where $\mathbf{u}^\top \nabla_\theta \ln \pi(y | x, \theta_0) \neq 0$ is open and has probability zero, so it is empty, that is,

$$\mathbf{u}^\top \nabla_\theta \ln \pi(y | x, \theta_0) = 0 \quad \text{for all } x, y > 0. \quad (4)$$

Next, we show the scores to be linearly independent. Since $\mathbf{u}$ does not depend on (x, y) and equation 4 holds for every $y > 0$, it also holds in the limit $y \rightarrow 0$ at any fixed $x > 0$, where the scores take an explicit form.

Fix $x > 0$ and let $y \rightarrow 0$, so that $a \rightarrow 0$. By the proof of Lemma A.1, $k(k + \nu)\mathbb{P}(Z = k) = \frac{a^2}{4}\mathbb{P}(Z = k - 1)$ for $k \geq 1$. Since $k + \nu \geq \nu + 1 = \frac{n}{2}$, summing $k\mathbb{P}(Z = k) \leq \frac{a^2}{2n}\mathbb{P}(Z = k - 1)$ over $k \geq 1$ gives $\mu \leq \frac{a^2}{2n}$. Moreover, $\psi(z + 1) = \psi(z) + 1/z$ gives $0 \leq \psi(k + \nu + 1) - \psi(\nu + 1) = \sum_{i=0}^{k-1} \frac{1}{\nu+1+i} \leq \frac{2k}{n}$, so $0 \leq \mathbb{E}\{\psi(Z + \nu + 1)\} - \psi\left(\frac{n}{2}\right) \leq \frac{2\mu}{n}$. Hence $\mu \rightarrow 0$, $\mathbb{E}\{\psi(Z + \nu + 1)\} \rightarrow \psi\left(\frac{n}{2}\right)$ and, as $y \rightarrow 0$,

$$\begin{aligned} \frac{\partial}{\partial \tau} \ln \pi(y | x, \boldsymbol{\theta}_0) &\rightarrow -\frac{n}{2\tau} + \frac{\rho x}{2\tau^2}, & \frac{\partial}{\partial \rho} \ln \pi(y | x, \boldsymbol{\theta}_0) &\rightarrow -\frac{x}{2\tau}, \\ \frac{\partial}{\partial \frac{n}{2}} \ln \pi(y | x, \boldsymbol{\theta}_0) &= \ln y - \ln(2\tau) - \psi\left(\frac{n}{2}\right) + o(1) &\rightarrow -\infty. \end{aligned}$$

By equation 4, $u_{n/2} \frac{\partial}{\partial \frac{n}{2}} \ln \pi(y | x, \boldsymbol{\theta}_0) = -u_\tau \frac{\partial}{\partial \tau} \ln \pi(y | x, \boldsymbol{\theta}_0) - u_\rho \frac{\partial}{\partial \rho} \ln \pi(y | x, \boldsymbol{\theta}_0)$, where the right-hand side remains bounded as $y \rightarrow 0$ while $\frac{\partial}{\partial \frac{n}{2}} \ln \pi(y | x, \boldsymbol{\theta}_0) \rightarrow -\infty$. Hence $u_{n/2} = 0$. Letting $y \rightarrow 0$ in equation 4 then gives, for every $x > 0$,

$$u_\tau \left(\frac{\rho x}{2\tau^2} - \frac{n}{2\tau} \right) - u_\rho \frac{x}{2\tau} = 0.$$

The left-hand side is affine in x , so its intercept, $-u_\tau \frac{n}{2\tau}$, and its slope, $u_\tau \frac{\rho}{2\tau^2} - u_\rho \frac{1}{2\tau}$, are both zero. Therefore $u_\tau = 0$ and then $u_\rho = 0$, that is, $\mathbf{u} = \mathbf{0}$. $\square$

Proof of Theorem 3.

Consistency is achieved by satisfying the requirements of Theorem 5.7 in Van der Vaart (1998). Lemma A.3 shows uniform convergence of the sample likelihood and Lemma A.2 shows the CIR process to be globally identified, which in turn implies a unique maximum at θ_0 of the population likelihood, see Lemma 5.35 in Van der Vaart (1998). Compactness of $\Theta^\dagger$, continuity of the population likelihood and a unique maximum imply a *well-separated* maximum. Therefore $\hat{\theta}_T$ is consistent.

Asymptotic normality is achieved by satisfying the requirements of Theorem 5.39 in Van der Vaart (1998). For almost every (y, x) , $\ln \pi(y | x, \theta)$ is differentiable at θ_0 by the smoothness of the non-central χ^2 transition density. For the local Lipschitz condition, let $\theta_1, \theta_2 \in \bar{U}(\theta_0, \epsilon)$. By the mean value theorem and Cauchy–Schwarz,

$$|\ln \pi(y | x, \theta_1) - \ln \pi(y | x, \theta_2)| \leq \left(\sup_{\theta \in \bar{U}(\theta_0, \epsilon)} \|\nabla_{\theta} \ln \pi(y | x, \theta)\| \right) \|\theta_1 - \theta_2\|,$$

and the envelope is bounded by compactness of $\bar{U}(\theta_0, \epsilon)$ and the integrable bound obtained in the verification of condition D3 in Theorem 2. The envelope is also square integrable under the invariant density of (X_1^*, X_0^*) , since it involves second and squared logarithmic moments of the Gamma distribution. Lastly, Lemma A.4 implies the Fisher information matrix is non-singular. Steps where the i.i.d. assumption is employed can be replaced by conditions D1, D2 and D3. Therefore,

$$\sqrt{T} \left(\hat{\theta}_T - \theta_0 \right) \rightarrow \mathcal{N} \left(0, \mathbf{K}_{\theta_0}^{-1} \right) \quad P_{\theta_0, x_0}\text{-weakly.} \quad \square$$

Proof of Theorem 4.

