symmetry

MDPI

Article

Robust Bayesian Estimation of SARX Models Under Scale-Mixtures of Normal Errors: Application to Najran Solar Irradiance

Ayman A. Amin [1,2,*] and Shuhrah A. Alghamdi [3]

[1] Department of Mathematics, College of Arts and Sciences, Najran University, Najran 66462, Saudi Arabia
[2] Science and Engineering Research Center, Najran University, Najran 66462, Saudi Arabia
[3] Department of Mathematical Sciences, College of Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11564, Saudi Arabia; shaalghamdi@pnu.edu.sa
* Correspondence: aaabdellah@nu.edu.sa

Abstract

Seasonal autoregressive models with exogenous variables (SARX) combine a multiplicative seasonal autoregressive backbone with a distributed-lag regression on observed drivers. These models are the workhorse specification for series such as energy demand, tourism arrivals, and retail sales that are simultaneously seasonal and causally linked to external covariates. Nearly all existing Bayesian treatments of such models continue to impose the Gaussian error assumption, a convenient yet empirically fragile choice. Seasonal series with exogenous drivers are routinely exposed to compounding sources of tail risk—including seasonal spikes, extreme weather realisations, and ordinary measurement error—which together produce residuals with kurtosis far exceeding the Gaussian benchmark. We close this gap by developing a robust Bayesian estimation framework for SARX models whose innovations follow a scale-mixtures of normal (SMN) distribution. Exploiting the conditionally Gaussian hierarchical representation of the SMN family, we augment the parameter space with observation-specific latent precision weights. Under three prior families, i.e., normal-gamma, Zellner's g-prior, and Jeffreys' prior, we derive closed-form full conditional posterior distributions for the nonseasonal and seasonal autoregressive coefficient blocks, the exogenous regression coefficients, and the error scale. For each SMN special case, we further derive the conditional posteriors of the latent weights and shape parameters, some of which require embedding a Metropolis–Hastings step within the Gibbs sampler. We combine these results into a unified hybrid MCMC algorithm and assess its performance through a simulation study covering multiple SARX specifications and all non-Gaussian members of the SMN family. We further illustrate the methodology through a real application to hourly solar irradiance at Najran, Saudi Arabia, driven by the physically determined clear-sky ceiling.

Keywords: multiplicative SARX; exogenous regressors; SMN distribution; data augmentation; Gibbs sampler; Metropolis–Hastings; robust Bayesian inference

MSC: 62F15

Check for updates

Academic Editors: Qiang Xia, Kai Yang and Jiazhu Pan

Received: 18 October 2026
Revised: 19 November 2026
Accepted: 23 November 2026
Published: 25 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Seasonal time series that are additionally driven by observable external forces are ubiquitous in applied forecasting. For instance, energy demand responds to calendar seasonality and to temperature, retail sales respond to the yearly shopping calendar and to promotional or macroeconomic indicators, and solar irradiance responds to the astronomical day cycle

and to the clear-sky ceiling. The seasonal autoregressive model with exogenous variables (SARX) is the natural linear-dynamics device for such series. It couples a multiplicative seasonal autoregressive (AR) polynomial, of the kind popularised by Box et al. [1] for purely endogenous seasonal dynamics, with a distributed-lag regression on one or more exogenous drivers. Applied evidence for the practical value of this coupling is abundant and recent. De Castro and Montini [2] use a SARX specification to forecast Brazilian industrial electricity consumption from tariff, exchange-rate, and activity regressors. Alharbi and Csala [3] and Almuhaini and Sultana [4] both build seasonal exogenous-regression forecasts of the Saudi Arabian electricity sector. Serrano et al. [5] compare seasonal time-series specifications against machine-learning competitors for Brazilian energy demand using climatic and industrial exogenous variables, concluding that a well-specified seasonal-exogenous linear model remains highly competitive. In every one of these applications, the exogenous channel is precisely the transmission belt through which extreme, non-Gaussian shocks—heat waves, currency shocks, commodity-price spikes, and policy interventions—enter the series, compounding the tail risk that seasonal AR (SAR) dynamics already inherit from occasional regime shifts and structural breaks.

Despite this, the Bayesian literature on SARX-type models has been built almost exclusively on the assumption of Gaussian innovations. This is a substantive limitation, not merely a technical convenience. Under Gaussian errors, the likelihood penalises large residuals quadratically so that a handful of atypical seasonal or exogenous shocks can dominate the posterior and drag every coefficient estimate away from the bulk of the data. The resulting credible intervals are simultaneously too wide where the majority of the sample is concerned and too narrow to acknowledge the possibility of a further extreme observation. This behaviour is well documented in the univariate regression and autoregressive literatures. West [6] showed that Gaussian linear regression lacks whole robustness against outliers, motivating explicit heavy-tailed error models. Zellner [7] and Lange et al. [8] formalised the Student's t linear model as a Bayesian-robust alternative. Harvey and Luati [9] demonstrate, in a filtering context, that score-based robust updating driven by heavy-tailed conditional densities materially improves signal extraction relative to the Gaussian benchmark. The empirical domains in which SARX models are most used are exactly those in which such heavy tails have been repeatedly documented. Stock and Watson [10] and Stock and Watson [11] attribute part of the post-2000 difficulty of forecasting US inflation to outlier-generating episodes, and Tsay [12] catalogues pervasive excess kurtosis in economic and financial series. Once a SAR backbone is combined with exogenous regressors that themselves carry such shocks, the case for abandoning the Gaussian assumption is, if anything, strengthened rather than weakened.

Bayesian computation for such models has a long pedigree. Data augmentation [13] and the Gibbs sampler [14] freed Bayesian time-series analysis from reliance on analytic conjugacy. A substantial Bayesian literature has developed around SAR structures [15,16] and, more recently, of double and triple seasonal autoregressions [17], in all of which the error distribution is Gaussian. On the exogenous side, fully Bayesian treatments of the AR model with exogenous variables (ARX) are likewise almost exclusively Gaussian [18–20], and the few relaxations of the Gaussian assumption within ARX-type identification are recursive or online rather than fully Bayesian [21,22].

A theoretically coherent and computationally tractable route to heavy-tailed robustness is furnished by the scale-mixtures of normal (SMN) family, characterised by Andrews and Mallows [23] and further developed by West [24], and deployed for regression models by Fernandez and Steel [25]. An SMN random variable is obtained by scaling a Gaussian variate by the inverse square root of a positive, observation-specific latent weight, whose own distribution—the mixing distribution—governs the tail behaviour of the resulting

marginal law. Varying the mixing distribution recovers, among others, the Student's t distribution (mixing with a gamma weight), the slash distribution (mixing with a beta weight), and the two-point contaminated normal distribution (mixing with a Bernoulli-type weight), together with the Gaussian distribution itself as a degenerate special case in which the weight is unity. Crucially, conditioning on the latent weights restores an exactly Gaussian complete-data likelihood, which is what makes the SMN family so attractive for Gibbs-sampler-based Bayesian computation. Robustness is purchased not by abandoning conjugacy but by augmenting the parameter space. This device has already proved effective for partial-linear AR models [26], AR processes with SMN innovations [27], finite-mixture autoregressions [28], and Student's t-mixture autoregressions [29]. Recent research has established Bayesian frameworks for estimation, prediction, and subset selection under SMN innovations in SAR models lacking exogenous regressors [30,31] and in ARX models without seasonal structure [32,33]. The methodology developed herein builds most directly upon these contributions.

The literature and empirical applications leave a well-defined methodological gap. Seasonality and exogenous regression have each been combined with heavy-tailed SMN errors in separate strands of the autoregressive literature, while the applied SARX literature has combined seasonal and exogenous structures under Gaussian innovations. To the best of our knowledge, however, multiplicative seasonality, exogenous regression, and SMN robustness have not been brought together within a unified Bayesian estimation framework admitting closed-form posterior updating. Filling this gap is not a routine concatenation of two existing derivations. The multiplicative seasonal structure makes the SARX coefficient vector a nonlinear function of the primitive nonseasonal and seasonal autoregressive coefficients since their pairwise products enter the design matrix. Consequently, unlike in a conventional ARX–SMN model, the autoregressive coefficients cannot be updated jointly within a single conjugate block. At the same time, the exogenous regression coefficients enter the model linearly and interact with both autoregressive blocks through the residual structure. Constructing an efficient MCMC sampler therefore requires isolating three mutually dependent but conditionally conjugate parameter blocks: nonseasonal autoregressive coefficients, seasonal autoregressive coefficients, and exogenous regression coefficients. This decomposition must be embedded within a data-augmented weighted-Gaussian likelihood and implemented consistently under the three prior specifications and the SMN error cases. This is precisely the derivation we undertake.

The specific contributions of this paper are as follows.

(i) Under three prior families—normal-gamma, Zellner's g-prior, and Jeffreys' prior—we prove a general weighted-Gaussian conjugacy lemma and use it to derive closed-form full conditional posteriors for the nonseasonal autoregressive coefficients, the seasonal autoregressive coefficients, the exogenous regression coefficients, and the error scale parameter. This requires an exact block decomposition of the multiplicative mean function and a three-block scheme, which has no counterpart in ARX–SMN models, where a single block suffices. It also shows that the shape of the error scale update depends on the number of free parameters under the normal-gamma prior but on the design dimension under the g-prior.

(ii) For each of the three non-Gaussian SMN special cases (Student's t, slash, contaminated normal), we derive the full conditional posteriors of the latent mixing weights and shape parameters, formally identifying which are standard and which require a Metropolis–Hastings correction. For the contaminated normal, the update is built on contamination indicators, which gives an exact full conditional for the contamination parameter.

(iii) We assemble these results into a hybrid MCMC algorithm for Bayesian estimation of SARX models under SMN errors. We validate this algorithm through a simulation study spanning several SARX configurations and all non-Gaussian SMN cases, together with a real application to hourly solar irradiance at Najran, Saudi Arabia. We state the Metropolis–Hastings acceptance ratios explicitly, give an invariance and irreducibility argument, and examine the sensitivity of the results to the prior family, to the Beta hyperparameters, and to rarer and more severe contamination.

Relation to ARX–SMN and SAR–SMN.

The proposed estimation framework of SARX models is closest to the ARX–SMN framework of Amin and Alghamdi [32] and to the SAR–SMN framework of Amin [30,31]. Three ingredients are inherited from these works and are not claimed as new: the SMN hierarchy, the priors of the mixing parameters, and the conditionals of the latent weights and shape parameters, which depend on the model only through the residuals. What is new follows from combining a multiplicative seasonal polynomial with exogenous regressors. When no seasonal polynomial is present, the model reduces to the ARX–SMN case, but once the seasonal component is included the ARX–SMN derivation cannot be transferred because the coefficient vector no longer enters the mean function linearly. This changes the structure of the coefficient updates, the form of the update for the scale parameter, the treatment of contaminated observations, and the design of the sampler, so the seasonal and nonseasonal cases require a separate treatment rather than a direct extension of the existing ARX–SMN results.

The remainder of the paper is organised as follows. Section 2 introduces the SMN distribution and formalises the SARX model. Section 3 develops the augmented likelihood, the three prior families, and the joint posterior. Section 4 contains the main theoretical contribution of the full conditional posterior distributions. Section 5 presents the proposed hybrid MCMC algorithm. Section 6 reports the simulation study. Section 7 presents the real application. Section 8 concludes the paper.

2. SARX Models Under SMN Errors

We first summarise the SMN distribution and then formalise the SARX model and its notation.

2.1. Scale-Mixtures of Normal Distribution

Definition 1. *A random variable y follows a scale-mixtures of normal (SMN) distribution with location μ, scale σ^2, and mixing parameter $\boldsymbol{\nu}$, written $y \sim \mathrm{SMN}(\mu, \sigma^2; \boldsymbol{\nu})$, if it admits the hierarchical stochastic representation*

$$y \mid u \sim N(\mu,\, u^{-1}\sigma^2), \qquad u \sim p(u \mid \boldsymbol{\nu}), \tag{1}$$

where $u > 0$ is a latent precision-mixing variable, independent of the Gaussian component, whose distribution $p(u \mid \boldsymbol{\nu})$ is indexed by $\boldsymbol{\nu}$.

Small realisations of u inflate the conditional variance and generate heavy marginal tails, while u near unity recovers near-Gaussian behaviour. Integrating u out of (1) produces a marginal density whose tails are at least as heavy as those of the Gaussian distribution. Conversely, conditioning on u restores an exactly Gaussian likelihood, a property that makes the SMN family particularly attractive for data-augmented Gibbs sampling [23,24]. Three non-Gaussian members of the family are used throughout this paper, together with the Gaussian model itself as a limiting case.

- **SMN-t (Student's t).** $u \sim \text{Gamma}(\nu/2, \nu/2)$, $\nu > 2$, yields the Student's t_ν distribution; heavier tails correspond to smaller ν, with $\nu \to \infty$ recovering the Gaussian.
- **SMN-sl (Slash).** $u \sim \text{Beta}(\nu, 1)$, $\nu > 0$, yields the slash distribution, whose polynomial tail decay makes it suitable for series with severe, isolated outliers [24].
- **SMN-cn (Contaminated normal).** u is discrete with

$$p(u \mid \nu, \gamma) = \nu \, \mathbb{1}_{\{\gamma\}}(u) + (1-\nu) \, \mathbb{1}_{\{1\}}(u), \qquad \nu, \gamma \in (0,1), \tag{2}$$

  representing a clean Gaussian component (probability $1-\nu$, precision 1) contaminated with probability ν by an inflated-variance component of precision $\gamma < 1$. Equivalently, with contamination indicators $I_t = \mathbb{1}\{u = \gamma\} \overset{\text{iid}}{\sim} \text{Bernoulli}(\nu)$ and $u = \gamma^{I_t}$; this indicator form is used in Section 4.5.
- **SMN-n (Gaussian).** Setting $p(u = 1) = 1$ recovers the Gaussian model as the degenerate limiting member of the family.

Remark 1. *Because conditioning on u preserves a Gaussian complete-data likelihood, the SMN hierarchy converts a difficult marginal robustness problem into a standard weighted-least-squares problem given the latent weights. The entire Bayesian machinery developed in Sections 3–5 exploits exactly this property.*

2.2. Formulation of SARX Models

Let $\{y_t\}_{t=1}^n$ denote the response series, let s be a known seasonal period, and let $\{x_{k,t}\}_{t=1}^n$, $k = 1, \ldots, r$, denote r observed exogenous series. Writing B for the backshift operator, $B^m y_t = y_{t-m}$, define the nonseasonal and seasonal autoregressive polynomials of orders p and P,

$$\Phi_p(B) = 1 - \phi_{11} B - \phi_{12} B^2 - \cdots - \phi_{1p} B^p, \tag{3}$$

$$\Psi_P(B^s) = 1 - \phi_{21} B^s - \phi_{22} B^{2s} - \cdots - \phi_{2P} B^{Ps}. \tag{4}$$

Definition 2. *A time series $\{y_t\}$ generated by a $\text{SARX}(p, P; d_1, \ldots, d_r)_s$ model satisfies*

$$\Phi_p(B) \, \Psi_P(B^s) \, y_t = \xi_0 + \sum_{k=1}^{r} \sum_{j=0}^{d_k} \xi_{kj} \, x_{k,t-j} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \text{SMN}(0, \sigma^2; \boldsymbol{\nu}), \tag{5}$$

where ξ_0 is an intercept, $d_k \geq 0$ is the maximum lag of the k-th exogenous regressor, and ξ_{kj}, $j = 0, \ldots, d_k$, are its lag-j coefficients.

Expanding the product $\Phi_p(B)\Psi_P(B^s)$ in (5) and moving all autoregressive terms to the right-hand side gives the explicit form

$$y_t = \sum_{i=1}^{p} \phi_{1i} y_{t-i} + \sum_{j=1}^{P} \phi_{2j} y_{t-js} - \sum_{i=1}^{p} \sum_{j=1}^{P} \phi_{1i} \phi_{2j} y_{t-i-js} + \xi_0 + \sum_{k=1}^{r} \sum_{j=0}^{d_k} \xi_{kj} x_{k,t-j} + \varepsilon_t. \tag{6}$$

As in the pure SAR case, the pairwise products $\phi_{1i}\phi_{2j}$ that arise from expanding the multiplicative polynomial are exactly the interaction lags $s+1, \ldots, s+p, 2s+1, \ldots$ of the classical Box–Jenkins airline-type specification [1].

Collect the seasonal autoregressive design entries in

$$\boldsymbol{a}_t = \left(y_{t-1}, \ldots, y_{t-p}, \; y_{t-s}, y_{t-s-1}, \ldots, y_{t-s-p}, \; \ldots, \; y_{t-Ps}, \ldots, y_{t-Ps-p}\right)^\top \in \mathbb{R}^{p+P+pP}, \tag{7}$$

the exogenous design entries in

$$\boldsymbol{w}_t = \left(1,\ x_{1,t},\ldots,x_{1,t-d_1},\ \ldots,\ x_{r,t},\ldots,x_{r,t-d_r}\right)^\top \in \mathbb{R}^q, \qquad q = 1+\sum_{k=1}^{r}(d_k+1), \tag{8}$$

and set $\boldsymbol{x}_t = (\boldsymbol{a}_t^\top, \boldsymbol{w}_t^\top)^\top \in \mathbb{R}^m$, $m = p+P+pP+q$. The corresponding SARX coefficient vector is

$$\boldsymbol{\beta} = \left(\phi_{11},\ldots,\phi_{1p},\ \phi_{21},-\phi_{11}\phi_{21},\ldots,-\phi_{1p}\phi_{21},\ \ldots,\ \phi_{2P},\ldots,-\phi_{1p}\phi_{2P},\ \boldsymbol{\xi}^\top\right)^\top \in \mathbb{R}^m, \tag{9}$$

where $\boldsymbol{\xi} = (\xi_0,\xi_{10},\ldots,\xi_{1d_1},\ldots,\xi_{r0},\ldots,\xi_{rd_r})^\top \in \mathbb{R}^q$ collects the intercept and exogenous regression coefficients. $\boldsymbol{\phi}_2 = (\phi_{21},\ldots,\phi_{2P})^\top$ is the seasonal coefficients vector, paired with the nonseasonal coefficients vector $\boldsymbol{\phi}_1 = (\phi_{11},\ldots,\phi_{1p})^\top$. Model (6) then takes the compact form $y_t = \boldsymbol{x}_t^\top\boldsymbol{\beta} + \varepsilon_t$, and stacking observations $t = 1,\ldots,n$ gives the matrix representation

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{10}$$

where $\boldsymbol{X}$ is the $n \times m$ design matrix with t-th row $\boldsymbol{x}_t^\top$.

Remark 2 (Free parameters versus design dimension). *Because the entries of $\boldsymbol{\beta}$ associated with the cross lags $-\phi_{1i}\phi_{2j}$ are deterministic products of the primitive coefficients $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ rather than free parameters, model (6) contains only*

$$k \ = \ p+P+q \tag{11}$$

genuinely free structural parameters, strictly fewer than the design dimension $m = p+P+pP+q$ whenever $p, P \geq 1$. While this distinction has no practical consequence in the purely linear ARX case, where $m = k$, it becomes essential when deriving the full conditional posterior of σ^2 in Section 4.4.

Model (5) is stationary in its autoregressive part if and only if the roots of $\Phi_p(z) = 0$ and $\Psi_P(z^s) = 0$ lie outside the unit circle. Stationarity is governed exclusively by the autoregressive polynomials and is unaffected by the exogenous regressors [34,35].

3. Bayesian Framework of SARX Models

3.1. Augmented Likelihood

Under $\varepsilon_t \overset{\text{iid}}{\sim} \text{SMN}(0,\sigma^2;\boldsymbol{\nu})$, the marginal likelihood of $\boldsymbol{y}$ is

$$\mathcal{L}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\nu} \mid \boldsymbol{y}) \ \propto \ \prod_{t=1}^{n}\int_0^\infty u_t^{1/2}\, p(u_t \mid \boldsymbol{\nu})\,(\sigma^2)^{-1/2}\exp\Big\{-\frac{u_t}{2\sigma^2}\left(y_t-\boldsymbol{x}_t^\top\boldsymbol{\beta}\right)^2\Big\}du_t, \tag{12}$$

which is generally intractable because u_t and $\boldsymbol{\beta}$ are coupled inside the integral. Following the data-augmentation principle of Tanner and Wong [13], we treat the latent weight vector $\boldsymbol{u} = (u_1,\ldots,u_n)^\top$ as auxiliary data, giving the complete-data joint density

$$\mathcal{L}_c(\boldsymbol{\beta},\sigma^2,\boldsymbol{\nu} \mid \boldsymbol{y},\boldsymbol{u}) \ \propto \ (\sigma^2)^{-n/2}\exp\left\{-\frac{(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top\boldsymbol{U}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2}\right\}\prod_{t=1}^{n}\Big\{u_t^{1/2}\,p(u_t \mid \boldsymbol{\nu})\Big\}, \tag{13}$$

with $\boldsymbol{U} = \text{diag}(u_1,\ldots,u_n)$. The complete-data likelihood (13) is a heteroscedastic Gaussian likelihood with precision weights $\{u_t\}$ and is the foundation for the conjugate derivations of Section 4.

Remark 3 (Adaptive downweighting). *Writing $\omega_t = u_t/\sigma^2$ for the observation-specific effective precision, (13) shows that each residual is scaled by ω_t rather than by the common precision σ^{-2}. Observations with large standardised residuals $\frac{y_t-\boldsymbol{x}_t^\top\boldsymbol{\beta}}{\sigma}$ tend to generate small draws of u_t through*

the full conditional posterior distribution derived in Section 4.5. Consequently, the associated weights ω_t are also reduced in subsequent MCMC iterations, automatically downweighting the influence of these observations on $\boldsymbol{\beta}$.

3.2. Prior Distributions

Let $\Theta = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\nu})$ collect all model parameters, with hyperparameters $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$. We assume prior independence between the structural parameters and the SMN mixing parameters,

$$\zeta(\Theta \mid \boldsymbol{H}) = \zeta(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2 \mid \boldsymbol{H}_1) \times \zeta(\boldsymbol{\nu} \mid \boldsymbol{H}_2), \tag{14}$$

and, under the normal-gamma prior below, that $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}$ are a priori mutually independent given σ^2; the g-prior instead couples the three blocks through a single joint Gaussian law on $\boldsymbol{\beta}$, and under Jeffreys' prior no Gaussian structure is imposed.

Normal-gamma prior. When genuine prior information is available,

$$\begin{aligned} \boldsymbol{\phi}_1 \mid \sigma^2 &\sim N_p(\boldsymbol{\mu}_{\phi_1}, \sigma^2 \boldsymbol{\Sigma}_{\phi_1}), \quad \boldsymbol{\phi}_2 \mid \sigma^2 \sim N_P(\boldsymbol{\mu}_{\phi_2}, \sigma^2 \boldsymbol{\Sigma}_{\phi_2}), \\ \boldsymbol{\xi} \mid \sigma^2 &\sim N_q(\boldsymbol{\mu}_{\xi}, \sigma^2 \boldsymbol{\Sigma}_{\xi}), \qquad \sigma^2 \sim IG\Big(\tfrac{a_0}{2}, \tfrac{b_0}{2}\Big), \end{aligned} \tag{15}$$

with joint density

$$\begin{aligned} \zeta(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2 \mid \boldsymbol{H}_1) \;\propto\; & (\sigma^2)^{-\left(\frac{a_0+p+P+q}{2}+1\right)} \exp\Bigg\{ -\frac{1}{2\sigma^2}\Big[b_0 \\ & + (\boldsymbol{\phi}_1 - \boldsymbol{\mu}_{\phi_1})^\top \boldsymbol{\Sigma}_{\phi_1}^{-1} (\boldsymbol{\phi}_1 - \boldsymbol{\mu}_{\phi_1}) + (\boldsymbol{\phi}_2 - \boldsymbol{\mu}_{\phi_2})^\top \boldsymbol{\Sigma}_{\phi_2}^{-1} (\boldsymbol{\phi}_2 - \boldsymbol{\mu}_{\phi_2}) \\ & + (\boldsymbol{\xi} - \boldsymbol{\mu}_{\xi})^\top \boldsymbol{\Sigma}_{\xi}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}_{\xi}) \Big] \Bigg\}. \end{aligned} \tag{16}$$

Zellner's g-prior. When prior precision should scale with the information content of the design, we place a single joint g-prior on the *full* coefficient vector $\boldsymbol{\beta}$ of (9),

$$\zeta(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2 \mid \boldsymbol{H}_1) \;\propto\; (\sigma^2)^{-(m/2+1)} \exp\Big\{ -\frac{g}{2\sigma^2} (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^\top (\boldsymbol{X}^\top \boldsymbol{X}) (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) \Big\}, \tag{17}$$

where $\bar{\boldsymbol{\beta}}$ is a prior centre for $\boldsymbol{\beta}$ and $g > 0$ is typically set as $g = 1/n$ [36,37]. In this precision parameterisation, $g = 1/n$ corresponds to the unit-information choice.

Jeffreys' prior. In the absence of prior information,

$$\zeta(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2) \;\propto\; (\sigma^2)^{-1}, \qquad \sigma^2 > 0, \tag{18}$$

which is improper but yields a proper posterior provided $n > k$ and the conditional design matrices of Table 1 have full rank [38].

Choice among the three priors. The three families encode different amounts of prior information. The normal-gamma prior is the only one that lets beliefs be set separately for the nonseasonal, seasonal and exogenous blocks, and it gives σ^2 a proper prior, at the price of the hyperparameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, a_0, b_0)$. It suits settings in which earlier studies or physical arguments inform these blocks. The g-prior needs only $(g, \bar{\boldsymbol{\beta}})$, because its precision is built from $\boldsymbol{X}^\top \boldsymbol{X}$, and $g = 1/n$ makes it worth one observation. It penalises the full vector $\boldsymbol{\beta}$, however, so that $\bar{\boldsymbol{\beta}}$ enters the σ^2 update through $g\|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\bar{\boldsymbol{\beta}}\|^2$ (Theorem 2); a centre far from the fitted mean function inflates σ^2 when the series is not centred. Jeffreys' prior has no hyperparameters, so the analysis is entirely data driven and reproducible across replicates. We therefore use it as the default in the simulation study and the application, where no genuine prior information is available.

Table 1. Substitutions specialising Theorem 1 to each SARX coefficient block and prior family. Entries are listed in precision form $(\boldsymbol{\Sigma}_0^{-1}, \boldsymbol{\theta}_0, g, z_g)$, where $\mathbf{0}$ denotes a zero prior precision matrix and "–" an entry that is not used; NG, g, and Jeffreys refer to the normal-gamma, g-, and Jeffreys' priors of Section 3.2.

Block $\boldsymbol{\theta}$	Design $\boldsymbol{C}$	Response $\boldsymbol{z}$	Prior-Specific $(\boldsymbol{\Sigma}_0^{-1}, \boldsymbol{\theta}_0, g, z_g)$
$\boldsymbol{\phi}_1$ $(k = p)$	$\boldsymbol{Z}_{\phi_1}$	$\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2$	NG: $(\boldsymbol{\Sigma}_{\phi_1}^{-1}, \boldsymbol{\mu}_{\phi_1}, 0, -)$ g: $(\mathbf{0}, -, g, \boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2 - \boldsymbol{W}\boldsymbol{\xi})$ Jeffreys: $(\mathbf{0}, -, 0, -)$
$\boldsymbol{\phi}_2$ $(k = P)$	$\boldsymbol{Z}_{\phi_2}$	$\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1$	NG: $(\boldsymbol{\Sigma}_{\phi_2}^{-1}, \boldsymbol{\mu}_{\phi_2}, 0, -)$ g: $(\mathbf{0}, -, g, \boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1 - \boldsymbol{W}\boldsymbol{\xi})$ Jeffreys: $(\mathbf{0}, -, 0, -)$
$\boldsymbol{\xi}$ $(k = q)$	$\boldsymbol{W}$	$\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\phi}$	NG: $(\boldsymbol{\Sigma}_{\xi}^{-1}, \boldsymbol{\mu}_{\xi}, 0, -)$ g: $(\mathbf{0}, -, g, \boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{A}\boldsymbol{\phi})$ Jeffreys: $(\mathbf{0}, -, 0, -)$

Priors for the SMN mixing parameters. Adapting the work of Cabral et al. [39], we adopt, for each SMN case: (i) SMN-t: ν truncated-exponential with rate $\lambda > 0$ on $(2, \infty)$; (ii) SMN-sl: the same truncated-exponential structure on $(1, \infty)$; and (iii) SMN-cn: $\nu \sim \text{Beta}(\nu_0, \nu_1)$ and $\gamma \sim \text{Beta}(\gamma_0, \gamma_1)$ independently, with $\nu_0, \nu_1, \gamma_0, \gamma_1 > 0$. For (i) and (ii), write $a = 2$ and $a = 1$, respectively. The pair (ν, λ) has the joint prior $\zeta(\nu, \lambda) \propto \lambda e^{-\lambda\nu}\, \mathbb{1}_{(a,\infty)}(\nu)\, \mathbb{1}_{(c,d)}(\lambda)$, which is the kernel used in all computations of this paper. Normalising it gives $\nu \mid \lambda \sim a + \text{Exp}(\lambda)$, so that $\text{E}[\nu \mid \lambda] = a + 1/\lambda$, and a marginal prior $\zeta(\lambda) \propto e^{-a\lambda}$ on (c, d), i.e., a truncated exponential law rather than a uniform one.

3.3. Joint Posterior Distribution

Combining (13) with (14) yields the joint posterior of all unobservables,

$$\zeta(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2, \boldsymbol{u}, \nu \mid \boldsymbol{y}, \boldsymbol{H}) \propto \zeta(\Theta \mid \boldsymbol{H}) \times (\sigma^2)^{-n/2} \exp\left\{ -\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{U} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2} \right\} \times \prod_{t=1}^{n} \left\{ u_t^{1/2}\, p(u_t \mid \nu) \right\}. \tag{19}$$

Because $\boldsymbol{\beta}$ is a nonlinear function of $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$, i.e., bilinear through the cross terms $\phi_{1i}\phi_{2j}$, the joint posterior (19) is analytically intractable in $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ jointly, and the marginal posteriors of the SARX parameters admit no closed form. We resolve this by deriving the full conditional posteriors and driving a Gibbs sampler augmented by Metropolis–Hastings steps where required. Deriving the full conditional posteriors is the subject of the next section.

4. Full Conditional Posterior Distributions of SARX Models

This section derives every full conditional posterior distribution required by the hybrid MCMC algorithm introduced in Section 5.

4.1. Exact Decomposition and General Update Lemma of SARX Models

For a fixed value of $\boldsymbol{\phi}_2$, define the $n \times p$ matrix $\boldsymbol{Z}_{\phi_1}$ with (t, i) entry $z_{\phi_1, ti} = y_{t-i} - \sum_{j=1}^{P} \phi_{2j} y_{t-i-js}$ and the $n \times P$ matrix $\boldsymbol{L}_{\phi_1}$ with t-th row $\boldsymbol{l}_{\phi_1, t} = (y_{t-s}, \ldots, y_{t-Ps})$. Symmetrically, for a fixed value of $\boldsymbol{\phi}_1$, define the $n \times P$ matrix $\boldsymbol{Z}_{\phi_2}$ with (t, j) entry $z_{\phi_2, tj} = y_{t-js} - \sum_{i=1}^{p} \phi_{1i} y_{t-i-js}$ and the $n \times p$ matrix $\boldsymbol{L}_{\phi_2}$ with t-th row $\boldsymbol{l}_{\phi_2, t} = (y_{t-1}, \ldots, y_{t-p})$. Let $\boldsymbol{W}$ be the $n \times q$ matrix with t-th row $\boldsymbol{w}_t^\top$ as in (8), and let $\boldsymbol{A}\boldsymbol{\phi}$ denote the n-vector with t-th entry $\sum_{i=1}^{p} \phi_{1i} y_{t-i} + \sum_{j=1}^{P} \phi_{2j} y_{t-js} - \sum_{i=1}^{p} \sum_{j=1}^{P} \phi_{1i}\phi_{2j} y_{t-i-js}$, i.e., the autoregressive part

of (6) evaluated at the current $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$. $\boldsymbol{A\phi}$ is shorthand for this nonlinear map, the symbol $\boldsymbol{A}$ alone not denoting a matrix.

Proposition 1 (Exact block decomposition)**.** *For every* $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi})$,

$$\boldsymbol{X\beta} = \boldsymbol{Z}_{\phi_1}\boldsymbol{\phi}_1 + \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2 + \boldsymbol{W\xi}, \tag{20}$$

$$\boldsymbol{X\beta} = \boldsymbol{Z}_{\phi_2}\boldsymbol{\phi}_2 + \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1 + \boldsymbol{W\xi}, \tag{21}$$

$$\boldsymbol{X\beta} = \boldsymbol{A\phi} + \boldsymbol{W\xi}. \tag{22}$$

Proof. By (6) and (9), the t-th entry of $\boldsymbol{X\beta}$ equals

$$\sum_{i=1}^{p} \phi_{1i} y_{t-i} + \sum_{j=1}^{P} \phi_{2j} y_{t-js} - \sum_{i=1}^{p}\sum_{j=1}^{P} \phi_{1i}\phi_{2j} y_{t-i-js} + \xi_0 + \sum_{k=1}^{r}\sum_{j=0}^{d_k} \xi_{kj} x_{k,t-j},$$

which is $(\boldsymbol{A\phi})_t + \boldsymbol{w}_t^\top \boldsymbol{\xi}$, giving (22) directly. For (20), regroup the autoregressive part as

$$\sum_{i=1}^{p} \phi_{1i}\Big(y_{t-i} - \sum_{j=1}^{P} \phi_{2j} y_{t-i-js}\Big) + \sum_{j=1}^{P} \phi_{2j} y_{t-js} = \sum_{i=1}^{p} \phi_{1i}\, z_{\phi_1,ti} + \boldsymbol{l}_{\phi_1,t}^\top \boldsymbol{\phi}_2 = (\boldsymbol{Z}_{\phi_1}\boldsymbol{\phi}_1)_t + (\boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2)_t,$$

which is an algebraic identity, valid for every fixed $\boldsymbol{\phi}_2$; adding $\boldsymbol{w}_t^\top \boldsymbol{\xi}$ gives (20). Identity (21) follows by the symmetric regrouping with the roles of $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$ exchanged. □

Proposition 1 is the algebraic engine behind Sections 4.2–4.4. Because $\boldsymbol{X\beta}$ can be written exactly as an affine function of any one coefficient block with the other two held fixed, the sum of squares $(\boldsymbol{y} - \boldsymbol{X\beta})^\top \boldsymbol{U}(\boldsymbol{y} - \boldsymbol{X\beta})$ appearing in (19), and likewise the g-prior quadratic form $(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^\top (\boldsymbol{X}^\top \boldsymbol{X})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) = \|\boldsymbol{X\beta} - \boldsymbol{X}\bar{\boldsymbol{\beta}}\|^2$, reduce exactly to ordinary weighted quadratic forms in that block.