By Van der Vaart (1998) Theorem 10.1, the Bernstein-von Mises Theorem holds provided that (i) the model is LAN at θ_0 with a non-singular Fisher information matrix, (ii) there exist uniformly consistent tests for θ_0 and the prior is absolutely continuous in a neighborhood of θ_0 with a continuous positive density at θ_0 (assumption 4 of Theorem 4). Condition (i) is shown in Theorem 2 and condition (ii) follows from Lemmas 10.6 and 10.4 in Van der Vaart (1998), with Lemma A.2 providing global identification, continuity of $\pi(y | x, \theta)$ in θ , $\Theta^\dagger$ compact and the CIR process being geometrically ergodic uniformly over $\Theta^\dagger$ replacing the i.i.d. assumption. $\square$

B Posterior Distribution with a Spike-and-Slab Prior for ρ

The derivation here follows closely that of Geweke (2005), Section 5.3.2.

Assume the prior for ρ is $\mathbb{P}(\rho = 1 \mid \Delta, x_0) = \underline{p}$ and if $\rho \neq 1$ then $\rho \mid \Delta, x_0 \sim \mathcal{U}(0, 1)$. So,

$$\pi(\rho \mid \Delta, x_0) = \underline{p}\delta_1(\rho) + (1 - \underline{p})\mathbb{I}\{0 < \rho < 1\}.$$

From the above equation it follows that if $\rho \neq 1$ then,

$$\rho \mid \tilde{x}_T, \Delta, \tilde{h}_T, \tau, \rho \neq 1 \sim \mathcal{G}(\bar{a}, \bar{b}) \mathbb{I}\{0 < \rho < 1\},$$

where $\bar{a}$ and $\bar{b}$ are given in Proposition 1 with $\underline{a} = 1$ and $\underline{b} = \infty$. Furthermore,

$$\begin{aligned} \pi(\rho = 1 \mid \tilde{x}_T, \Delta, \tilde{h}_T, \tau) &\propto \underline{p} \exp \left\{ -\frac{1}{2\tau} \sum_{t=0}^{T-1} x_t \right\} \quad \text{and} \\ \pi(\rho \mid \rho \neq 1, \tilde{x}_T, \Delta, \tilde{h}_T, \tau) &\propto (1 - \underline{p}) \int_0^1 \rho^{\sum_{t=1}^T h_t} \exp \left\{ -\left(\frac{1}{2\tau} \sum_{t=0}^{T-1} x_t \right) \rho \right\} d\rho \\ &\propto (1 - \underline{p}) \bar{b}^{\bar{a}} \gamma(\bar{a}, 1/\bar{b}). \end{aligned}$$

Thus,

$$\mathbb{P}(\rho = 1 \mid \tilde{x}_T, \Delta, \tilde{h}_T, \tau) = \frac{\underline{p} \exp \left\{ -\frac{1}{2\tau} \sum_{t=0}^{T-1} x_t \right\}}{\underline{p} \exp \left\{ -\frac{1}{2\tau} \sum_{t=0}^{T-1} x_t \right\} + (1 - \underline{p}) \bar{b}^{\bar{a}} \gamma(\bar{a}, 1/\bar{b})}.$$

C Alternative Simulation Schemes

In this appendix we develop and compare three alternative simulation schemes.

C.1 Collapsed Gibbs Sampler

In this subsection we develop a collapsed Gibbs sampler that iterates over the following steps:

1. Update $\tau^{(m+1)}$ with a tailored independent Metropolis-Hastings step that leaves the collapsed conditional $\pi(\tau \mid \tilde{x}_T, \rho^{(m)}, (n/2)^{(m)})$ invariant.
2. Sample $(\tilde{s}_T^{(m+1)}, \tilde{h}_T^{(m+1)})$ from $\pi(\tilde{s}_T \mid \tilde{h}_T, (n/2)^{(m)}) \times \pi(\tilde{h}_T \mid \tilde{x}_T, \tau^{(m+1)}, \rho^{(m)}, (n/2)^{(m)})$.
3. Sample $(\rho^{(m+1)}, (n/2)^{(m+1)})$ from $\pi(\rho \mid \tilde{x}_T, \tilde{h}_T^{(m+1)}, \tau^{(m+1)}) \times \pi(\frac{n}{2} \mid \tilde{x}_T, \tilde{s}_T^{(m+1)}, \tau^{(m+1)})$.

Steps 2 and 3 are identical to those detailed in Proposition 1. Step 1 is the collapsed step as the contribution of $\tilde{h}_T$ is marginalized. The conditional posterior density $\pi(\tau \mid \tilde{x}_T, \rho, (n/2))$ is given by

$$\pi(\tau \mid \tilde{x}_T, \rho, \frac{n}{2}) \propto \tau^{-(\alpha+T+1)} \exp\{-\bar{\beta}/\tau\} \prod_{t=1}^T \mathbf{I}_{n/2-1}\left(\frac{\sqrt{\rho x_t x_{t-1}}}{\tau}\right).$$

For the modified Bessel function of the first kind, $\mathbf{I}_\nu(a)$, with ν fixed and $a \rightarrow \infty$, the Hankel expansion gives

$$\mathbf{I}_\nu(a) \sim \frac{e^a}{\sqrt{2\pi a}} \sum_{k \geq 0} \frac{(-1)^k b_k(\nu)}{a^k} \quad a \rightarrow \infty,$$

where $b_0(\nu) = 1$ and $b_k(\nu) = (4\nu^2 - 1^2)(4\nu^2 - 3^2) \cdots (4\nu^2 - (2k-1)^2)/(k!8^k)$. Employing the first two terms of the expansion and $1 - z \approx e^{-z}$ for small z , the posterior density of τ

can be approximated as $\tau \rightarrow 0$ by

$$\pi\left(\tau \mid \tilde{x}_T, \rho, \frac{n}{2}\right) \approx \tau^{-(\underline{\alpha}+T/2+1)} \exp \left\{ -\tau \frac{(n-1)(n-3)}{8} \sum_{t=1}^T \frac{1}{\sqrt{\rho x_t x_{t-1}}} - \frac{1}{\tau} \left(\underline{\beta} + \frac{1}{2} \sum_{t=1}^T (\sqrt{x_t} - \sqrt{\rho x_{t-1}})^2 \right) \right\}.$$