Lemma 1 (Weighted Gaussian conjugate update with prior and pseudo-data)**.** *Let* $\boldsymbol{\theta} \in \mathbb{R}^k$, *let* $\boldsymbol{C}$ *be a known* $n \times k$ *matrix, and let* $\boldsymbol{U}$ *be a known* $n \times n$ *diagonal, positive-definite weight matrix. Suppose*

$$\boldsymbol{z} = \boldsymbol{C\theta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{U}^{-1}), \tag{23}$$

and that $\boldsymbol{\theta}$ *is additionally assigned, a priori, a (possibly improper,* $\boldsymbol{\Sigma}_0^{-1} = \boldsymbol{0}$*) Gaussian prior* $\boldsymbol{\theta} \mid \sigma^2 \sim N_k(\boldsymbol{\theta}_0, \sigma^2 \boldsymbol{\Sigma}_0)$ *together with an optional Gaussian pseudo-observation (present iff* $g > 0$*)*

$$\boldsymbol{z}_g = \boldsymbol{C\theta} + \boldsymbol{\epsilon}_g, \qquad \boldsymbol{\epsilon}_g \sim N_n(\boldsymbol{0}, \sigma^2 g^{-1} \boldsymbol{I}_n), \tag{24}$$

independent of $\boldsymbol{\epsilon}$*. Then the conditional posterior of* $\boldsymbol{\theta}$ *given* $\boldsymbol{z}$, $\boldsymbol{z}_g$, σ^2, $\boldsymbol{U}$ *is*

$$\boldsymbol{\theta} \mid \cdot \;\sim\; N_k(\boldsymbol{\mu}^\star, \boldsymbol{V}^\star), \qquad \boldsymbol{V}^\star = \sigma^2\big[\boldsymbol{C}^\top(\boldsymbol{U} + g\boldsymbol{I}_n)\boldsymbol{C} + \boldsymbol{\Sigma}_0^{-1}\big]^{-1}, \tag{25}$$

$$\boldsymbol{\mu}^\star = \boldsymbol{V}^\star \sigma^{-2}\big[\boldsymbol{C}^\top \boldsymbol{U}\boldsymbol{z} + g\,\boldsymbol{C}^\top \boldsymbol{z}_g + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_0\big]. \tag{26}$$

Proof. By Bayes' theorem and the conditional independence of $\boldsymbol{z}$, $\boldsymbol{z}_g$, and the prior given $\boldsymbol{\theta}$,

$$\begin{aligned}\zeta(\boldsymbol{\theta} \mid \cdot) \;\propto\; & \exp\Big\{-\frac{1}{2\sigma^2}(\boldsymbol{z} - \boldsymbol{C\theta})^\top \boldsymbol{U}(\boldsymbol{z} - \boldsymbol{C\theta})\Big\} \exp\Big\{-\frac{g}{2\sigma^2}(\boldsymbol{z}_g - \boldsymbol{C\theta})^\top(\boldsymbol{z}_g - \boldsymbol{C\theta})\Big\} \\ & \times \exp\Big\{-\frac{1}{2\sigma^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\Big\}.\end{aligned}$$

Collecting the exponents and retaining only terms that depend on $\boldsymbol{\theta}$,

$$-2\sigma^2 \log \zeta(\boldsymbol{\theta} \mid \cdot) \quad \doteq \quad \boldsymbol{\theta}^\top \left[\boldsymbol{C}^\top \boldsymbol{U}\boldsymbol{C} + g\boldsymbol{C}^\top \boldsymbol{C} + \boldsymbol{\Sigma}_0^{-1}\right]\boldsymbol{\theta} - 2\boldsymbol{\theta}^\top \left[\boldsymbol{C}^\top \boldsymbol{U}\boldsymbol{z} + g\boldsymbol{C}^\top \boldsymbol{z}_g + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_0\right],$$

where $\doteq$ denotes equality up to an additive constant not involving $\boldsymbol{\theta}$. Writing $\boldsymbol{K} = \boldsymbol{C}^\top(\boldsymbol{U} + g\boldsymbol{I}_n)\boldsymbol{C} + \boldsymbol{\Sigma}_0^{-1}$ and $\boldsymbol{h} = \boldsymbol{C}^\top \boldsymbol{U}\boldsymbol{z} + g\boldsymbol{C}^\top \boldsymbol{z}_g + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_0$, this is $\boldsymbol{\theta}^\top \boldsymbol{K}\boldsymbol{\theta} - 2\boldsymbol{\theta}^\top \boldsymbol{h} = (\boldsymbol{\theta} - \boldsymbol{K}^{-1}\boldsymbol{h})^\top \boldsymbol{K}(\boldsymbol{\theta} - \boldsymbol{K}^{-1}\boldsymbol{h}) - \boldsymbol{h}^\top \boldsymbol{K}^{-1}\boldsymbol{h}$ by completing the square, so that

$$\zeta(\boldsymbol{\theta} \mid \cdot) \;\propto\; \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\theta} - \boldsymbol{K}^{-1}\boldsymbol{h})^\top \boldsymbol{K}(\boldsymbol{\theta} - \boldsymbol{K}^{-1}\boldsymbol{h})\right\},$$

which is the kernel of $N_k(\boldsymbol{K}^{-1}\boldsymbol{h}, \sigma^2\boldsymbol{K}^{-1})$. Identifying $\boldsymbol{\mu}^\star = \boldsymbol{K}^{-1}\boldsymbol{h}$ and $\boldsymbol{V}^\star = \sigma^2\boldsymbol{K}^{-1}$ gives (25)–(26). □

4.2. The Generic Block Posterior and Prior-Specific Substitutions

Theorem 1 (Generic block posterior)**.** *Let $\boldsymbol{\theta} \in \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}\}$ denote any one of the three SARX coefficient blocks, with the remaining two blocks fixed at their current MCMC values. Then, under any of the three prior families of Section 3.2, the full conditional posterior of $\boldsymbol{\theta}$ given $\boldsymbol{y}$, the remaining blocks, σ^2, and $\boldsymbol{U}$ is $N_k(\boldsymbol{\mu}^\star, \boldsymbol{V}^\star)$ with $(\boldsymbol{\mu}^\star, \boldsymbol{V}^\star)$ given by (25)–(26), upon substituting the block-specific $(\boldsymbol{C}, \boldsymbol{z})$ pair from Table 1 together with the prior-specific $(\boldsymbol{\Sigma}_0^{-1}, \boldsymbol{\theta}_0, g, \boldsymbol{z}_g)$ listed in the same table.*

Proof. Fix $\boldsymbol{\theta} = \boldsymbol{\phi}_1$, with $\boldsymbol{\phi}_2, \boldsymbol{\xi}$ held at their current values. By Proposition 1, (20), $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi}) - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2 - \boldsymbol{Z}_{\phi_1}\boldsymbol{\phi}_1$ so that the likelihood contribution to (19), as a function of $\boldsymbol{\phi}_1$ alone, is exactly $\exp\{-\frac{1}{2\sigma^2}(\boldsymbol{z} - \boldsymbol{Z}_{\phi_1}\boldsymbol{\phi}_1)^\top \boldsymbol{U}(\boldsymbol{z} - \boldsymbol{Z}_{\phi_1}\boldsymbol{\phi}_1)\}$ with $\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2$—precisely (23) with $\boldsymbol{C} = \boldsymbol{Z}_{\phi_1}$. Under the normal-gamma prior (15), the prior factor contributes exactly the Gaussian term of Lemma 1 with $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_{\phi_1}$, $\boldsymbol{\theta}_0 = \boldsymbol{\mu}_{\phi_1}$, $g = 0$. Under the g-prior (17), using $(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^\top \boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) = \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\bar{\boldsymbol{\beta}}\|^2$ and Proposition 1, $\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\bar{\boldsymbol{\beta}} = \boldsymbol{Z}_{\phi_1}\boldsymbol{\phi}_1 - (\boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2 - \boldsymbol{W}\boldsymbol{\xi})$, so the g-prior term is exactly (24) with $\boldsymbol{C} = \boldsymbol{Z}_{\phi_1}$ and $\boldsymbol{z}_g = \boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2 - \boldsymbol{W}\boldsymbol{\xi}$, while $\boldsymbol{\Sigma}_0^{-1} = \boldsymbol{0}$. Under Jeffreys' prior (18) both $\boldsymbol{\Sigma}_0^{-1} = \boldsymbol{0}$ and $g = 0$. Applying Lemma 1 in each case yields the stated result for $\boldsymbol{\phi}_1$. The derivations for $\boldsymbol{\theta} = \boldsymbol{\phi}_2$ and $\boldsymbol{\theta} = \boldsymbol{\xi}$ are identical upon using (21) and (22), respectively, in place of (20). □

Remark 4 (Blockwise Gaussianity under the g-prior)**.** *Under the g-prior (17), the penalty $(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^\top(\boldsymbol{X}^\top \boldsymbol{X})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) = \|\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\bar{\boldsymbol{\beta}}\|^2$ is quadratic—and the pseudo-observation representation (24) exact—in each coefficient block separately once the other two blocks are fixed, even though it is quartic jointly in $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ through the cross-products $\phi_{1i}\phi_{2j}$. This blockwise Gaussianity is what allows Lemma 1 to be applied under the g-prior exactly as under the conjugate priors, and it is a second reason—besides the likelihood itself—why the three coefficient blocks cannot be updated jointly in a single conjugate step.*

4.3. Conditional Posteriors of the SARX Coefficient Blocks

The three corollaries below instantiate Theorem 1, and they are implemented as Steps 1–3 of Algorithm 1 in Section 5.

Corollary 1 (Conditional posterior of $\boldsymbol{\phi}_1$)**.** $\boldsymbol{\phi}_1 \mid \cdot \sim N_p(\boldsymbol{\mu}_{\phi_1}^\star, \boldsymbol{V}_{\phi_1}^\star)$ *with, under the normal-gamma prior,*

$$\boldsymbol{V}_{\phi_1}^\star \;=\; \sigma^2\left(\boldsymbol{Z}_{\phi_1}^\top \boldsymbol{U}\boldsymbol{Z}_{\phi_1} + \boldsymbol{\Sigma}_{\phi_1}^{-1}\right)^{-1}, \tag{27}$$

$$\boldsymbol{\mu}_{\phi_1}^\star \;=\; \boldsymbol{V}_{\phi_1}^\star \sigma^{-2}\left[\boldsymbol{Z}_{\phi_1}^\top \boldsymbol{U}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2) + \boldsymbol{\Sigma}_{\phi_1}^{-1}\boldsymbol{\mu}_{\phi_1}\right]; \tag{28}$$

under the g-prior, with $\boldsymbol{G} = \boldsymbol{U} + g\boldsymbol{I}_n$,

$$\boldsymbol{V}^{\star}_{\phi_1} = \sigma^2\left(\boldsymbol{Z}^{\top}_{\phi_1}\boldsymbol{G}\boldsymbol{Z}_{\phi_1}\right)^{-1}, \tag{29}$$

$$\boldsymbol{\mu}^{\star}_{\phi_1} = \left(\boldsymbol{Z}^{\top}_{\phi_1}\boldsymbol{G}\boldsymbol{Z}_{\phi_1}\right)^{-1}\left[g\boldsymbol{Z}^{\top}_{\phi_1}\left(\boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2 - \boldsymbol{W}\boldsymbol{\xi}\right) + \boldsymbol{Z}^{\top}_{\phi_1}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2\right)\right]; \tag{30}$$

and under Jeffreys' prior,

$$\boldsymbol{V}^{\star}_{\phi_1} = \sigma^2\left(\boldsymbol{Z}^{\top}_{\phi_1}\boldsymbol{U}\boldsymbol{Z}_{\phi_1}\right)^{-1}, \qquad \boldsymbol{\mu}^{\star}_{\phi_1} = \left(\boldsymbol{Z}^{\top}_{\phi_1}\boldsymbol{U}\boldsymbol{Z}_{\phi_1}\right)^{-1}\boldsymbol{Z}^{\top}_{\phi_1}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_1}\boldsymbol{\phi}_2\right). \tag{31}$$

Proof. Immediate from Theorem 1 upon substituting the $\boldsymbol{\phi}_1$ row of Table 1 into (25)–(26). □

Corollary 2 (Conditional posterior of $\boldsymbol{\phi}_2$). $\boldsymbol{\phi}_2 \mid \cdot \sim N_P(\boldsymbol{\mu}^{\star}_{\phi_2}, \boldsymbol{V}^{\star}_{\phi_2})$, *with, under the normal-gamma prior,*

$$\boldsymbol{V}^{\star}_{\phi_2} = \sigma^2\left(\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{U}\boldsymbol{Z}_{\phi_2} + \boldsymbol{\Sigma}^{-1}_{\phi_2}\right)^{-1}, \tag{32}$$

$$\boldsymbol{\mu}^{\star}_{\phi_2} = \boldsymbol{V}^{\star}_{\phi_2}\sigma^{-2}\left[\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1\right) + \boldsymbol{\Sigma}^{-1}_{\phi_2}\boldsymbol{\mu}_{\phi_2}\right]; \tag{33}$$

under the g-prior, with $\boldsymbol{G} = \boldsymbol{U} + g\boldsymbol{I}_n$,

$$\boldsymbol{V}^{\star}_{\phi_2} = \sigma^2\left(\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{G}\boldsymbol{Z}_{\phi_2}\right)^{-1}, \tag{34}$$

$$\boldsymbol{\mu}^{\star}_{\phi_2} = \left(\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{G}\boldsymbol{Z}_{\phi_2}\right)^{-1}\left[g\boldsymbol{Z}^{\top}_{\phi_2}\left(\boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1 - \boldsymbol{W}\boldsymbol{\xi}\right) + \boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1\right)\right]; \tag{35}$$

and under Jeffreys' prior,

$$\boldsymbol{V}^{\star}_{\phi_2} = \sigma^2\left(\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{U}\boldsymbol{Z}_{\phi_2}\right)^{-1}, \qquad \boldsymbol{\mu}^{\star}_{\phi_2} = \left(\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{U}\boldsymbol{Z}_{\phi_2}\right)^{-1}\boldsymbol{Z}^{\top}_{\phi_2}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\xi} - \boldsymbol{L}_{\phi_2}\boldsymbol{\phi}_1\right). \tag{36}$$

Proof. Immediate from Theorem 1 upon substituting the $\boldsymbol{\phi}_2$ row of Table 1. □

Corollary 3 (Conditional posterior of $\boldsymbol{\xi}$). $\boldsymbol{\xi} \mid \cdot \sim N_q(\boldsymbol{\mu}^{\star}_{\xi}, \boldsymbol{V}^{\star}_{\xi})$, *with, under the normal-gamma prior,*

$$\boldsymbol{V}^{\star}_{\xi} = \sigma^2\left(\boldsymbol{W}^{\top}\boldsymbol{U}\boldsymbol{W} + \boldsymbol{\Sigma}^{-1}_{\xi}\right)^{-1}, \qquad \boldsymbol{\mu}^{\star}_{\xi} = \boldsymbol{V}^{\star}_{\xi}\sigma^{-2}\left[\boldsymbol{W}^{\top}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\phi}\right) + \boldsymbol{\Sigma}^{-1}_{\xi}\boldsymbol{\mu}_{\xi}\right]; \tag{37}$$

under the g-prior, with $\boldsymbol{G} = \boldsymbol{U} + g\boldsymbol{I}_n$,

$$\boldsymbol{V}^{\star}_{\xi} = \sigma^2\left(\boldsymbol{W}^{\top}\boldsymbol{G}\boldsymbol{W}\right)^{-1}, \ \boldsymbol{\mu}^{\star}_{\xi} = \left(\boldsymbol{W}^{\top}\boldsymbol{G}\boldsymbol{W}\right)^{-1}\left[g\boldsymbol{W}^{\top}\left(\boldsymbol{X}\bar{\boldsymbol{\beta}} - \boldsymbol{A}\boldsymbol{\phi}\right) + \boldsymbol{W}^{\top}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\phi}\right)\right] \tag{38}$$

and under Jeffreys' prior,

$$\boldsymbol{V}^{\star}_{\xi} = \sigma^2\left(\boldsymbol{W}^{\top}\boldsymbol{U}\boldsymbol{W}\right)^{-1}, \qquad \boldsymbol{\mu}^{\star}_{\xi} = \left(\boldsymbol{W}^{\top}\boldsymbol{U}\boldsymbol{W}\right)^{-1}\boldsymbol{W}^{\top}\boldsymbol{U}\left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\phi}\right). \tag{39}$$