The previous equation suggests that for small values of τ , the collapsed posterior distribution of τ is approximated by a Generalized Inverse Gaussian (GIG) distribution. $\mathcal{GIG}(\bar{a}, \bar{b}, \bar{p})$ with parameters $\bar{p} = -(\underline{\alpha} + T/2)$, $\bar{a} = \max \left\{ \frac{(n-1)(n-3)}{4} \sum_{t=1}^T 1/\sqrt{\rho x_t x_{t-1}}, 0 \right\}$ and $\bar{b} = 2\underline{\beta} + \sum_{t=1}^T (\sqrt{x_t} - \sqrt{\rho x_{t-1}})^2$. For the collapsed step we will employ a tailored independent Metropolis-Hastings algorithm, where the proposal will be given by the previous GIG. Given the current value $\theta^{(m)}$, generate a candidate $\tau^\dagger \sim \mathcal{GIG}(\bar{a}, \bar{b}, \bar{p})$ and set $\tau^{(m+1)} = \tau^\dagger$ with probability

$$\begin{aligned} \alpha\left(\theta^{(m)}, \tau^\dagger\right) &= \min \left\{ 1, \frac{\pi\left(\tau^\dagger \mid \tilde{x}_T, \rho^{(m)}, (n/2)^{(m)}\right) f_{\mathcal{GIG}}\left(\tau^{(m)} \mid \bar{a}, \bar{b}, \bar{p}\right)}{\pi\left(\tau^{(m)} \mid \tilde{x}_T, \rho^{(m)}, (n/2)^{(m)}\right) f_{\mathcal{GIG}}\left(\tau^\dagger \mid \bar{a}, \bar{b}, \bar{p}\right)} \right\} \\ &= \min \left\{ 1, \left(\frac{\tau^\dagger}{\tau^{(m)}} \right)^{-T/2} \exp \left\{ \frac{\bar{a}}{2} (\tau^\dagger - \tau^{(m)}) \right\} \prod_{t=1}^T \frac{\tilde{\mathbf{I}}_{(n/2)^{(m)}-1}(\sqrt{\rho^{(m)} x_t x_{t-1}} / \tau^\dagger)}{\tilde{\mathbf{I}}_{(n/2)^{(m)}-1}(\sqrt{\rho^{(m)} x_t x_{t-1}} / \tau^{(m)})} \right\}, \end{aligned}$$

where $f_{\mathcal{GIG}}$ is a GIG density and $\tilde{\mathbf{I}}_\nu(a) = e^{-a} \mathbf{I}_\nu(a)$. Otherwise, set $\tau^{(m+1)} = \tau^{(m)}$.

C.2 Random-Walk Metropolis-Hastings

In this subsection we develop a random-walk Metropolis-within-Gibbs sampler that iterates over the following steps:

1. Sample $\tilde{h}_T^{(m+1)}$ from $\pi(\tilde{h}_T \mid \tilde{x}_T, \tau^{(m)}, \rho^{(m)}, (n/2)^{(m)})$.
2. Update $(\tau^{(m+1)}, \rho^{(m+1)}, (n/2)^{(m+1)})$ with a random-walk Metropolis-Hastings move that leaves $\pi(\tau, \rho, \frac{n}{2} \mid \tilde{x}_T, \tilde{h}_T^{(m+1)})$ invariant.

Step 1 is the Bessel conditional of the latent counts derived in Proposition 1. Step 2 replaces the three conjugate draws of Proposition 1 by a single joint update of $\theta = (\tau, \rho, n/2)$. The target density satisfies,

$$\begin{aligned} \ln \pi\left(\tau, \rho, \frac{n}{2} \mid \tilde{x}_T, \tilde{h}_T\right) &\propto \frac{n}{2} \sum_{t=1}^T \ln x_t - \left(\frac{n}{2}T + 2 \sum_{t=1}^T h_t\right) \ln(2\tau) - \frac{1}{2\tau} \sum_{t=1}^T (x_t + \rho x_{t-1}) \\ &\quad + \ln \rho \sum_{t=1}^T h_t - \sum_{t=1}^T \ln \Gamma\left(\frac{n}{2} + h_t\right) + \ln \pi\left(\tau, \rho, \frac{n}{2} \mid x_0\right), \end{aligned}$$

and the log-prior decomposes as

$$\ln \pi\left(\tau, \rho, \frac{n}{2} \mid x_0\right) = -(\underline{\alpha} + 1) \ln \tau - \frac{\underline{\beta}}{\tau} + (\underline{a} - 1) \ln \rho - \frac{\rho}{\underline{b}} + \underline{\nu} \left[\frac{n}{2} \ln \underline{\lambda} - \ln \Gamma\left(\frac{n}{2}\right) \right],$$

matching the $\mathcal{JG}(\underline{\alpha}, \underline{\beta})$, $\mathcal{G}(\underline{a}, \underline{b})$ and $\mathcal{GC1}(\underline{\lambda}, \underline{\nu})$ priors of Proposition 1.

The three parameters are restricted to be strictly positive, $\tau, \rho, n/2 > 0$. To run an unconstrained random walk we reparametrize employing a componentwise logarithm,

$$\xi = (\ln \tau, \ln \rho, \ln(n/2))^{\top}.$$

The reparametrized target involves the Jacobian $|\partial\theta/\partial\xi| = \tau \rho (n/2)$, so that

$$\ln \tilde{\pi}(\xi) = \ln \pi(\theta(\xi) | \tilde{x}_T, \tilde{h}_T) + \ln \tau + \ln \rho + \ln (n/2).$$

Given the current state $\xi^{(m)}$, we draw a symmetric Gaussian random-walk candidate $\xi^\dagger = \xi^{(m)} + \varepsilon$, with $\varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$, and set $\theta^{(m+1)} = \exp(\xi^\dagger)$, componentwise, with probability

$$\min \left\{ 1, \exp \left\{ \ln \tilde{\pi}(\xi^\dagger) - \ln \tilde{\pi}(\xi^{(m)}) \right\} \right\},$$

and set $\theta^{(m+1)} = \theta^{(m)}$ otherwise. To lower the autocorrelation, the move is repeated 10 times within each sweep. The proposal covariance Σ is fixed in a preliminary trial run by adaptive Metropolis-Hastings, cf. Andrieu and Thoms (2008), so as to target an average acceptance probability of 0.234, the asymptotically optimal rate, cf. Roberts et al. (1997) and Roberts and Rosenthal (2001). Adaptation is confined to the trial run. After the trial run Σ is held fixed, so the chain is a standard non-adaptive Metropolis-within-Gibbs.