Proof. Immediate from Theorem 1 upon substituting the $\boldsymbol{\xi}$ row of Table 1. □

Remark 5. *Corollary 3 shows that, conditionally on* $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$, *the exogenous-coefficient block* $\boldsymbol{\xi}$ *obeys exactly the same conjugate update as in an ordinary nonseasonal ARX–SMN regression of the autoregressive-adjusted response* $\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\phi}$, *in which the entire autoregressive component (nonseasonal, seasonal, and interaction lags) has been removed, on* $\boldsymbol{W}$. *The seasonal AR structure affects* $\boldsymbol{\xi}$*'s conditional posterior only through the working response, never through the design. This is a direct, and previously unavailable, generalisation of the ARX–SMN conjugate result to the seasonal setting.*

4.4. Conditional Posterior of the Scale Parameter

Theorem 2 (Conditional posterior of σ^2)**.** *Given $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \boldsymbol{u})$, the full conditional posterior of σ^2 is $IG(a/2, b/2)$, where we have the following:*

Normal-gamma prior:

$$\begin{aligned} a &= a_0 + n + k, && (40)\\ b &= b_0 + (\boldsymbol{\phi}_1 - \boldsymbol{\mu}_{\phi_1})^\top \boldsymbol{\Sigma}_{\phi_1}^{-1} (\boldsymbol{\phi}_1 - \boldsymbol{\mu}_{\phi_1}) + (\boldsymbol{\phi}_2 - \boldsymbol{\mu}_{\phi_2})^\top \boldsymbol{\Sigma}_{\phi_2}^{-1} (\boldsymbol{\phi}_2 - \boldsymbol{\mu}_{\phi_2}) \\ &\quad + (\boldsymbol{\xi} - \boldsymbol{\mu}_{\xi})^\top \boldsymbol{\Sigma}_{\xi}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}_{\xi}) + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{U} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}); && (41) \end{aligned}$$

g-prior:

$$a = n + m, \qquad b = g(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^\top (\boldsymbol{X}^\top \boldsymbol{X})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) + (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{U} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}); \tag{42}$$

Jeffreys' prior:

$$a = n, \qquad b = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{U} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{43}$$

Proof. Collect, from (19), every factor of σ^2. The likelihood contributes $(\sigma^2)^{-n/2} \exp\{-(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{U} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})/(2\sigma^2)\}$ regardless of the prior. Under the normal-gamma prior, (16) contributes $(\sigma^2)^{-(p+P+q)/2}$ from the three independent Gaussian blocks together with $(\sigma^2)^{-(a_0/2+1)} \exp\{-b_0/(2\sigma^2)\}$ from the $IG(a_0/2, b_0/2)$ prior on σ^2 itself, and the three quadratic forms in the exponent. Summing exponents, the total power of σ^2 is $-\big[n/2 + (p+P+q)/2 + a_0/2 + 1\big] = -(a_0 + n + k)/2 - 1$, which is the kernel of $IG(a/2, b/2)$ with $a = a_0 + n + k$ and b as in (41). Under the g-prior, (17) places a single Gaussian prior directly on the full vector $\boldsymbol{\beta} \in \mathbb{R}^m$, contributing $(\sigma^2)^{-m/2}$; combined with the likelihood's $(\sigma^2)^{-n/2}$, the total power is $-(n+m)/2 - 1$, giving $a = n + m$ as in (42). Under Jeffreys' prior (18), the likelihood alone contributes $(\sigma^2)^{-n/2}$, and the flat $(\sigma^2)^{-1}$ prior adds one further power, giving $a = n$ as in (43). □

Remark 6. *Theorem 2 is where the distinction of Remark 2 between the design dimension $m = p + P + pP + q$ and the number $k = p + P + q$ of free structural parameters becomes operative. Under the normal-gamma prior, each of the k independent Gaussian prior blocks contributes one factor $(\sigma^2)^{-1/2}$, so the free-parameter count k enters the shape; under the g-prior, the single Gaussian prior is placed on the full vector $\boldsymbol{\beta} \in \mathbb{R}^m$, so the ambient dimension m enters; and under Jeffreys' prior no Gaussian block is present, so only n enters. In all three cases the rate b is positive with probability one, so the conditional posterior of σ^2 is proper.*

4.5. Conditional Posteriors of the SMN Latent Weights and Mixing Parameters

Write $e_t(\boldsymbol{\beta}) = y_t - \boldsymbol{x}_t^\top \boldsymbol{\beta}$ for the SARX residual; because $\boldsymbol{x}_t^\top \boldsymbol{\beta} = (\boldsymbol{A}\boldsymbol{\phi})_t + \boldsymbol{w}_t^\top \boldsymbol{\xi}$ by Proposition 1, $e_t(\boldsymbol{\beta})$ depends on $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi})$ only through this scalar fitted value. All full conditionals in this subsection are derived from the augmented joint posterior (19) given the current values of the remaining unobservables. Because the complete-data kernel (13) factors across t given $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\nu})$, the latent quantities u_t (respectively, the indicators I_t of Proposition 4 below) are mutually conditionally independent and may be drawn in parallel. Throughout, $\mathrm{TGamma}(a, b; S)$ denotes the $\mathrm{Gamma}(a, b)$ distribution with shape a and rate b truncated to $S \subseteq (0, \infty)$. $\mathrm{TN}(\mu, \tau^2; S)$ denotes the $N(\mu, \tau^2)$ distribution truncated to S, and $\mathbb{1}_S(\cdot)$ the indicator function of S. The truncated-exponential priors of Section 3.2 are understood through their proper joint kernels $\zeta(\nu, \lambda) \propto \lambda e^{-\lambda\nu}\, \mathbb{1}_{(2,\infty)}(\nu) \mathbb{1}_{(c,d)}(\lambda)$ under SMN-t and $\zeta(\nu, \lambda) \propto \lambda e^{-\lambda\nu}\, \mathbb{1}_{(1,\infty)}(\nu) \mathbb{1}_{(c,d)}(\lambda)$ under SMN-sl; the ν-conditional of each kernel is the truncated exponential with rate λ described in Section 3.2, and both induce the truncated-gamma conditionals for λ derived below.

Proposition 2 (SMN-t case). *For $t = 1, \dots, n$,*

$$u_t \mid \cdot \;\sim\; \text{Gamma}\!\left(\frac{\nu+1}{2},\; \frac{e_t(\boldsymbol{\beta})^2}{2\sigma^2} + \frac{\nu}{2}\right), \tag{44}$$

and the conditional posterior of ν satisfies

$$\zeta(\nu \mid \cdot) \;\propto\; \left(\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\right)^n \exp\left\{-\nu\left(\frac{1}{2}\sum_{t=1}^{n} u_t + \lambda\right)\right\} \prod_{t=1}^{n} u_t^{(\nu/2-1)}, \qquad \nu > 2, \tag{45}$$

which is non-standard and is sampled by the Metropolis–Hastings step of Section 5.2 with a $(2,\infty)$-truncated-normal proposal; the conditional posterior of the prior rate is $\lambda \mid \nu, \cdot \sim \text{TGamma}(2, \nu; (c,d))$.

Proof. Under $u_t \sim \text{Gamma}(\nu/2, \nu/2)$, the complete-data kernel (13) restricted to u_t is $u_t^{1/2} \cdot u_t^{\nu/2-1} e^{-\nu u_t/2} \cdot \exp\{-u_t e_t(\boldsymbol{\beta})^2/(2\sigma^2)\} = u_t^{(\nu+1)/2-1} \exp\{-u_t[e_t(\boldsymbol{\beta})^2/(2\sigma^2) + \nu/2]\}$, the kernel of the stated gamma distribution, giving (44). Collecting all n terms $p(u_t \mid \nu)$ together with the joint prior kernel $\lambda e^{-\lambda\nu} \mathbb{1}_{(2,\infty)}(\nu) \mathbb{1}_{(c,d)}(\lambda)$ yields (45) directly. Since (45) is not a recognisable standard kernel in ν, the Metropolis–Hastings correction of Section 5.2 is required. Viewing the same joint kernel as a function of λ alone gives $\zeta(\lambda \mid \cdot) \propto \lambda\, e^{-\lambda\nu} \mathbb{1}_{(c,d)}(\lambda)$, the kernel of a Gamma$(2, \nu)$ density truncated to (c,d). □

Proposition 3 (SMN-sl case). *For $t = 1, \dots, n$,*

$$u_t \mid \cdot \;\sim\; \text{TGamma}\!\left(\nu + \tfrac{1}{2},\; \frac{e_t(\boldsymbol{\beta})^2}{2\sigma^2};\; (0,1)\right), \tag{46}$$

and

$$\nu \mid \cdot \;\sim\; \text{TGamma}\!\left(n+1,\; \lambda - \sum_{t=1}^{n} \ln u_t;\; (1,\infty)\right), \qquad \lambda \mid \nu, \cdot \sim \text{TGamma}(2, \nu; (c,d)). \tag{47}$$

Proof. Under $u_t \sim \text{Beta}(\nu, 1)$, $p(u_t \mid \nu) = \nu u_t^{\nu-1} \mathbb{1}_{(0,1)}(u_t)$, so the complete-data kernel in u_t is $u_t^{1/2} \cdot u_t^{\nu-1} \cdot \exp\{-u_t e_t(\boldsymbol{\beta})^2/(2\sigma^2)\} \mathbb{1}_{(0,1)}(u_t)$, the kernel of (46). Collecting the n factors $\nu u_t^{\nu-1}$ together with the joint prior kernel gives $\zeta(\nu \mid \cdot) \propto \nu^n \exp\{-\nu(\lambda - \sum_t \ln u_t)\} \mathbb{1}_{(1,\infty)}(\nu)$, the kernel of (47). Note that the rate is strictly positive because $u_t \in (0,1)$ almost surely implies $\sum_{t=1}^{n} \ln u_t < 0$. The conditional posterior of λ follows as in Proposition 2. Unlike the SMN-t case, both conditionals are of standard truncated-gamma form, so no Metropolis–Hastings correction is needed. □

Proposition 4 (SMN-cn case). *Parameterize the contaminated-normal latent structure through the contamination indicators $\boldsymbol{I} = (I_1, \dots, I_n)^\top$, $I_t \in \{0,1\}$, defined by the one-to-one map $u_t = \gamma^{I_t}$, so that $I_t = 1$ if and only if $u_t = \gamma$.* (i) *For $t = 1, \dots, n$,*

$$I_t \mid \cdot \;\sim\; \text{Bernoulli}(\pi_t), \qquad \pi_t = \frac{\nu^*_{1,t}}{\nu^*_{1,t} + \nu^*_{2,t}}, \tag{48}$$

where

$$\nu^*_{1,t} = \nu\gamma^{1/2} \exp\left\{-\frac{\gamma\, e_t(\boldsymbol{\beta})^2}{2\sigma^2}\right\}, \qquad \nu^*_{2,t} = (1-\nu) \exp\left\{-\frac{e_t(\boldsymbol{\beta})^2}{2\sigma^2}\right\}; \tag{49}$$

equivalently, $u_t \mid \cdot$ is discrete on $\{\gamma, 1\}$ with $\Pr(u_t = \gamma \mid \cdot) = \pi_t$. (ii) *With $m_\gamma = \sum_{t=1}^{n} I_t$,*

$$\nu \mid \cdot \;\sim\; \mathrm{Beta}(\nu_0 + m_\gamma,\, \nu_1 + n - m_\gamma). \tag{50}$$

(iii) *The full conditional posterior of γ, given the indicators, is*

$$\zeta(\gamma \mid \boldsymbol{I}, \cdot) \;\propto\; \gamma^{\gamma_0 + m_\gamma/2 - 1}\,(1-\gamma)^{\gamma_1 - 1}\,\exp\!\left\{-\frac{\gamma\, S_\gamma}{2\sigma^2}\right\}, \qquad \gamma \in (0,1), \tag{51}$$

where $S_\gamma = \sum_{t:\, I_t = 1} e_t(\boldsymbol{\beta})^2$ is the sum of squared residuals of the contaminated observations. The conditionals (48) and (50) are exact draws; (51) is not of standard form and is sampled by the Metropolis–Hastings step of Section 5.2.

Proof. Under the indicator parameterisation, the joint complete-data density of (y_t, I_t) given the parameters is

$$p(y_t \mid I_t, \boldsymbol{\beta}, \sigma^2, \gamma)\,\Pr(I_t \mid \nu) = (2\pi\sigma^2)^{-1/2}\,\gamma^{I_t/2}\exp\!\left\{-\frac{\gamma^{I_t} e_t(\boldsymbol{\beta})^2}{2\sigma^2}\right\}\nu^{I_t}(1-\nu)^{1-I_t},$$

which coincides with the per-observation factor of (13) upon substituting $u_t = \gamma^{I_t}$. For part (i), the two possible values of I_t therefore have kernels $\nu^*_{1,t} = \nu\gamma^{1/2}\exp\{-\gamma e_t(\boldsymbol{\beta})^2/(2\sigma^2)\}$ (at $I_t = 1$) and $\nu^*_{2,t} = (1-\nu)\exp\{-e_t(\boldsymbol{\beta})^2/(2\sigma^2)\}$ (at $I_t = 0$), the common factor $(2\pi\sigma^2)^{-1/2}$ cancelling in the Bayes ratio; this gives (48)–(49). Since $u_t = \gamma^{I_t}$ is a bijection between $\{0,1\}$ and $\{\gamma, 1\}$, the induced law of u_t is the stated two-point law—the draw used in the SMN-cn branch of Algorithm 1. For part (ii), collecting the factors $\nu^{I_t}(1-\nu)^{1-I_t}$ over $t = 1, \ldots, n$ gives the binomial likelihood $\nu^{m_\gamma}(1-\nu)^{n-m_\gamma}$, which combined with the $\mathrm{Beta}(\nu_0, \nu_1)$ prior yields (50) by conjugacy. For part (iii), the γ-dependent factors are $\gamma^{1/2}\exp\{-\gamma e_t(\boldsymbol{\beta})^2/(2\sigma^2)\}$ for each contaminated observation ($I_t = 1$) and unity for each clean observation ($I_t = 0$); their product is $\gamma^{m_\gamma/2}\exp\{-\gamma S_\gamma/(2\sigma^2)\}$, and multiplying by the $\mathrm{Beta}(\gamma_0, \gamma_1)$ prior gives (51). This kernel involves $(1-\gamma)$ only through the prior and is not recognisable as any standard distribution on $(0,1)$; it is therefore sampled by the logit-normal Metropolis–Hastings step of Section 5.2. □

Remark 7 (Indicator parameterisation for the contaminated normal)**.** *Three comments on Proposition 4 are in order. First, the indicator and latent-weight parameterisations are statistically equivalent. $u_t = \gamma^{I_t}$ is a one-to-one map between $\{0,1\}$ and the two-point support $\{\gamma, 1\}$, and the induced joint law of $(\boldsymbol{y}, \boldsymbol{I})$ coincides with that of $(\boldsymbol{y}, \boldsymbol{u})$ under (2), so both induce the same marginal posterior of $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2, \nu, \gamma)$. Second, the indicator form is not merely cosmetic but necessary for a valid MCMC scan on γ. The support $\{\gamma, 1\}^n$ of $\boldsymbol{u}$ depends on the unknown γ, so that conditioned on a realised $\boldsymbol{u}$ containing contaminated values, the conditional law of γ degenerates to a point mass at the current value (any contaminated observation forces $\gamma = u_t$), whereas the support $\{0,1\}^n$ of $\boldsymbol{I}$ is γ-free and (51) is a proper density on $(0,1)$. With the indicator parameterisation, every update in the SMN-cn branch of Algorithm 1 is an exact full-conditional draw or a Metropolis–Hastings step targeting the corresponding full conditional, so the resulting sampler is exactly invariant for the augmented posterior (19) in its indicator form. Third, because the clean component has fixed precision 1 and the contaminated component precision $\gamma \in (0,1)$, the two components are distinguished by construction and no label-switching problem arises.*

Finally, in the Gaussian case (SMN-n), $p(u = 1) = 1$, so $\boldsymbol{u} \equiv \mathbf{1}$ is fixed, no latent-weight or shape updates are required, and the sampler reduces to the plain Gibbs scheme built from the exact conditional draws of Sections 4.2–4.4.

5. Hybrid MCMC Algorithm for Bayesian Estimation of SARX Models

Section 4 supplies every full conditional posterior distribution required by a hybrid MCMC scheme. This section assembles them into a systematic-scan Metropolis–Hastings-within-Gibbs sampler targeting the augmented posterior (19).

5.1. Design of the Sampler and Summary of the Hybrid MCMC Algorithm

The sampler operates on the augmented space $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2, \boldsymbol{u}, \boldsymbol{\nu})$ with target (19). Its blockwise architecture is dictated by the SARX mean function. No single conjugate update of all three coefficient blocks exists, and Proposition 1 is what restores exact conditional Gaussianity one block at a time, with the remaining blocks absorbed into the working responses of Table 1. Algorithm 1 presents the complete sampler.

Three design points are worth recording explicitly.

(i) *The chain must move on the space of primitive coefficients.* By Remark 2, the multiplicative structure confines $\boldsymbol{\beta}$ to a k-dimensional manifold of $\mathbb{R}^m$, and an unconstrained m-dimensional Gaussian draw of $\boldsymbol{\beta}$ would violate the constraint almost surely. The exact decompositions (20)–(22) are therefore not merely a computational convenience. They are the mechanism that makes sequential, constraint-respecting, exactly conjugate updates possible.

(ii) *Systematic scan with exact conditional draws.* Each iteration updates $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, $\boldsymbol{\xi}$, and σ^2 in this order, each step conditioning on the most recent values of all other quantities, followed by the SMN-specific updates of the latent weights and shape parameters. Conditional on $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$, the $\boldsymbol{\xi}$-update depends on the autoregressive blocks only through the fitted values $\boldsymbol{A\phi}$, never through the design (Corollary 3). Conversely, the $\boldsymbol{\phi}_1$- and $\boldsymbol{\phi}_2$-updates interact through the adjusted designs $\boldsymbol{Z}_{\phi_1}$ and $\boldsymbol{Z}_{\phi_2}$, each of which depends on the current value of the other seasonal block.

(iii) *Only two scalar Metropolis–Hastings corrections are ever needed.* Steps 1–4 are exact draws under all three prior families and all four SMN cases, and so are all Step-5 draws except ν under SMN-t and γ under SMN-cn. Under SMN-sl and SMN-n the entire sampler is tuning-free; under SMN-t and SMN-cn exactly one proposal scale (τ_ν or τ_γ) must be selected.

Summary of the Algorithm

In words, iteration r of Algorithm 1 consists of: (a) three exact multivariate normal draws updating, in sequence, the nonseasonal autoregressive, seasonal autoregressive, and exogenous coefficient blocks; (b) one exact inverse-gamma draw of the error scale σ^2; (c) exact draws of the latent weights u_t from their SMN-specific conditional distributions; and (d) updates of the mixing shape parameters, exact in all but two cases.

Every individual update is either an exact draw from the corresponding full conditional posterior or a Metropolis–Hastings step targeting that full conditional. Consequently each update—and hence their systematic-scan composition—leaves the augmented posterior (19) invariant, and the marginal chains of the structural parameters target their marginal posteriors [40]. Irreducibility holds provided the conditional precision matrices of Section 4.3 are nonsingular and the two MH proposals place positive density on their supports, $(2, \infty)$ and $(0, 1)$. Since the coefficient-block conditionals have full support on $\mathbb{R}^p$, $\mathbb{R}^P$, and $\mathbb{R}^q$, and that of σ^2 has full support on $(0, \infty)$, the sampler can reach the entire posterior support from any interior starting point such as the OLS initialisation of Section 5.3.

Algorithm 1 Hybrid Gibbs–Metropolis–Hastings sampler for Bayesian estimation of SARX models under SMN errors.

Inputs: data $(\boldsymbol{y}, \{x_{k,t}\}_{t=1}^{n}; k = 1, \ldots, r)$; orders $(p, P; d_1, \ldots, d_r)_s$; SMN case; prior hyperparameters $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$, including the SMN mixing hyperparameters (c, d) or $(\nu_0, \nu_1, \gamma_0, \gamma_1)$.
Specify the run lengths $(N_{\text{sim}}, N_{\text{burn}}, N_{\text{thin}})$ and the MH proposal scales τ_ν, τ_γ (the latter only under SMN-t and SMN-cn).
Construct $\boldsymbol{X}$ and $\boldsymbol{W}$ with rows $\boldsymbol{x}_t^\top, \boldsymbol{w}_t^\top$ as in (7)–(8); compute $\boldsymbol{X}\tilde{\boldsymbol{\beta}}$ once if the g-prior is used.
Initialise (Section 5.3): set $(\boldsymbol{\phi}_1^{(0)}, \boldsymbol{\phi}_2^{(0)}, \boldsymbol{\xi}^{(0)}, (\sigma^2)^{(0)})$ by OLS on (6), projected onto the multiplicative structure; set $u_t^{(0)} = 1$ for all t; set $\nu^{(0)}, \gamma^{(0)}$ at their prior modes and $\lambda^{(0)}$ at the centre of (c, d).
for $r = 1, \ldots, N_{\text{sim}}$ **do**
    [Step 1] Draw $\boldsymbol{\phi}_1^{(r)} \sim N_p(\boldsymbol{\mu}_{\phi_1}^\star, \boldsymbol{V}_{\phi_1}^\star)$ from Corollary 1, using $(\boldsymbol{\phi}_2^{(r-1)}, \boldsymbol{\xi}^{(r-1)}, \boldsymbol{U}^{(r-1)}, (\sigma^2)^{(r-1)})$.
    [Step 2] Draw $\boldsymbol{\phi}_2^{(r)} \sim N_P(\boldsymbol{\mu}_{\phi_2}^\star, \boldsymbol{V}_{\phi_2}^\star)$ from Corollary 2, using $(\boldsymbol{\phi}_1^{(r)}, \boldsymbol{\xi}^{(r-1)}, \boldsymbol{U}^{(r-1)}, (\sigma^2)^{(r-1)})$.
    [Step 3] Draw $\boldsymbol{\xi}^{(r)} \sim N_q(\boldsymbol{\mu}_\xi^\star, \boldsymbol{V}_\xi^\star)$ from Corollary 3, using $(\boldsymbol{\phi}_1^{(r)}, \boldsymbol{\phi}_2^{(r)}, \boldsymbol{U}^{(r-1)}, (\sigma^2)^{(r-1)})$.
    [Step 4] Draw $(\sigma^2)^{(r)} \sim IG(a/2, b/2)$ from Theorem 2, using $(\boldsymbol{\phi}_1^{(r)}, \boldsymbol{\phi}_2^{(r)}, \boldsymbol{\xi}^{(r)}, \boldsymbol{u}^{(r-1)})$.
    [Step 5: SMN-specific latent weights and shape parameters]
    if SMN-t **then**
        Draw $u_t^{(r)} \sim \text{Gamma}\big(\frac{\nu^{(r-1)}+1}{2}, \frac{e_t(\boldsymbol{\beta}^{(r)})^2}{2(\sigma^2)^{(r)}} + \frac{\nu^{(r-1)}}{2}\big)$, $t = 1, \ldots, n$ [Proposition 2].
        Draw $\nu^{(r)}$ by MH with proposal $\nu^\dagger \sim \text{TN}(\nu^{(r-1)}, \tau_\nu^2; (2, \infty))$ and log-acceptance ratio (53), using $\lambda^{(r-1)}$.
        Draw $\lambda^{(r)} \sim \text{TGamma}(2, \nu^{(r)}; (c, d))$.
    else if SMN-sl **then**
        Draw $u_t^{(r)} \sim \text{TGamma}\big(\nu^{(r-1)} + \frac{1}{2}, \frac{e_t(\boldsymbol{\beta}^{(r)})^2}{2(\sigma^2)^{(r)}}; (0, 1)\big)$, $t = 1, \ldots, n$ [Proposition 3].
        Draw $\nu^{(r)} \sim \text{TGamma}\big(n + 1, \lambda^{(r-1)} - \sum_{t=1}^{n} \ln u_t^{(r)}; (1, \infty)\big)$.
        Draw $\lambda^{(r)} \sim \text{TGamma}(2, \nu^{(r)}; (c, d))$.
    else if SMN-cn **then**
        Draw indicators $I_t^{(r)} \in \{0, 1\}$ with $\Pr(I_t^{(r)}{=}1) = \nu_{1,t}^* / (\nu_{1,t}^* + \nu_{2,t}^*)$ from (49), evaluated at $\gamma^{(r-1)}$, and set $m_\gamma^{(r)} = \sum_{t=1}^{n} I_t^{(r)}$ [Proposition 4].
        Draw $\nu^{(r)} \sim \text{Beta}(\nu_0 + m_\gamma^{(r)}, \nu_1 + n - m_\gamma^{(r)})$.
        Draw $\gamma^{(r)}$ by MH with logit-normal proposal targeting (51), using $(\boldsymbol{\beta}^{(r)}, (\sigma^2)^{(r)}, \boldsymbol{I}^{(r)})$ and log-acceptance ratio (54); set $u_t^{(r)} = (\gamma^{(r)})^{I_t^{(r)}}$, $t = 1, \ldots, n$.
    else
        Set $u_t^{(r)} = 1$, $t = 1, \ldots, n$ (Gaussian case; no shape parameters).
    end if
    Record the deviance value $D^{(r)}$ used in the DIC computations of Section 7.
end for
Adaptation (burn-in only): for $r \le N_{\text{burn}}$, adapt τ_ν and/or τ_γ toward an acceptance rate in $[0.2, 0.5]$ every 100 iterations (Section 5.2); freeze both scales for $r > N_{\text{burn}}$.
Output: discard draws $r \le N_{\text{burn}}$; retain every N_{thin}-th draw thereafter ($N_{\text{ret}} = (N_{\text{sim}} - N_{\text{burn}})/N_{\text{thin}}$ draws); report posterior means, standard deviations, and 95% equal-tailed credible intervals for $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2, \nu)$, Monte Carlo standard errors (Section 5.3), the post-burn-in average acceptance rates of the MH steps, and the convergence diagnostics of Section 5.3.

In Algorithm 1, steps 1–3 are executed sequentially rather than jointly because $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ enter $\boldsymbol{X}\boldsymbol{\beta}$ bilinearly. Conditioning on any two of the three blocks restores exact conjugacy in the third, per Theorem 1. Under the g-prior, the centre $\boldsymbol{X}\tilde{\boldsymbol{\beta}}$ is computed once before the

run, while each pseudo-response z_g of Table 1 is formed with the current values of the other blocks at the corresponding step. OLS initialisation places the chain near the posterior mode and does not depend on which SMN case is subsequently fitted, since $u_t^{(0)} = 1$ for all t is common to all cases and recovers the Gaussian starting point. The SMN-specific branch of Step 5 then adjusts the latent weights to the fitted error law within the first few iterations. Finally, the deviance values recorded at every iteration feed the model-comparison analysis of Section 7.

5.2. Metropolis–Hastings Steps: Proposals and Acceptance Ratios

Two scalar updates require a Metropolis–Hastings correction: the shape parameter ν under SMN-t and the contamination parameter γ under SMN-cn. Both are random-walk proposals on a single parameter, so the tuning burden is minimal. For completeness, and to make the algorithm fully reproducible, this section states the proposals and the exact acceptance ratios used in Algorithm 1. All ratios are evaluated on the logarithmic scale.

5.2.1. Update of ν (SMN-t)

The full conditional of ν is given in (45); its logarithmic kernel is

$$\ln \zeta(\nu \mid \cdot) \;=\; n\Big[\tfrac{\nu}{2}\ln\big(\tfrac{\nu}{2}\big) - \ln\Gamma\big(\tfrac{\nu}{2}\big)\Big] - \nu\Big(\tfrac{1}{2}\sum_{t=1}^{n} u_t + \lambda\Big) + \Big(\tfrac{\nu}{2} - 1\Big)\sum_{t=1}^{n}\ln u_t + \text{const},\ \nu > 2. \tag{52}$$

We propose $\nu^\dagger \sim \mathrm{TN}(\nu, \tau_\nu^2; (2,\infty))$. The truncated-normal proposal density, $q(x \mid \nu) = \tau_\nu^{-1}\varphi\big((x-\nu)/\tau_\nu\big)/\big[1 - \Phi\big((2-\nu)/\tau_\nu\big)\big]$ for $x > 2$, has a normalising constant that depends on its conditioning value; the proposal is therefore not exactly symmetric, and the Hastings correction must be retained. The proposal is accepted with probability $\min\{1, A_\nu\}$, where

$$\begin{aligned}\log A_\nu \;=\;& n\Big[\tfrac{\nu^\dagger}{2}\ln\big(\tfrac{\nu^\dagger}{2}\big) - \ln\Gamma\big(\tfrac{\nu^\dagger}{2}\big) - \tfrac{\nu}{2}\ln\big(\tfrac{\nu}{2}\big) + \ln\Gamma\big(\tfrac{\nu}{2}\big)\Big] - (\nu^\dagger - \nu)\Big(\tfrac{1}{2}\sum_{t=1}^{n} u_t + \lambda\Big) \\ &+ \frac{\nu^\dagger - \nu}{2}\sum_{t=1}^{n}\ln u_t \;+\; \ln\frac{1 - \Phi\big((2-\nu)/\tau_\nu\big)}{1 - \Phi\big((2-\nu^\dagger)/\tau_\nu\big)},\end{aligned} \tag{53}$$

φ and Φ denoting the standard normal density and distribution functions. The final ratio is the Hastings correction for the truncation normalisers of the two proposal densities.

5.2.2. Update of γ (SMN-cn)

The CN updates are implemented through contamination indicators $\boldsymbol{I} = (I_1, \ldots, I_n)^\top$, $I_t \in \{0, 1\}$, with $u_t = \gamma^{I_t}$ (Proposition 4). Drawing u_t from the two-point law (49) is equivalent to drawing I_t with $\Pr(I_t = 1) = \nu_{1,t}^*/(\nu_{1,t}^* + \nu_{2,t}^*)$ and setting $u_t = \gamma^{I_t}$. The indicator parameterisation is used because the location of the contaminated atom, γ, is itself unknown. Conditioning on the indicators $\boldsymbol{I}$ rather than on the realised weights $\boldsymbol{u}$ yields the proper full conditional (51) of Proposition 4, with $m_\gamma = \sum_{t=1}^{n} I_t$ and $S_\gamma = \sum_{t:\, I_t = 1} e_t(\boldsymbol{\beta})^2$ the sum of squared residuals of the contaminated observations. Because the two atoms are ordered, $\gamma < 1$, no label-switching problem arises. We propose $\gamma^\dagger$ on the logit scale, $\mathrm{logit}(\gamma^\dagger) = \mathrm{logit}(\gamma) + \epsilon$, $\epsilon \sim N(0, \tau_\gamma^2)$, which automatically respects the $(0, 1)$ support. The logit-normal proposal is asymmetric, and its Jacobian contributing the multiplicative correction $\gamma^\dagger(1-\gamma^\dagger)/[\gamma(1-\gamma)]$. The proposal is accepted with probability $\min\{1, A_\gamma\}$, where

$$\begin{aligned}\log A_\gamma \;=\;& \Big(\gamma_0 + \tfrac{m_\gamma}{2} - 1\Big)\big(\ln\gamma^\dagger - \ln\gamma\big) + (\gamma_1 - 1)\big(\ln(1-\gamma^\dagger) - \ln(1-\gamma)\big) \\ &- \frac{S_\gamma}{2\sigma^2}\big(\gamma^\dagger - \gamma\big) \;+\; \ln\frac{\gamma^\dagger(1-\gamma^\dagger)}{\gamma(1-\gamma)}.\end{aligned} \tag{54}$$

5.2.3. Proposal Tuning

Only the two scales τ_ν and τ_γ require tuning. During burn-in they are adapted by a simple interval rule. Every 100 iterations, the running acceptance rate is compared with the band $[0.2, 0.5]$, whose upper end is close to the 0.44 optimum of one-dimensional random-walk Metropolis moves [41]. The scale is multiplied by 1.1 if the rate exceeds 0.5, by 0.9 if it falls below 0.2, and left unchanged otherwise. Both scales are frozen after burn-in so that the monitored chain is generated by a fixed transition kernel. Acceptance rates persistently above this band preserve correctness but indicate overly conservative moves that inflate autocorrelation. They are improved by enlarging the scale and repeating the burn-in, whereas rates persistently below it call for a smaller scale.

5.3. Initialisation and Convergence Diagnostics

5.3.1. Initialisation

OLS applied to the expanded regression (10) yields $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$, whose cross-lag entries need not equal the products of its main-lag entries. Starting values are therefore obtained by projecting $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ onto the multiplicative structure. $\boldsymbol{\phi}_1^{(0)}$ and $\boldsymbol{\phi}_2^{(0)}$ are taken from the main-lag entries, $\boldsymbol{\xi}^{(0)}$ is refitted by OLS of $\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{\phi}_1^{(0)}, \boldsymbol{\phi}_2^{(0)})$ on $\boldsymbol{W}$, and $(\sigma^2)^{(0)}$ is set to the residual variance or to the prior mode. The latent weights start at the Gaussian configuration $u_t^{(0)} = 1$ for all t. The shape parameters start at their prior modes. $\nu^{(0)} = 2$ under SMN-t, $\nu^{(0)} = 1$ under SMN-sl, and the interior Beta modes for $\nu^{(0)}$ and $\gamma^{(0)}$ under SMN-cn. Under SMN-t and SMN-sl, $\lambda^{(0)} = (c + d)/2$ is the centre of (c, d).

5.3.2. Convergence Diagnostics

Convergence is assessed on the post-burn-in thinned chains using the following complementary diagnostics.

- *Geweke z-scores* [42]: The difference between the means of the first 10% and the last 50% of each chain, standardised by spectral-density estimates. Convergence is declared when all z-scores lie in $(-1.96, 1.96)$.
- *Raftery–Lewis diagnostic* [43]: The total run length required for the target precision of estimated quantiles, the minimum number of essentially independent draws, and the dependence factor, which flags insufficient thinning when large.
- *Autocorrelation functions and effective sample size*: Sample ACFs at lags 1, 5, 10, and 50 are inspected for every parameter, and the effective sample size $\mathrm{ESS} = N_{\mathrm{ret}} / \left(1 + 2 \sum_{l \geq 1} \rho_l\right)$ is reported. Values of $|\rho_1|$ above roughly 0.1, or ESS below a few hundred, indicate that N_{thin} or N_{sim} should be increased.
- *Monte Carlo standard errors*: MCSEs of the reported posterior means, estimated by batch means, are required to be at most 5% of the corresponding posterior standard deviations.