C.3 Hybrid Sampler

The hybrid sampler updates the scale τ by employing the collapsed independent Metropolis-Hastings step of Section C.1, and the remaining parameters $(\rho, n/2)$ are updated with the random-walk move of Section C.2. It iterates over the following steps:

1. Draw $\tau^{(m+1)}$ with the collapsed step of Section C.1.
2. Sample $\tilde{h}_T$ from $\pi(\tilde{h}_T | \tilde{x}_T, \tau^{(m+1)}, \rho^{(m)}, (n/2)^{(m)})$.
3. Update $(\rho^{(m+1)}, (n/2)^{(m+1)})$ with a random-walk Metropolis-Hastings move on $\xi = (\ln \rho, \ln(n/2))$ that leaves $\pi(\rho, \frac{n}{2} | \tilde{x}_T, \tilde{h}_T^{(m+1)}, \tau^{(m+1)})$ invariant.

Step 3 is the parameter block of Section C.2 restricted to $(\rho, n/2)$, holding τ and $\tilde{h}_T$ fixed. As the random walk is two-dimensional, the proposal covariance is tuned to an acceptance probability of about 0.35, the finite-dimensional optimum for $d = 2$ in Roberts and Rosenthal (2001).

C.4 Comparison

To compare the different simulation schemes, they are all executed on the simulated example of Section 2.1. Table 1 reports the effective sample size (ESS) per minute for the Gibbs sampler of Proposition 1 and the three alternative schemes. Figure 10 displays the autocorrelation functions of the draws of the three alternative schemes; those of the Gibbs sampler are displayed in Figure 3.

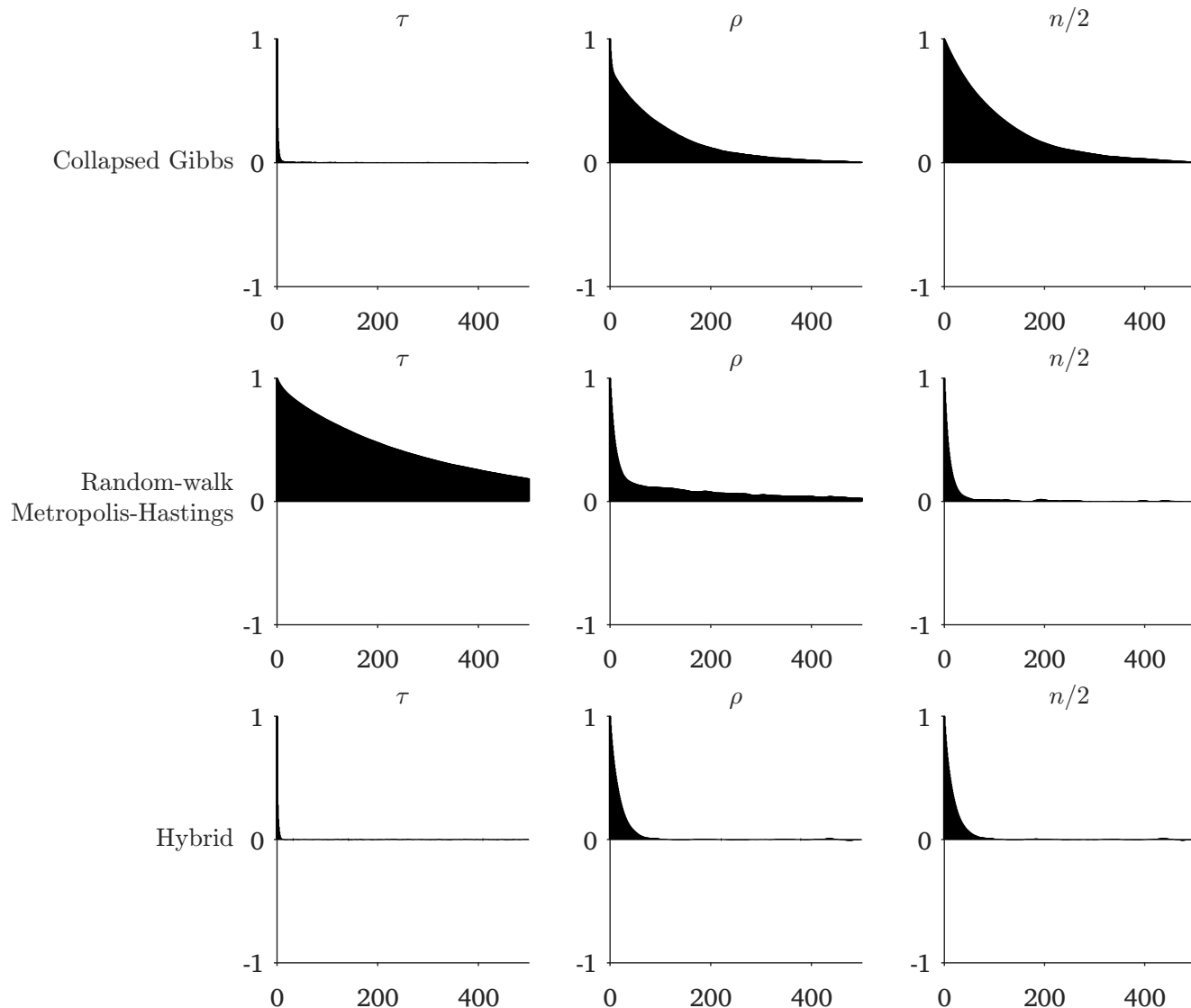


Figure 10: Autocorrelation plots of θ under the alternative simulation schemes.

References

- Ahn, Dong-Hyun, and Bin Gao. (1999). "A parametric nonlinear model of term structure dynamics". *The Review of Financial Studies*, 12(4), pp. 721-762. <https://doi.org/10.1093/rfs/12.4.721>
- Ait-Sahalia, Yacine. (1996). "Testing continuous-time models of the spot interest rate". *The Review of Financial Studies*, 9(2), pp. 385-426. <https://doi.org/10.1093/rfs/9.2.385>
- Andrieu, Christophe, and Johannes Thoms. (2008). "A tutorial on adaptive MCMC". *Statistics and Computing*, 18(4), pp. 343-373. <https://doi.org/10.1007/s11222-008-9110-y>
- Ben Alaya, Mohamed, and Ahmed Kebaier. (2012). "Parameter estimation for the squareroot diffusions: Ergodic and nonergodic cases". *Stochastic Models*, 28(4), pp. 609-634. <https://doi.org/10.1080/15326349.2012.726042>
- Ben Alaya, Mohamed, and Ahmed Kebaier. (2013). "Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions". *Stochastic Analysis and Applications*, 31(4), pp. 552-573. <https://doi.org/10.1080/07362994.2013.798175>
- Bauwens, Luc, and Pierre Giot. (2000). "The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks". *Annales d'Economie et de Statistique*, 60, pp. 117-149. <https://doi.org/10.2307/20076257>
- Berger, James O., and Ruo-Yong Yang. (1994). "Noninformative priors and Bayesian testing for the AR(1) model". *Econometric Theory*, 10(3-4), pp. 461-482. <https://doi.org/10.1017/S026646660000863X>
- Bielecki, Tomasz R., Monique Jeanblanc and Marek Rutkowski. (2011). "Hedging of a credit default swaption in the CIR default intensity model". *Finance and Stochastics*, 15(3), pp. 541-572. <https://doi.org/10.1007/s00780-010-0143-7>
- Box, George E. P., and George C. Tiao. (1973). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons. <https://doi.org/10.1002/9781118033197>
- Brunnermeier, Markus K., Sebastian Merkel and Yuliy Sannikov. (2024). "Safe assets". *Journal of Political Economy*, 132(11), pp. 3603-3657. <https://doi.org/10.1086/730547>
- Bühlmann, Hans, and Alois Gisler. (2005). *A Course in Credibility Theory and its Applications* (1st ed.). Springer Berlin, Heidelberg. <https://doi.org/10.1007/3-540-29273-X>
- Cheng, Yuzhong, Nicole Hufnagel and Hiroki Masuda. (2024). "Estimation of ergodic square-root diffusion under high-frequency sampling". *Econometrics and Statistics*, 32, pp. 73-87. <https://doi.org/10.1016/j.ecosta.2022.05.003>
- Chib, Siddhartha. (1995). "Marginal likelihood from the Gibbs output". *Journal of the American Statistical Association*, 90(432), pp. 1313-1321. <https://doi.org/10.1080/01621459.1995.10476635>
- Chib, Siddhartha, and Bakhodir Ergashev. (2009). "Analysis of multifactor affine yield curve models". *Journal of the American Statistical Association*, 104(488), pp. 1324-1337. <https://doi.org/10.1198/jasa.2009.ap08029>
- Cox, John C., Jonathan E. Ingersoll and Stephen A. Ross. (1985). "A theory of the term structure of interest rates". *Econometrica*, 53(2), pp. 385-407. <https://doi.org/10.2307/1911242>

- Creal, Drew D. (2017). "A class of non-Gaussian state space models with exact likelihood inference". *Journal of Business & Economic Statistics*, 35(4), pp. 585-597. <https://doi.org/10.1080/07350015.2015.1092977>
- Damsleth, Eivind. (1975). "Conjugate classes for gamma distributions". *Scandinavian Journal of Statistics*, 2(2), pp. 80-84. <https://www.jstor.org/stable/4615580>
- Devroye, Luc. (2002). "Simulating Bessel random variables". *Statistics & probability letters*, 57(3), pp. 249-257. [https://doi.org/10.1016/S0167-7152\(02\)00055-X](https://doi.org/10.1016/S0167-7152(02)00055-X)
- Devroye, Luc. (2014). "Random variate generation for the generalized inverse Gaussian distribution". *Statistics and Computing*, 24, pp. 239-246. <https://doi.org/10.1007/s11222-012-9367-z>
- Durrett, Rick. (1991). *Probability: Theory and Examples* (5th ed.). Cambridge University Press. <https://doi.org/10.1017/9781108591034>
- Elerian, Ola, Siddhartha Chib and Neil Shephard. (2001). "Likelihood inference for discretely observed nonlinear diffusions". *Econometrica*, 69, pp. 959-993. <https://doi.org/10.1111/1468-0262.00226>
- Eraker, Bjørn. (2001). "MCMC analysis of diffusion models with application to finance". *Journal of Business & Economic Statistics*, 19(2), pp. 177-191. <https://doi.org/10.1198/073500101316970403>
- Feller, William. (1951a). "Diffusion processes in genetics". In Jerzy Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp. 227-247. <https://doi.org/10.1525/9780520411586-018>
- Feller, William. (1951b). "Two singular diffusion problems". *Annals of Mathematics*, 54(1), pp. 173-182. <https://doi.org/10.2307/1969318>
- Feller, William. (1975). *An introduction to probability theory and its applications*. John Wiley & Sons. <https://www.climet.com/toolbox/feller-correction-calculator/Feller-1968.pdf>
- Gelman, Andrew, David A. van Dyk, Zaiying Huang and John W. Boscardin. (2008). "Using redundant parameterizations to fit hierarchical models". *Journal of Computational and Graphical Statistics*, 17(1), pp. 95-122. <https://doi.org/10.1198/106186008X287337>
- Geweke, John. (2005). *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons. <https://doi.org/10.1002/0471744735>
- Geyer, Charles J. (1992). "Practical Markov Chain Monte Carlo". *Statistical Science*, 7(4), pp. 473-483. <https://doi.org/10.1214/ss/1177011137>
- Gourieroux, Christian, and Joann Jasiak. (2006). "Autoregressive gamma processes". *Journal of Forecasting*, 25(2), pp. 129-152. <https://doi.org/10.1002/for.978>
- Gourieroux, Christian, and Alain Monfort. (1995). *Statistics and Econometric Models*, 1. Cambridge University Press. <https://doi.org/10.1017/CBO9780511751967>
- Guo, Bin, Fuzhe Huang and Kai Li. (2022). "Time to build and bond risk premia". *Journal of Economic Dynamics and Control*, 136(104080). <https://doi.org/10.1016/j.jedc.2021.104080>
- Hamilton, James D. (1994). *Time Series Analysis*. Princeton University Press. <https://doi.org/10.1515/9780691218632>