Because every coefficient-block and scale update is an exact draw from its full conditional, the only conceivable sources of slow mixing are the bilinear coupling between $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, handled by the sequential scan, and a single scalar MH step in two of the four SMN cases. Thinning serves to reduce autocorrelation and storage and to supply the approximately independent draws used in the DIC computations of Section 7. Ergodic averages remain valid without thinning [40,44]. The diagnostics reported in Section 6 follow these guidelines, and further implementation guidance is given in LeSage [45].

6. Simulation Study

We evaluate the proposed hybrid MCMC algorithm on three SARX configurations, each paired with a different non-Gaussian SMN error case. Collectively, these designs

span low-order and higher-order seasonal structure, and moderate to severe tail heaviness. Table 2 summarises the design.

Table 2. Simulation study design: SARX model configurations and true parameter values.

Model	$(p, P; d_1, \ldots)_s$	$\boldsymbol{\phi_1}$	$\boldsymbol{\phi_2}$	$\boldsymbol{\xi}$	n	σ^2	Error Distribution (ν^{true})
I	$(1,1;1,0)_{12}$	0.6	0.3	$\xi_0 = 1.5, \xi_{10} = 0.6, \xi_{11} = 0.3, \xi_{20} = 0.5$	300	1.0	SMN-cn ($\nu = 0.2, \gamma = 0.1$)
II	$(2,1;0,0)_{12}$	(0.5, 0.3)	0.4	$\xi_0 = 1.8, \xi_{10} = 0.5, \xi_{20} = 0.4$	300	1.0	SMN-sl ($\nu = 2.0$)
III	$(1,1;1,0)_4$	0.4	0.6	$\xi_0 = 2.0, \xi_{10}$=0.5, $\xi_{11} = 0.4, \xi_{20} = 0.6$	300	1.0	SMN-t ($\nu = 4.0$)

Model I is a SARX$(1,1;1,0)_{12}$ specification suited to monthly data, carrying two exogenous regressors—the first entering with a one-period lag (ξ_{10}, ξ_{11}) and the second contemporaneously (ξ_{20})—together with contaminated normal errors ($\nu = 0.2$, $\gamma = 0.1$). Model II is a SARX$(2,1;0,0)_{12}$ specification with two nonseasonal and one seasonal autoregressive lag and two purely contemporaneous exogenous regressors, paired with slash-distributed innovations ($\nu = 2.0$). Among the three designs, this is the one with the heaviest polynomial tail decay, and hence the most demanding test of the sampler's robustness. Model III is a SARX$(1,1;1,0)_4$ specification—appropriate to quarterly seasonality—again with one lagged and one contemporaneous exogenous regressor, but with moderately heavy-tailed Student's t innovations ($\nu = 4.0$). For every model the exogenous regressors are generated as independent stationary AR(1) processes with autoregressive coefficient 0.7 and standard Gaussian innovations, uncorrelated with the SARX errors.

For each model, we generate 500 independent replicate datasets, each using its own reproducible random stream. We then apply Algorithm 1 with Jeffreys' prior (18) for $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2)$, which involves no hyperparameters (the default prior), and the conjugate SMN priors of Section 3.2 with the following hyperparameter values. Under SMN-cn (Model I), $\nu \sim \text{Beta}(\nu_0, \nu_1)$ and $\gamma \sim \text{Beta}(\gamma_0, \gamma_1)$ with $\nu_0 = \nu_1 = \gamma_0 = \gamma_1 = 2$, i.e., prior means of $1/2$ for both contamination parameters. Under SMN-t (Model III) and SMN-sl (Model II), (ν, λ) receives the joint truncated-exponential prior of Section 3.2, with $a = 2$ and $a = 1$, respectively, and $(c, d) = (0.02, 0.5)$ and $(c, d) = (0.5, 1.0)$, respectively. The implied conditional prior means $\text{E}[\nu \mid \lambda] = a + 1/\lambda$ range over $(4, 52)$ and $(2, 3)$. The Metropolis–Hastings proposal scales are initialised at $\tau_\nu = 1.0$ (SMN-t) and $\tau_\gamma = 0.3$ (SMN-cn) and adapted during burn-in toward the 20–50% acceptance band of Section 5.2, then frozen. MCMC settings are $N_{\text{sim}} = 25{,}000$, $N_{\text{burn}} = 5{,}000$, $N_{\text{thin}} = 20$, giving 1,000 retained draws per replicate. Performance is summarised, across replicates, by the average posterior mean (μ), posterior standard deviation (σ), 95% equal-tailed credible interval (L, U), root mean square error (RMSE), and coverage probability (CP), the proportion of replicates whose credible interval contains the true value. For the single-replicate illustration below, the full convergence battery of Section 5.3 is additionally reported.

6.1. Illustrative Example

Before turning to the aggregated results, we illustrate the sampler's behaviour on a single representative dataset generated from Model I. Table 3 reports the resulting Bayesian estimates, while Table 4 reports the convergence diagnostics and Figure 1 displays the trace plots and marginal posterior distributions.

From Table 3, the posterior means of the autoregressive, exogenous-regression, and scale parameters lie close to the generating values, and all nine 95% credible intervals cover the corresponding true values, including those of the contamination parameters ν and γ. The largest point deviations occur for the intercept ($\hat{\xi}_0 = 1.826$ against $\xi_0 = 1.5$) and the contamination probability ($\hat{\nu} = 0.149$ against $\nu = 0.2$). The deviations remain well inside the posterior uncertainty rather than indicating bias. The Metropolis–Hastings step embed-

ded in the sampler for this SMN-cn replicate—used to update the contamination-variance parameter γ, whose conditional posterior (51) is non-standard—registers an acceptance rate of 35.6%, comfortably within the 20–50% target band of Section 5.2.

Table 3. Bayesian estimates for one replicate of Model I (SMN-cn). μ: posterior mean; σ: posterior SD; $[L, U]$: 95% credible interval.

Parameter	True	μ	σ	L	U
ϕ_{11}	0.6	0.560	0.031	0.500	0.624
ϕ_{21}	0.3	0.243	0.035	0.172	0.314
ξ_0	1.5	1.826	0.176	1.488	2.179
ξ_{10}	0.6	0.549	0.075	0.406	0.695
ξ_{11}	0.3	0.405	0.082	0.245	0.569
ξ_{20}	0.5	0.599	0.055	0.494	0.704
σ^2	1.0	1.093	0.142	0.843	1.394
ν	0.2	0.149	0.053	0.071	0.269
γ	0.1	0.126	0.038	0.065	0.214

Table 4. Convergence diagnostics of the MCMC chain for one replicate of Model I (SMN-cn).

	Raftery–Lewis			Geweke		Autocorrelations				Efficiency		
Param	**Ntotal**	**Nmin**	**Factor**	**Z-Score**	***p*-Value**	**Lag 1**	**Lag 5**	**Lag 10**	**Lag 50**	**ESS**	**MCSE**	**MCSE/SD (%)**
ϕ_{11}	892	936	0.953	−0.192	0.848	0.001	−0.044	0.055	0.001	907	0.00101	3.32
ϕ_{21}	892	936	0.953	0.308	0.758	−0.036	−0.062	0.020	0.026	1077	0.00108	3.05
ξ_0	968	936	1.033	0.325	0.746	0.022	−0.026	0.066	0.015	929	0.00579	3.28
ξ_{10}	968	936	1.033	1.560	0.119	−0.034	−0.036	0.019	0.024	1047	0.00233	3.09
ξ_{11}	892	936	0.953	0.181	0.856	−0.044	0.011	0.027	0.015	1076	0.00249	3.05
ξ_{20}	968	936	1.033	−0.376	0.707	0.008	−0.062	0.014	−0.021	893	0.00183	3.35
σ^2	971	936	1.037	−0.174	0.862	−0.067	0.034	−0.071	0.032	1154	0.00419	2.94
ν	892	936	0.953	0.768	0.443	0.084	0.033	0.001	0.034	733	0.00196	3.69
γ	892	936	0.953	0.596	0.551	0.083	−0.043	−0.039	0.015	819	0.00134	3.49

Note: Raftery–Lewis: $q = 0.025$, $r = 0.01$, $s = 0.95$. Geweke: first 10% vs last 50%. ESS and MCSE use the Geyer initial-monotone-sequence estimator.

Convergence of this representative chain is supported by the diagnostics reported in Table 4. Every Geweke z-score lies well inside the $(-1.96, 1.96)$ acceptance band, giving no indication that the means of the early and late portions of the post-burn-in chain differ. The Raftery–Lewis diagnostic recommends, after accounting for autocorrelation, a total chain length of between 892 and 971 iterations to estimate the 2.5th posterior quantile to the target accuracy, against a theoretical independence-chain minimum of 936. The dependence factors range from 0.953 to 1.037, so the retained 1000-draw chain meets the requirement for all nine parameters. The factors below unity reflecting the mildly antithetic dependence that thinning can induce. This near-independence is corroborated by the reported autocorrelations themselves, the largest of which is 0.084 in absolute value (for ν at lag 1), confirming that the chosen thinning interval $N_{\text{thin}} = 20$ removes essentially all residual serial dependence. The efficiency columns of Table 4 translate this into sampling accuracy. Effective sample sizes range from 733 (ν) to 1154 (σ^2) out of the 1000 retained draws. The Monte Carlo standard errors amount to between 2.94% and 3.69% of the posterior standard deviations, comfortably below the 5% gate adopted in Section 5.3. These numerical diagnostics are echoed visually by the trace plots and marginal posterior densities of Figure 1. The trace plots mix rapidly with no discernible trend, drift, or sustained excursion, consistent with the near-zero autocorrelations of Table 4. The marginal densities are unimodal and approximately symmetric for the autoregressive and exogenous-regression coefficients and moderately right-skewed for the strictly positive scale-type parameters σ^2, ν, and γ, exactly as theory predicts. No marginal density shows

any sign of multimodality or of switching between competing regions of the parameter space, indicating that the chain concentrates reliably on the single mode corresponding to the true generating model. Table 4 and Figure 1 thus provide concrete, single-dataset evidence for the practical reliability of Algorithm 1 ahead of the aggregated evaluation that follows.

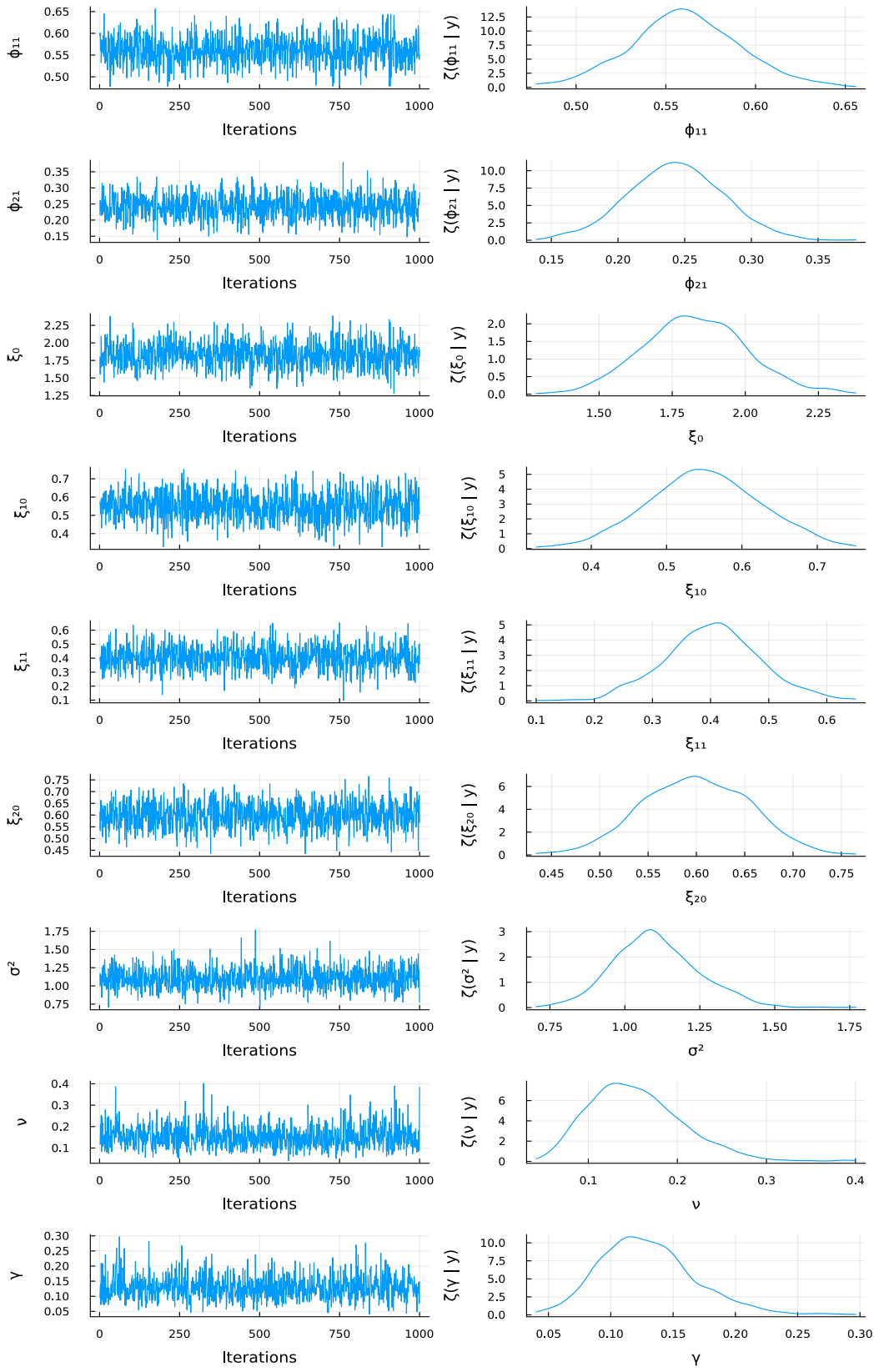


Figure 1. Trace plots and marginal posterior distributions for one replicate of Model I (SMN-cn).

6.2. Aggregated Results

Tables 5–7 report averaged posterior summaries over the 500 replicates for each model under its SMN error case.

Table 5. Aggregated simulation results for Model I (SARX$(1,1;1,0)_{12}$, SMN-cn).

Parameter	True	μ	σ	L	U	CP (%)	RMSE
ϕ_{11}	0.6	0.595	0.028	0.541	0.650	94.6	0.028
ϕ_{21}	0.3	0.296	0.034	0.230	0.362	95.0	0.033
ξ_0	1.5	1.513	0.148	1.227	1.815	94.8	0.152
ξ_{10}	0.6	0.597	0.076	0.449	0.746	95.4	0.073
ξ_{11}	0.3	0.308	0.085	0.144	0.470	96.0	0.083
ξ_{20}	0.5	0.503	0.057	0.391	0.613	96.6	0.054
σ^2	1.0	0.971	0.162	0.680	1.322	97.6	0.153
ν	0.2	0.234	0.064	0.126	0.380	94.2	0.072
γ	0.1	0.111	0.028	0.067	0.177	93.0	0.037

Table 6. Aggregated simulation results for Model II (SARX$(2,1;0,0)_{12}$, SMN-sl).

Parameter	True	μ	σ	L	U	CP (%)	RMSE
ϕ_{11}	0.5	0.496	0.050	0.399	0.592	94.4	0.051
ϕ_{12}	0.3	0.296	0.048	0.203	0.389	94.0	0.047
ϕ_{21}	0.4	0.393	0.045	0.308	0.481	94.2	0.046
ξ_0	1.8	1.871	0.281	1.331	2.446	92.6	0.303
ξ_{10}	0.5	0.504	0.060	0.387	0.621	95.8	0.061
ξ_{20}	0.4	0.404	0.059	0.286	0.519	93.8	0.063
σ^2	1.0	1.019	0.179	0.719	1.428	95.4	0.174
ν	2.0	2.064	0.658	1.327	3.753	96.4	0.597

Table 7. Aggregated simulation results for Model III (SARX$(1,1;1,0)_4$, SMN-t).