- Hamura, Yasuyuki, Kaoru Irie and Shonosuke Sugasawa. (2023). "On data augmentation for models involving reciprocal gamma functions". *Journal of Computational and Graphical Statistics*, 32(3), pp. 908-916. <https://doi.org/10.1080/10618600.2022.2119988>
- Hefter, Mario, and Arnulf Jentzen. (2019). "On arbitrarily slow convergence rates for strong numerical approximations of Cox-Ingersoll-Ross processes and squared Bessel processes". *Finance and Stochastics*, 23, pp. 139-172. <https://doi.org/10.1007/s00780-018-0375-5>
- Heston, Steven L. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options". *The Review of Financial Studies*, 6(2), pp. 327-343. <https://doi.org/10.1093/rfs/6.2.327>
- Hiraki, Daichi, Siddhartha Chib and Yasuhiro Omori. (2025). "Stochastic volatility in mean: Efficient analysis by a generalized mixture sampler". *Journal of Econometrics*, 256-B(105949). <https://doi.org/10.1016/j.jeconom.2025.105949>
- Ibragimov, Ildar Abdulovich, and Rafail Zalmanovich Hasminskii. (1981). *Statistical estimation: Asymptotic theory*, 16. Springer New York. <https://doi.org/10.1007/978-1-4899-0027-2>
- Iliopoulos, George, and Dimitris Karlis. (2003). "Simulation from the bessel distribution with applications". *Journal of Statistical Computation and Simulation*, 73(7), pp. 491-506. <https://doi.org/10.1080/0094965031000123124>
- Jagannathan, Ravi, Andrew Kaplin and Steve Sun. (2003). "An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices". *Journal of Econometrics*, 116(1-2), pp. 113-146. [https://doi.org/10.1016/S0304-4076\(03\)00105-2](https://doi.org/10.1016/S0304-4076(03)00105-2)
- Kerman, Jouni. (2011). "Neutral noninformative and informative conjugate beta and gamma prior distributions". *Electronic Journal of Statistics*, 5, pp. 1450-1470. <https://doi.org/10.1214/11-EJS648>
- León-González, Roberto. (2019). "Efficient Bayesian inference in generalized inverse gamma processes for stochastic volatility". *Econometric Reviews*, 38(8), pp. 899-920. <https://doi.org/10.1080/07474938.2018.1485614>
- Lin, Xiaoji, Chong Wang, Neng Wang and Jinqiang Yang. (2018). "Investment, Tobin's q , and interest rates". *Journal of Financial Economics*, 130(3), pp. 620-640. <https://doi.org/10.1016/j.jfineco.2017.05.013>
- Makarov, Roman N., and Devin Glew. (2010). "Exact simulation of Bessel diffusions". *Monte Carlo Methods and Applications*, 16(3-4), pp. 283-306. <https://doi.org/10.1515/mcma.2010.010>
- Muirhead, Robb J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons. <https://doi.org/10.1002/9780470316559>
- Newey, Whitney K., and Daniel McFadden. (1994). "Chapter 36 Large sample estimation and hypothesis testing". In *Handbook of Econometrics*, 4. Elsevier, pp. 2111-2245. [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4)
- Overbeck, Ludger. (1998). "Estimation for continuous branching processes". *Scandinavian Journal of Statistics*, 25(1), pp. 111-126. <https://doi.org/10.1111/1467-9469.00092>
- Overbeck, Ludger, and Tobias Ryden. (1997). "Estimation in the Cox-Ingersoll-Ross model". *Econometric Theory*, 13(3), pp. 430-461. <https://doi.org/10.1017/S0266466600005880>

- Phillips, P. C. B. (1991). "To criticize the critics: An objective bayesian analysis of stochastic trends". *Journal of Applied Econometrics*, 6(4), pp. 333-364. <https://doi.org/10.1002/jae.3950060402>
- Pitman, Jim, and Marc Yor. (1982). "A decomposition of Bessel Bridges". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59, pp. 425-457. <https://doi.org/10.1007/BF00532802>
- Revuz, Daniel, and Marc Yor. (1991). *Continuous Martingales and Brownian Motion* (3rd ed.). Springer Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-06400-9>
- Roberts, Gareth O., Andrew Gelman and Walter R. Gilks. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms". *The Annals of Applied Probability*, 7(1), pp. 110-120. <https://doi.org/10.1214/aoap/1034625254>
- Roberts, Gareth O., and Jeffrey S. Rosenthal. (2001). "Optimal scaling for various Metropolis–Hastings algorithms". *Statistical Science*, 16(4), pp. 351-367. <https://doi.org/10.1214/ss/1015346320>
- Roy, Vivekananda, Kshitij Khare and James P. Hobert. (2024). "The data augmentation algorithm". [Preprint]. <https://doi.org/10.48550/arXiv.2406.10464>
- Schotman, Peter, and Herman K. van Dijk. (1991). "A Bayesian analysis of the unit root in real exchange rates". *Journal of Econometrics*, 49(1-2), pp. 195-238. [https://doi.org/10.1016/0304-4076\(91\)90014-5](https://doi.org/10.1016/0304-4076(91)90014-5)
- Sims, Christopher A. (1988). "Bayesian skepticism on unit root econometrics". *Journal of Economic Dynamics and Control*, 12(1-2), pp. 463-474. [https://doi.org/10.1016/0165-1889\(88\)90050-4](https://doi.org/10.1016/0165-1889(88)90050-4)
- Stanton, Richard. (1997). "A nonparametric model of term structure dynamics and the market price of interest rate risk". *The Journal of Finance*, 52(5), pp. 1973-2002. <https://doi.org/10.1111/j.1540-6261.1997.tb02748.x>
- Van der Vaart, Aad W. (1998). *Asymptotic Statistics*. (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press. <https://doi.org/10.1017/CBO9780511802256>
- Vasicek, Oldrich. (1977). "An equilibrium characterization of the term structure". *Journal of Financial Economics*, 5(2), pp. 177-188. [https://doi.org/10.1016/0304-405X\(77\)90016-2](https://doi.org/10.1016/0304-405X(77)90016-2)
- Wachter, Jessica A. (2013). "Can time-varying risk of rare disasters explain aggregate stock market volatility?". *The Journal of Finance*, 68(3), pp. 987-1035. <https://doi.org/10.1111/jofi.12018>
- Yuan, Lin, and John D. Kalbfleisch. (2000). "On the Bessel distribution and related problems". *Annals of the Institute of Statistical Mathematics*, 52, pp. 438-447. <https://doi.org/10.1023/A:1004152916478>

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

- 2530 MIGUEL ACOSTA-HENAO, MARÍA ALEJANDRA AMADO, MONTSERRAT MARTÍ and DAVID PÉREZ-REYNA: Heterogeneous UIPDs across Firms: Spillovers from U.S. Monetary Policy Shocks.
- 2531 LUIS HERRERA and JESÚS VÁZQUEZ: Learning from news.
- 2532 MORTEZA GHOMI, JOCHEN MANKART, RIGAS OIKONOMOU and ROMANOS PRIFTIS: Debt maturity and government spending multipliers.
- 2533 MARINA DIAKONOVA, CORINNA GHIRELLI and JAVIER J. PÉREZ: Political polarization in Europe.
- 2534 NICOLÁS FORTEZA and SERGIO PUENTE: Measuring non-workers' labor market attachment with machine learning.
- 2535 GERGELY GANICS and LLUC PUIG CODINA: Simple Tests for the Correct Specification of Conditional Predictive Densities.
- 2536 HENRIQUE S. BASSO and OMAR RACHEDI: Robot adoption and inflation dynamics.
- 2537 PABLO GARCIA, PASCAL JACQUINOT, ČRT LENARČIČ, KOSTAS MAVROMATIS, NIKI PAPADOPOULOU and EDGAR SILGADO-GÓMEZ: Green transition in the Euro area: domestic and global factors.
- 2538 MARÍA ALEJANDRA AMADO, CARLOS BURGA and JOSÉ E. GUTIÉRREZ: Cross-border spillovers of bank regulations: Evidence of a trade channel.
- 2539 ALEJANDRO CASADO and DAVID MARTÍNEZ-MIERA: Banks' specialization and private information.
- 2540 CHRISTIAN E. CASTRO, ÁNGEL ESTRADA GARCÍA and GONZALO FERNÁNDEZ DIONIS: Diversifying sovereign risk in the Euro area: empirical analysis of different policy proposals.
- 2541 RAFAEL GUNTIN and FEDERICO KOCHEN: The Origins of Top Firms.
- 2542 ÁLVARO FERNÁNDEZ-GALLARDO: Natural disasters, economic activity, and property insurance: evidence from weekly U.S. state-level data.
- 2543 JOSÉ ELÍAS GALLEGOS, ESTEBAN GARCÍA-MIRALLES, IVÁN KATARYNIUK and SUSANA PÁRRAGA RODRÍGUEZ: Fiscal Announcements and Households' Beliefs: Evidence from the Euro Area.
- 2544 LUIS HERRERA, MARA PIROVANO and VALERIO SCALONE: From risk to buffer: Calibrating the positive neutral CCyB rate.
- 2545 ESTEBAN GARCÍA-MIRALLES et al.: Fiscal drag in theory and in practice: A European perspective.
- 2546 TATSURO SENGU and IACOPO VAROTTO: Investment Irreversibility in a Granular World.
- 2547 OLYMPIA BOVER, NEZIH GUNER, YULIYA KULIKOVA, ALESSANDRO RUGGIERI and CARLOS SANZ: Family-friendly policies and fertility: What firms have to do with it?
- 2548 ADINA-ELENA FUDULACHE and MARIA DEL CARMEN CASTILLO LOZOYA: Demand drivers of central bank liquidity: A time-to-exit TLTRO analysis.
- 2549 ERIK ANDRES-ESCAYOLA, LUIS MOLINA, JAVIER J. PÉREZ and ELENA VIDAL: How economic policy uncertainty spreads across borders: the case of Latin America.
- 2550 MATTHIAS BURGERT, MATTHIEU DARRACQ PARIÈS, LUIGI DURAND, MARIO GONZÁLEZ, ROMANOS PRIFTIS, OKE RÖHE, MATTHIAS ROTTNER, EDGAR SILGADO-GÓMEZ, NIKOLAI STÄHLER and JANOS VARGA: Macroeconomic effects of carbon-intensive energy price changes: A model comparison.
- 2601 IACOPO VAROTTO: Blocking the Blockers? Diversity Matters.
- 2602 CARLOS CAÑIZARES MARTÍNEZ, ADRIANA LOJSCHOVÁ and ALICIA AGUILAR: Non-linear effects of monetary policy shocks on housing: Evidence from a CESEE country.
- 2603 DIEGO BONELLI: Inflation risk and yield spread changes.
- 2604 MARÍA ALEJANDRA AMADO: Macroprudential FX Regulations and Small Firms: Unintended Consequences for Credit Growth.
- 2605 FERNANDO ÁVALOS, BORIS HOFMANN and JOSE M. SERENA: Monetary policy and private equity acquisitions: tracing the links.
- 2606 RICARD GREBOL, MARGARITA MACHELETT, JAN STUHLER and ERNESTO VILLANUEVA: Assortative Mating, Inequality, and Rising Educational Mobility in Spain.
- 2607 PABLO AGUILAR, RUBÉN DOMÍNGUEZ-DÍAZ, JOSÉ-ELÍAS GALLEGOS and JAVIER QUINTANA: The Transmission of Foreign Shocks in a Networked Economy.
- 2608 ERWAN GAUTIER, CRISTINA CONFLITTI, DANIEL ENDERLE, LUDMILA FADEJEVA, ALEX GRIMAUD, EDUARDO GUTIÉRREZ, VALENTIN JOUVANCEAU, JAN-OLIVER MENZ, ALARI PAULUS, PAVLOS PETROULAS, PAU ROLDAN-BLANCO and ELISABETH WIELAND: Consumer price stickiness in the euro area during an inflation surge.
- 2609 MORTEZA GHOMI and SAMUEL HURTADO: RAUI: Uncertainty Indicators Built With Artificial Intelligence.
- 2610 MORTEZA GHOMI and EVI PAPPAS: Stimulating avenues: EIB loans and returns to Public Investment.