Parameter	True	μ	σ	L	U	CP (%)	RMSE
ϕ_{11}	0.4	0.396	0.037	0.326	0.467	93.2	0.039
ϕ_{21}	0.6	0.598	0.027	0.544	0.651	94.8	0.028
ξ_0	2.0	2.029	0.177	1.690	2.373	94.6	0.182
ξ_{10}	0.5	0.501	0.070	0.365	0.637	94.8	0.069
ξ_{11}	0.4	0.409	0.080	0.252	0.563	94.8	0.082
ξ_{20}	0.6	0.607	0.055	0.499	0.716	93.6	0.057
σ^2	1.0	1.013	0.154	0.752	1.357	96.0	0.148
ν	4.0	4.190	1.228	2.641	7.452	94.6	1.568

Across all three models, average posterior means track the true generating values closely and RMSEs are of the same magnitude as—and only rarely exceed—the corresponding average posterior standard deviations. This confirms that the sampler locates the posterior centre reliably despite the bilinear coupling between $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ induced by the multiplicative seasonal structure. The largest relative biases occur for the contamination parameters of Model I ($\hat{\nu} = 0.234$ against 0.2 and $\hat{\gamma} = 0.111$ against 0.1) and remain small relative to the associated posterior standard deviations. Coverage probabilities are close to the nominal 95% level throughout the 25 parameter–model combinations. They range from 93.0% to 97.6% for Model I, from 92.6% to 96.4% for Model II, and from 93.2% to 96.0% for Model III. The widest interval in the study relative to its signal is that for the intercept ξ_0 of Model II. Its coverage of 92.6% remains within two-and-a-half percentage points of nominal, and its posterior mean deviates from the truth by only about a quarter of a posterior standard deviation ($\hat{\xi}_0 = 1.871$ against $\xi_0 = 1.8$), indicating width-driven

calibration rather than bias. This width plausibly reflects the weaker identification of the level when both exogenous regressors enter only contemporaneously, together with the occasional very small latent weights u_t permitted by the slash mixing law.

Across all three SMN cases, the exogenous-regression coefficients $\boldsymbol{\xi}$ are recovered with a precision generally comparable to that of the autoregressive coefficients $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2$. Relative posterior standard deviations lie between 4.5% and 16.2% for every autoregressive coefficient, and below 20% for every exogenous coefficient. A single exception is ξ_{11} in Model I (27.6%), the weakest-signal coefficient in the design, whose posterior variance is inflated by the collinearity (coefficient 0.7) between the contemporaneous and lagged values of the first regressor. This corroborates, at the level of a full simulation study, the exact block-conjugacy result of Corollary 3. The seasonal autoregressive structure affects the exogenous coefficients only through the working response of (37)–(39), never through their design. So, it does not degrade the quality of exogenous-effect estimation. As in the single-replicate illustration, the SMN shape and contamination parameters are estimated with substantially larger relative posterior standard deviations than the regression coefficients. Specifically, the relative posterior standard deviations are 27.4% for the contamination probability ν and 25.2% for γ in Model I, 31.9% for the slash parameter ν in Model II, and 29.3% for the degrees-of-freedom parameter ν in Model III. This behaviour is expected and reflects the comparatively weak identifiability of tail-shape parameters in samples of moderate size [8]. This added uncertainty is, however, correctly reflected in, rather than concealed from, the reported credible intervals. Coverage probabilities for $\boldsymbol{\nu}$ and γ remain close to the nominal level in every model (93.0–96.4%).

Across all simulated replicates, MCMC convergence was confirmed using the full battery of Section 5.3 illustrated for the single replicate above. For the two SMN cases whose shape or contamination parameters require a Metropolis–Hastings correction (Propositions 2 and 4; Section 5.2), the sampler maintained efficient acceptance rates throughout. The average Metropolis–Hastings acceptance rate across replicates was 39.0% for the SMN-cn contamination parameter γ (Model I) and 34.0% for the SMN-t degrees-of-freedom parameter ν (Model III), both comfortably within the 20–50% target band of Section 5.2. The simulation evidence establishes the operating characteristics of Algorithm 1 under known data-generating processes.

6.3. Sensitivity Analyses

This subsection examines how the results depend on the prior family, on the Beta hyperparameters of the contamination parameters, and on the frequency and severity of the contamination. Every run uses the settings of Section 6 (500 replicates, $N_{\text{sim}} = 25{,}000$, $N_{\text{burn}} = 5000$, $N_{\text{thin}} = 20$) and, within a comparison, the same 500 datasets.

6.3.1. Prior Families

We repeat Model III under the normal-gamma and g priors; the results under Jeffreys' prior are in Table 7. The normal-gamma prior is weakly informative on the scale of σ: $\boldsymbol{\mu}_{\phi_1} = \boldsymbol{\mu}_{\phi_2} = \boldsymbol{\mu}_{\xi} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\phi_1} = \boldsymbol{\Sigma}_{\phi_2} = \boldsymbol{I}$, $\boldsymbol{\Sigma}_{\xi} = 100\,\boldsymbol{I}$, and $a_0 = b_0 = 1$. The g-prior uses $g = 1/n$ and takes $\bar{\boldsymbol{\beta}}$ to be the OLS estimate of the expanded regression, i.e., an empirical-Bayes centre. Table 8 reports the results. Across the parameter–model combinations, the average posterior means of the coefficients differ from those under Jeffreys' prior by at most 0.004, the coverage ranges from 92.6% to 96.0% (normal-gamma) and from 91.8% to 95.6% (g), and the RMSE changes by at most 0.015 among the regression coefficients (0.217 overall, driven by ν under the g-prior). The average posterior mean of σ^2 is 95.0% (normal-gamma) and 93.3% (g) of its value under Jeffreys' prior, in line with the shape parameters $a = a_0 + n + k$ and $a = n + m$ of Theorem 2.

Table 8. Aggregated simulation results for Model III (SARX$(1,1;1,0)_4$, SMN-t) under the normal-gamma and g priors.

		Normal-Gamma				g-prior			
Parameter	**True**	μ	σ	**CP (%)**	**RMSE**	μ	σ	**CP (%)**	**RMSE**
ϕ_{11}	0.4	0.393	0.035	94.6	0.036	0.394	0.035	93.6	0.036
ϕ_{21}	0.6	0.598	0.027	94.8	0.028	0.598	0.026	94.6	0.028
ξ_0	2.0	2.028	0.173	95.4	0.167	2.026	0.172	94.8	0.167
ξ_{10}	0.5	0.500	0.069	95.8	0.068	0.500	0.068	95.6	0.068
ξ_{11}	0.4	0.409	0.078	95.4	0.080	0.408	0.078	94.8	0.080
ξ_{20}	0.6	0.603	0.054	96.0	0.053	0.603	0.053	95.6	0.052
σ^2	1.0	0.962	0.144	92.6	0.146	0.945	0.143	91.8	0.152
ν	4.0	4.114	1.493	93.8	1.750	4.030	1.496	93.4	1.785

6.3.2. Beta Hyperparameters of the Contamination Parameters

Given the allocation count m_γ, the conditional posterior mean of ν is $(\nu_0 + m_\gamma)/(\nu_0 + \nu_1 + n)$ by (50) so that Beta$(2,2)$ and Beta$(1,1)$ act as four and two pseudo-observations against $n = 300$. For $m_\gamma \approx 60$, the two conditional means differ by $(300 - 2m_\gamma)/(304 \times 302) \approx 0.002$, an order of magnitude below the deviation $\hat{\nu} - \nu = 0.034$ in Table 5. For γ, the prior changes by one unit the exponent of the conditional kernel (51), which is about $m_\gamma/2 \approx 30$. The Beta$(2,2)$ prior is therefore unlikely to be the main source of the upward deviations in Table 5. Table 9 confirms this: replacing Beta$(2,2)$ by Beta$(1,1)$ changes the average posterior means of ν and γ in Model I by 0.010 and 0.001, respectively; $0.234 \to 0.224$ and $0.111 \to 0.110$ at reported precision.

Table 9. Aggregated simulation results for Model I (SARX$(1,1;1,0)_{12}$, SMN-cn) under Beta$(1,1)$ prior of hyperparameters.

Parameter	**True**	μ	σ	L	U	**CP (%)**	**RMSE**
ϕ_{11}	0.6	0.598	0.027	0.545	0.651	95.8	0.027
ϕ_{21}	0.3	0.296	0.033	0.231	0.360	93.8	0.034
ξ_0	1.5	1.513	0.145	1.236	1.801	95.4	0.143
ξ_{10}	0.6	0.600	0.074	0.457	0.743	93.8	0.075
ξ_{11}	0.3	0.306	0.083	0.144	0.467	94.2	0.083
ξ_{20}	0.5	0.499	0.056	0.390	0.609	94.0	0.058
σ^2	1.0	0.979	0.165	0.687	1.333	95.0	0.154
ν	0.2	0.224	0.066	0.120	0.374	95.4	0.067
γ	0.1	0.110	0.029	0.065	0.177	93.6	0.035

6.3.3. Rarer and More Severe Contamination

The empirical application gives $(\hat{\nu}, \hat{\gamma}) \approx (0.084, 0.010)$, i.e., contamination that is rarer and far more severe than in Model I, $(0.2, 0.1)$. We therefore add Model IV, which is Model I with $(\nu, \gamma) = (0.08, 0.01)$ at $n = 300$: on average $n\gamma = 3$ contaminated observations per sample, rather than about 30 in Model I, with a much larger standard deviation than the clean component. Table 10 reports the results under Beta$(1,1)$ priors, and Table 9 compares them with Model I under the same prior. The posterior means of the autoregressive and exogenous coefficients are within 0.003 of the true values, with coverage between 93.8% and 95.2%, whereas ν and γ have relative RMSEs of 25.0% and 50.0% (absolute RMSEs 0.020 and 0.005) and coverages of 95.2% and 92.0%. The average Metropolis–Hastings acceptance rate for γ is 39.34% in Model IV, compared with 38.78% in Model I.

Table 10. Aggregated simulation results for Model IV (SARX$(1,1;1,0)_{12}$, SMN-cn with $\nu = 0.08$, $\gamma = 0.01$) under Beta$(1,1)$ prior of hyperparameters.

Parameter	True	μ	σ	L	U	CP (%)	RMSE
ϕ_{11}	0.6	0.600	0.016	0.569	0.631	93.8	0.016
ϕ_{21}	0.3	0.298	0.019	0.261	0.335	95.2	0.019
ξ_0	1.5	1.501	0.097	1.312	1.692	94.4	0.096
ξ_{10}	0.6	0.600	0.063	0.476	0.723	94.0	0.067
ξ_{11}	0.3	0.303	0.067	0.172	0.433	95.0	0.069
ξ_{20}	0.5	0.499	0.047	0.407	0.591	95.0	0.049
σ^2	1.0	0.996	0.099	0.824	1.211	95.2	0.099
ν	0.08	0.084	0.020	0.051	0.128	95.2	0.020
γ	0.01	0.012	0.004	0.006	0.021	92.0	0.005

7. Empirical Application: Hourly Solar Irradiance in Najran, Saudi Arabia

The simulation study established that Algorithm 1 recovers SARX-SMN parameters accurately across configurations, orders, and tail regimes, with near-nominal coverage and confirmed convergence. This section asks the complementary question: does the heavy-tailed robustness matter in practice? We answer this question using a series for which all three defining features of the SARX framework are externally documented rather than imposed. Specifically, the series exhibits a deterministic daily cycle governed by solar geometry, a physically motivated exogenous ceiling given by clear-sky irradiance, and a contamination mechanism induced by cloud and dust episodes at a hyper-arid site.

7.1. Data Description and Preprocessing

The dataset consists of two hourly irradiance time series recorded over a 90-day period at the Najran monitoring station in Southwestern Saudi Arabia (approximately 17.5° N). Observations are restricted to a fixed 12 h daylight interval (06:00–17:00 local time), which is identical across all days because Saudi Arabia does not implement daylight saving time. This sampling scheme produces $n = 1080$ observations for each series, arranged as $n_d = 90$ daily blocks of length $s = 12$. Under this structure, the seasonal lag associates each observation with the corresponding hour of the preceding day. Restricting attention to daylight hours is standard in solar-energy studies since irradiance is identically zero at night and daytime conditions are the primary focus for photovoltaic forecasting and operation. The two variables analyzed are defined as follows:

- **Global Horizontal Irradiance (GHI)** $[y_t]$: The total incoming solar radiation, including both direct and diffuse components, received on a horizontal surface and measured in $\mathrm{Wh/m^2}$. The series is obtained from ground-based pyranometer measurements at the Najran station. Because GHI strongly influences photovoltaic output, it is the most widely used response variable in solar applications [46,47].
- **Clear-Sky GHI** $[x_{1,t}]$: The irradiance level predicted under ideal cloud-free atmospheric conditions by a radiative-transfer model. Its computation is based on solar-geometric variables, including latitude, hour angle, and Earth–Sun distance, together with standard atmospheric characteristics such as aerosol optical depth, precipitable water, and ozone concentration [47,48]. This series represents the slowly varying upper bound on available solar energy. Actual GHI fluctuates around this benchmark as cloud cover and aerosol conditions change, making the two series closely related but conceptually distinct.

The region ranks among the sunniest in the Arabian Peninsula, and its solar resource is a central asset of the Kingdom's utility-scale renewable programme [49]. Table 11 lists

summary statistics and Figure 2 displays the two series with their autocorrelation functions. In levels the two series are nearly indistinguishable in their broad pattern (contemporaneous correlation 0.9965), as physics requires, since measured GHI is the product of the clear-sky envelope and a stochastic atmospheric transmittance whose mean over the window is $\text{mean}(\text{GHI})/\text{mean}(\text{CS}) = 0.961$. The informative coupling emerges on the transformed scales used below, where the correlation falls to 0.864 (static-profile anomalies) and 0.557 (differenced anomalies).

Table 11. Summary statistics of the hourly Najran GHI and clear-sky GHI series (levels and differenced anomalies).

Series	Mean	Median	Std	Min	Max
GHI (Wh/m^2), levels	518.03	564.50	320.15	6.00	1022.00
Clear-sky GHI (Wh/m^2), levels	538.93	588.00	319.55	9.00	1043.00
GHI (Wh/m^2), differenced anomalies	−0.00	1.52	28.64	−211.58	240.42
Clear-sky (Wh/m^2), differenced anomalies	−0.00	1.20	16.87	−59.43	82.57

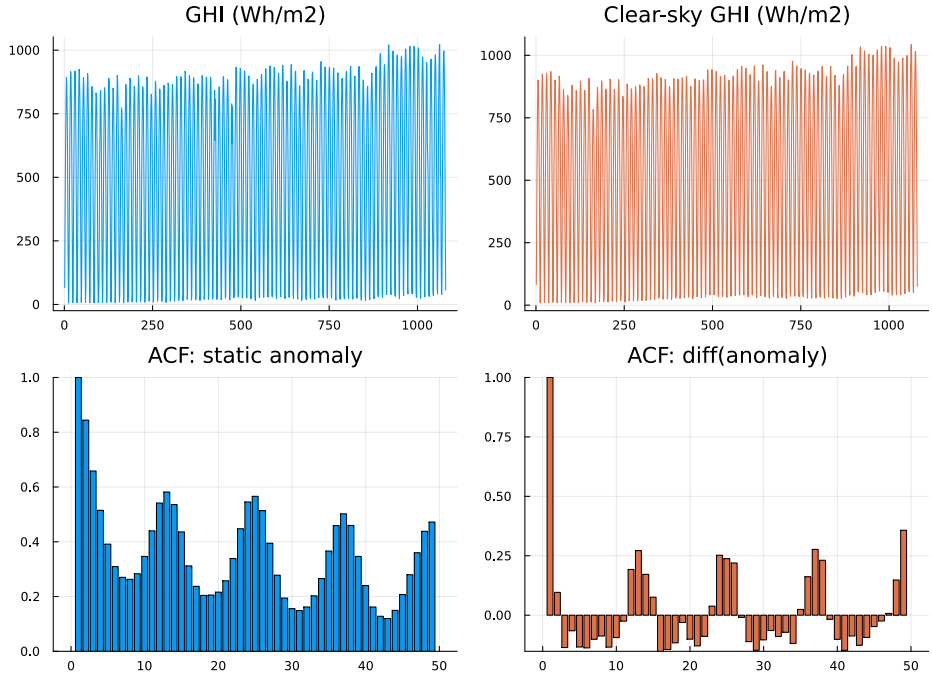


Figure 2. Hourly GHI and clear-sky GHI at Najran: raw series (**top**) and sample autocorrelation functions of the static-profile anomalies and their first differences to lag 48 (**bottom**).

The dominant feature of both series is a deterministic daily cycle imposed by solar geometry, which accounts for roughly 97% of the variance of GHI. Removing the static hourly profile alone reduces the GHI standard deviation from 320.1 to 51.3 Wh/m^2. The analysis is conducted on the differenced anomaly scale, obtained by subtracting the static per-hour profile from each series and then taking first differences, for three reasons. First, the resulting response—the hour-to-hour change of the irradiance anomaly—is the operationally relevant object in solar applications, where ramp events rather than levels drive grid management. Second, on levels, the exogenous channel and the seasonal autoregressive block are nearly collinear. The lag-s observation is approximately the clear-sky level scaled by a slowly varying clearness factor so that the levels correlation of 0.9965 leaves the two mechanisms identified only through day-to-day clearness variation. Differencing reduces the contemporaneous correlation between response and regressor to 0.557 and yields a well-conditioned design, under which the exact block updates of Section 4 mix efficiently. Third, stationarity of the transformed series is supported by the Augmented Dickey–Fuller

(ADF) test at the 5% significance level, providing a statistical justification for modelling the resulting process within a stationary autoregressive framework. The marginal kurtosis of the differenced anomalies is 20.48, while the conditional residuals of the fitted specification below display kurtosis 37.5. The heavy tails are thus an innovation-level phenomenon, invisible to marginal diagnostics and progressively exposed as deterministic and exogenous structure is stripped away—exactly the setting the SMN hierarchy of Section 2.1 addresses.