- 2611 JUAN S. MORA-SANGUINETTI, CRISTINA PEÑASCO and ROK SPRUK: The impact of "Green Regulation" on firms' innovation.
- 2612 ÁLVARO FERNÁNDEZ-GALLARDO and EVI PAPP: Natural disasters and fiscal shelters.
- 2613 NICOLÁS BONINO-GAYOSO and MÓNICA CORREA-LÓPEZ: Unexpecting the expected in real-time inflation forecasting: The inflation expectations channel?
- 2614 ANDRÉS AZQUETA-GAVALDÓN, MARINA DIAKONOVA, CORINNA GHIRELLI and JAVIER J. PÉREZ: Diverging signals from economic uncertainty measures: Uncovering coherence through news narratives.
- 2615 JOSÉ E. GUTIÉRREZ, ENRIC MARTORELL and MARIYA MELNYCHUK: The nexus between the deposit and risk-taking channels of monetary policy.
- 2616 MARTA GARCÍA-RODRÍGUEZ and CLEMENTE PINILLA-TORREMOCHA: The role of confidence measures in European unemployment dynamics.
- 2617 DANIEL DEJUAN-BITRIA: Processing forward-looking loan loss provisions: evidence from the adoption of the CECL model.
- 2618 RYAN BANERJEE, FRANCISCO GONZÁLEZ, JOSÉ E. GUTIÉRREZ and JOSE-MARÍA SERENA: Credit Supply in the Wake of Distressed Bank Acquisitions.
- 2619 ALEJANDRO CASADO, MARCO GIOMETTI, JOSÉ E. GUTIÉRREZ, DAVID MARTÍNEZ-MIERA, ALEXANDRA MATYUNINA and TAMMARO TERRACCiano: Digitalization and credit markets: Evidence from e-Invoicing.
- 2620 JORGE ABAD, SAKI BIGIO, SALOMÓN GARCÍA-VILLEGAS, JOËL MARBET and GALO NUÑO: The Heterogeneous Bank Lending Channel of Monetary Policy.
- 2621 NICOLÁS FORTEZA and JOSÉ M. LABEAGA: Pollution, Density and Low Emission Zones: European Evidence.
- 2622 ANDREA DE POLIS, ÁLVARO FERNÁNDEZ-GALLARDO and JOSÉ NICOLÁS ROSAS: Oil supply shocks and inflation tail risks.
- 2623 LUIS HERRERA, CATERINA MENDICINO, KALIN NIKOLOV and VALERIO SCALONE: Monetary Tightening, Financial Stability and the Role of Macroprudential Policy.
- 2624 MARINA GÓMEZ-GARCÍA: The impact of air pollution on the housing market: The case of Madrid.
- 2625 ANA AGUILAR, RAFAEL GUERRA, CAROLA MÜLLER and ALEXANDRE TOMBINI: The macro-financial impact of economic policy uncertainty in Latin America.
- 2626 ANDREA DE POLIS, LEONARDO MELOSI and IVAN PETRELLA: The Taming of the Skew: Asymmetric Inflation Risk and Monetary Policy.
- 2627 EOGHAN O'NEILL and SOFIA VELASCO: Let the Tree Decide: FABART. A Non-parametric factor model for nonlinear oil shock transmission.
- 2628 CRAIG A. CHIKIS, JONATHAN GOLDBERG and DAVID LÓPEZ-SALIDO: The innovator's risk premium: Sticky hurdle rates, the cost of capital, and creative destruction.
- 2629 LUCAS TER STEEGE and SOFIA VELASCO: What volatility reveals: agnostic identification of exchange rate and Treasury-market dynamics in inconvenient times.
- 2630 CALLUM JONES, DAVID LÓPEZ-SALIDO and THOMAS PHILIPPON: Is growth additive?
- 2631 JACOB T. HESS and XUFENG WANG: Measuring Merger Effects with Revenue Data.
- 2632 PABLO AGUILAR, MATTHIEU DARRACQ PARIÈS, ALISTAIR DIEPPE, RUBÉN DOMÍNGUEZ-DÍAZ, ANTONIO EUGENIO, JOSÉ-ELÍAS GALLEGOS and JAVIER QUINTANA: Tariffs, production networks, and spillovers: the case of a US-China Trade War.
- 2633 LUCA BARBAGLIA, MARINA DIAKONOVA, LUIS DÍEZ-CATALÁN, CORINNA GHIRELLI, ALICE LIA, JAVIER J. PÉREZ and STEFANO VERZILLO: Moving Out: Geoeconomic Policy and the Relocation of Foreign Affiliates.
- 2634 IRINA BALTEANU, MAXIMILIANO MORENO and FRANCESCA VIANI: Flooded by China? Trade diversion due to US tariffs during Trump's first mandate.
- 2635 PIERLAURO LOPEZ, FRANCISCO VÁZQUEZ-GRANDE and DAVID LÓPEZ-SALIDO: A New Keynesian model of the term structures of equity and bond returns.
- 2636 MARINA DIAKONOVA, JULIA ESTEFANIA-FLORES, CORINNA GHIRELLI, ANA PEÑALOSA, JAVIER J. PÉREZ and AITANA SIMÓ: Political Polarization and Legislative Gridlock: A Comparative High-Frequency Approach.
- 2637 MARIO ALLOZA, RUBÉN DOMÍNGUEZ-DÍAZ, PAU DURÁ and IACOPO VAROTTO: The macroeconomic effects of defence expenditure: Evidence from Spain.
- 2638 SANTIAGO ALVAREZ-BLASER: Inflation and Price Dispersion: New Cross-Sectoral and International Evidence.
- 2639 DAVID CUBERES, AITOR LACUESTA, CARLOS MORENO-PÉREZ, and DANIEL OTO-PERALÍAS: Land concentration and large renewable energy projects.
- 2640 DANTE AMENGUAL and LLUC PUIG-CODINA: A Bayesian analysis of the Cox-Ingersoll-Ross model.