7.2. SARX Order Selection, Bayesian Estimation, and Model Comparison

The SARX order is selected by minimising the Bayesian information criterion (BIC) over the grid $p, P \in \{1,2,3\}$ and $d_1 \in \{0,1,2\}$ at seasonal period $s = 12$. It is evaluated on a common estimation sample across configurations and penalised by the free-parameter count $k = p + P + q$ of Remark 2. The selected model is a SARX$(2,1;1)_{12}$ specification with BIC 6533.1. The top four configurations lie within one BIC point and the parsimony penalty selects the smallest. This specification gives $q = 3$ exogenous parameters (intercept, contemporaneous clear-sky effect, and one lag), $k = 6$ free structural parameters against a design dimension $m = 8$, and an aligned sample of $n = 1065$. Prior to fitting the SMN models, the OLS residuals of the selected specification display pronounced non-Gaussianity. Sample kurtosis 37.5 against the Gaussian value of 3.0, and a Jarque–Bera statistic of 5.3×10^4. These diagnostics motivate a heavy-tailed SMN specification, exactly the mechanism of Remark 3. Changes inside dust or cloud episodes should be automatically downweighted rather than allowed to dominate a quadratic Gaussian likelihood.

We apply Algorithm 1 with Jeffreys' prior (18) for $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\xi}, \sigma^2)$, which involves no hyperparameters, and the SMN mixing priors of Section 3.2 with the hyperparameter values used in the simulation study of Section 6. The prior for (ν, λ) is the joint truncated-exponential prior of Section 3.2, with $(c,d) = (0.02, 0.5)$ under SMN-t and $(c,d) = (0.5, 1.0)$ under SMN-sl, and $\nu \sim \text{Beta}(2,2)$, $\gamma \sim \text{Beta}(2,2)$ under SMN-cn. With $n = 1065$, Beta$(2,2)$ acts as four pseudo-observations; for an allocation count $m_\gamma \approx 90$ it changes the conditional posterior mean of ν by less than 0.001 relative to Beta$(1,1)$. The Metropolis–Hastings scales are adapted during burn-in toward the 20–50% band of Section 5.2 and frozen thereafter. Post-burn-in acceptance rates are 29.0% for ν (SMN-t) and 38.1% for γ (SMN-cn), both inside the target band. MCMC settings are $N_{\text{sim}} = 25{,}000$, $N_{\text{burn}} = 5000$, $N_{\text{thin}} = 20$ (1000 retained draws), run separately for each of the four error laws. Convergence was verified with the full battery of Section 5.3, and Table 12 reports the diagnostics for the fitted SARX$(2,1;1)_{12}$ model under SMN-t errors. Every Geweke z-score lies inside $(-1.96, 1.96)$ (maximum $|Z| = 1.21$), effective sample sizes range from 830 to 982 out of 1000 retained draws. All MCSE/SD ratios are between 3.2% and 3.5%—below the 5% gate—and the Raftery–Lewis dependence factors lie between 0.95 and 1.22.

Table 12. Convergence diagnostics of the MCMC sampler for the fitted SMN-t SARX$(2,1;1)_{12}$ model applied to the Najran GHI series on the differenced-anomaly scale.

	Raftery–Lewis			Geweke		Autocorrelations				Efficiency		
Param	**Ntotal**	**Nmin**	**Factor**	**Z-Score**	***p*-Value**	**Lag 1**	**Lag 5**	**Lag 10**	**Lag 50**	**ESS**	**MCSE**	**MCSE/SD (%)**
ϕ_{11}	1051	936	1.123	0.384	0.701	0.026	0.010	−0.006	−0.049	950	0.000705	3.24
ϕ_{12}	968	936	1.033	−1.036	0.300	0.021	−0.037	0.002	−0.064	928	0.00065	3.28
ϕ_{21}	892	936	0.953	−0.289	0.773	−0.014	−0.033	−0.024	−0.049	830	0.000501	3.47
ξ_0	971	936	1.037	−0.247	0.805	0.054	−0.016	0.001	−0.024	903	0.00855	3.33
ξ_{10}	968	936	1.033	−0.940	0.347	0.002	−0.025	−0.000	−0.005	982	0.000851	3.19
ξ_{11}	1058	936	1.130	0.788	0.430	0.054	−0.018	−0.015	−0.055	903	0.00135	3.33
σ^2	892	936	0.953	0.397	0.691	0.046	−0.019	−0.038	−0.024	861	0.0931	3.41
ν	1142	936	1.220	1.209	0.226	0.096	0.005	−0.001	0.016	834	0.00135	3.46

To evaluate the relative fit of the competing specifications, the deviance information criterion (DIC) of Spiegelhalter et al. [50] with the latent-weight-integrated deviance is computed for the SARX model under each of the four SMN error distributions. The corresponding results are summarised in Table 13.

Table 13. Deviance information criterion for the SARX$(2,1;1)_{12}$ Najran GHI model under different error distributions.

Error Distribution	$\bar{D}$	p_D	DIC	ΔDIC
SMN-t	8165.6	7.1	8172.7	—
SMN-cn	8165.5	8.9	8174.3	1.7
SMN-sl	8177.5	7.3	8184.7	12.1
SMN-n (Gaussian)	9667.8	7.1	9674.9	1502.3

Table 13 carries two messages of very different weight. The primary one is unambiguous: the Gaussian benchmark is penalised by ΔDIC $= 1502.3$—an overwhelming margin, the likelihood-level counterpart of the residual kurtosis of 37.5. Under a quadratic Gaussian likelihood the few dozen episode changes dominate the fit and inflate every variance estimate, whereas the SMN hierarchies isolate them and restore precision for the remaining $\sim 92\%$ of hours. The secondary message is that among the three non-Gaussian members the margins are 1.7 points between SMN-t and SMN-cn—a statistical tie, well within the resolution of the criterion —and 12.1 points against SMN-sl. We therefore report the two leading specifications side by side in Table 14 and regard the contaminated normal as the preferred interpretation. It fits equally well, keeps both shape parameters interior and far from their priors, reproduces the unconditional residual variance, and delivers the flagged-hour decomposition. We first examine the Bayesian estimation results obtained from the SARX$(2,1;1)_{12}$ under SMN-t errors.

Table 14. Bayesian posterior summary for the SARX$(2,1;1)_{12}$ Najran GHI model (differenced-anomaly scale) under SMN-t and SMN-cn errors.

	SMN-t				SMN-cn			
Parameter	μ	σ	L	U	μ	σ	L	U
ϕ_{11}	−0.062	0.022	−0.102	−0.016	−0.070	0.024	−0.118	−0.027
ϕ_{12}	−0.102	0.020	−0.143	−0.064	−0.082	0.020	−0.124	−0.047
ϕ_{21}	0.035	0.014	0.010	0.066	0.033	0.012	0.011	0.055
ξ_0	0.825	0.257	0.303	1.334	0.463	0.271	−0.054	0.991
ξ_{10}	0.886	0.027	0.834	0.938	0.918	0.025	0.869	0.963
ξ_{11}	0.149	0.041	0.072	0.229	0.149	0.037	0.076	0.223
σ^2	38.266	2.732	33.099	43.675	58.966	3.388	52.511	65.773
ν	2.038	0.039	2.001	2.147	0.084	0.011	0.064	0.108
γ	–	–	–	–	0.010	0.002	0.007	0.014

7.2.1. Exogenous Channel

The distributed-lag coefficients have a direct physical reading on this scale. Because the clear-sky ceiling drifts with solar declination over the 90-day window, its differenced anomaly traces the hour-to-hour evolution of the deterministic envelope. Under the SMN-t, $\hat{\xi}_{10} = 0.886$ [0.834, 0.938] measures the contemporaneous pass-through of that evolution into realised GHI, and the lagged term $\hat{\xi}_{11} = 0.149$ [0.072, 0.229] adds a small carryover. So, the cumulative response to a unit change in the ceiling is $\hat{\xi}_{10} + \hat{\xi}_{11} \approx 1.04$—near-unit pass-through completed within two hours, consistent with the transmittance-scale estimates obtained on the levels and anomaly specifications (0.93–0.95). The intercept is small on

the scale of the series (posterior mean 0.8 Wh/m^2 against a residual standard deviation of ≈ 22), absorbing the residual drift of the estimation window.

7.2.2. Autoregressive Blocks

The nonseasonal coefficients are negative and precisely estimated,

$$\hat{\phi}_{11} = -0.062\,[-0.102,\ -0.016], \qquad \hat{\phi}_{12} = -0.102\,[-0.143,\ -0.064],$$

and hour-scale changes in the irradiance anomaly partially reverse, the mean-reverting signature of cloud transients—a broken-cloud enhancement is typically followed by attenuation within the hour and conversely. The seasonal coefficient $\hat{\phi}_{21} = 0.035\,[0.010,\ 0.066]$ is modest but decisively nonzero and positive. The day-over-day evolution of the anomaly is persistent, reflecting multi-day atmospheric regimes—clear spells and dust episodes lasting several days impart same-sign day-over-day drifts. The seasonal block does exactly the work it exists for. The raw lag-12 autocorrelation of the modelled series, 0.272, is the dominant feature of its ACF, and after fitting, the residual autocorrelation at lag 12 is -0.016, statistically indistinguishable from zero. The fitted effective lag polynomial is covariance-stationary in 100% of posterior draws.

7.2.3. Tail Behaviour: The Single-t Fit and Its Boundary

The degrees-of-freedom posterior, $\hat{\nu} = 2.038$ with 95% credible interval $[2.001,\ 2.147]$ and $\Pr(\nu < 2.1 \mid \boldsymbol{y}) = 0.92$, concentrates at the variance-existence boundary $\nu > 2$ adopted in Section 2.1. A single symmetric heavy-tailed law is being asked to reconcile a very tight bulk—a typical hour predicted to within a core standard deviation of $\sqrt{\hat{\sigma}^2} \approx 6.2$ Wh/m^2—with rare changes of up to ± 240 Wh/m^2, and it does so by adopting the heaviest tails the family permits. Two consequences must be recorded. First, $\hat{\sigma}^2$ is the scale of the core component, not the marginal innovation variance. As $\nu \to 2$ the theoretical marginal variance $\sigma^2\,\nu/(\nu-2)$ diverges, so the marginal-variance reading of $\hat{\sigma}^2$ is formally lost at the boundary. Second, the boundary mass is itself evidence that one symmetric law is too rigid for this residual structure, which motivates the contaminated-normal alternative.

7.2.4. The Contaminated-Normal Reading

The SMN-cn fit, reported in Table 14, achieves the same likelihood-level fit with an interior, interpretable parameterisation. The contamination probability is $\hat{v} = 0.084\,[0.064,\ 0.108]$. Roughly, one change in twelve belongs to the inflated-variance component, and the responsibility-based flagging assigns 76 of the 1065 changes to the episode regime. Spread between 4 and 10 per clock hour across the entire daylight window, with no concentration at the window edges, the seam-crossing hour records the joint minimum. Episodes are therefore scattered transients of the cloud and dust field, not low-sun artefacts. The clean component has an estimated standard deviation of $\sqrt{\hat{\sigma}^2} \approx 7.7$ Wh/m^2, indicating that typical hour-to-hour changes are predicted with errors of only a few Wh/m^2. In contrast, the contaminated component is characterised by a precision parameter of $\hat{\gamma} = 0.010\,[0.007,\ 0.014]$. This corresponds to a variance approximately 95 times larger than that of the clean component, yielding a component standard deviation of roughly 75 Wh/m^2. This decomposition is not only interpretable but internally consistent in a way the single-t fit cannot be. The implied marginal innovation standard deviation, $\left[(1-\hat{v})\hat{\sigma}^2 + \hat{v}\,\hat{\sigma}^2/\hat{\gamma}\right]^{1/2} = 22.9$ Wh/m^2, reproduces almost exactly the OLS residual standard deviation of 22.4. Both error laws deliver the same conditional-mean estimates, and the exogenous and autoregressive posteriors agree to within a few hundredths across Table 14. This stability is precisely the robustness property that the SMN framework is designed to achieve. Rather than allowing extreme episodes to distort the regression re-

lationship, the model automatically downweights their influence while preserving the underlying conditional mean structure. Moreover, the contamination regime is a property of the series rather than of the preprocessing. The fit on the undifferenced anomaly scale delivers $\hat{v} = 0.080\,[0.062, 0.100]$ and $\hat{\gamma} = 0.007\,[0.005, 0.010]$ under a different selected order, essentially the same regime split.

7.2.5. Internal Diagnostics and Robustness

Three diagnostics qualify the preceding conclusions. First, the residual autocorrelation of the SMN-cn fit is -0.158 at lag 1 and -0.016 at lag 12. The lag-12 value is consistent with white noise, but the lag-1 value lies more than five standard errors from zero under the approximate white-noise band $\pm 1.96/\sqrt{1065} \approx \pm 0.06$, which contradicts the independent-innovation assumption of Definition 2. This is treated as a limitation below. Second, the heavy-tail finding is invariant to the scale of analysis. Conditional residual kurtosis is 41.6 on the levels specification and 37.5 on the differenced specification, against marginal values of 1.6–20.5, and the contamination regime $(\hat{v}, \hat{\gamma})$ is essentially unchanged between the anomaly and differenced-anomaly fits. Third, all convergence and efficiency diagnostics of Section 5.3 pass with wide margins.

7.2.6. Limitation: Residual Serial Dependence

The SARX-SMN model assumes serially independent innovations (Definition 2), whereas the lag-1 residual autocorrelation of -0.158 ($n = 1065$) shows that this assumption is not fully met on the differenced-anomaly scale. A plausible source is the first differencing of a series observed with measurement noise, which induces a negative lag-1 moving-average component (white measurement noise has a lag-1 autocorrelation of -0.5 once differenced); a heavy-tailed innovation law cannot account for such dependence. We have not fitted an extension with a moving-average term, which would require a SARMAX-type sampler and is left for future work. The implications for the inferential conclusions are as follows. First, the posterior intervals of Table 14 are conditional on independent innovations and may be miscalibrated; the nonseasonal autoregressive coefficients may partly absorb the neglected dependence and are best read as descriptive summaries of the short-run dynamics. Second, $(\hat{v}, \hat{\gamma})$ and the flagged hours describe the tails of the one-step residuals; because dependence is not modelled, part of the contaminated component may reflect it, so the reading of this component as cloud and dust episodes is indicative rather than confirmed. Third, the rejection of the Gaussian benchmark (ΔDIC$= 1502.3$, residual kurtosis 37.5) is unlikely to be reversed by dependence of this size: for Gaussian errors, a lag-1 correlation of -0.158 changes the deviance by about $-n \ln(1 - \hat{\rho}_1^2) \approx 27$.

Taken together, the application demonstrates the complete workflow proposed in this paper. The selected specification combines physically grounded seasonal dynamics, exogenous structure, and tail behaviour within a unified modeling framework, while admitting exact conjugate Bayesian estimation. The empirical results decisively reject the Gaussian benchmark and support an interpretable two-regime decomposition of the innovation process. This decomposition separates ordinary atmospheric variability from the transient cloud- and dust-related episodes that dominate the extreme observations.

8. Conclusions

This paper developed a complete Bayesian estimation framework for SARX models under SMN errors, closing a gap left open by the separate treatment, in the existing literature, of seasonality with SMN robustness on the one hand and seasonality with exogenous regression under Gaussianity on the other. Exploiting an exact algebraic decomposition of the SARX mean function and a single general weighted-conjugacy lemma, we derived

closed-form full conditional posteriors for the nonseasonal autoregressive coefficients, the seasonal autoregressive coefficients, and the exogenous regression coefficients under three prior families, together with the conditional posterior of the error scale and, for each SMN special case, of the associated latent weights and shape parameters. These results were integrated into a hybrid MCMC algorithm and evaluated through an extensive simulation study covering multiple SARX specifications, exogenous-regressor configurations, and SMN error models. Across all scenarios, the proposed algorithm exhibited near-nominal coverage probabilities, low RMSE values, and satisfactory convergence behaviour. Sensitivity analyses showed that these conclusions hold under the normal-gamma and g priors, under alternative Beta hyperparameters, and under rarer and more severe contamination. The practical usefulness of the methodology was further demonstrated through an application to hourly solar irradiance data from Najran, Saudi Arabia, using clear-sky irradiance as an exogenous driver. The analysis provided decisive evidence against the Gaussian benchmark. Moreover, the contaminated-normal specification decomposed the innovation process into a dominant clean regime accounting for approximately 92% of hourly observations and a smaller contaminated regime associated with dust- and cloud-induced transients. Beyond its immediate estimation results, the paper contributes a template that should extend readily to still richer seasonal-exogenous structures, including time-varying-coefficient and multivariate extensions. Natural directions for future work include Bayesian order and exogenous-lag identification for SARX models under SMN errors, multi-step Bayesian prediction under the SARX-SMN framework, SARMAX-type extensions with a moving-average component to model the residual serial dependence found in the empirical application, and asymmetric extensions of the SMN family to accommodate skewed innovations.

Author Contributions: Conceptualisation, A.A.A.; Methodology, A.A.A.; Software, A.A.A.; Validation, A.A.A. and S.A.A.; Writing—original draft, A.A.A.; Writing—review and editing, S.A.A.; Project administration, S.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R744), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, for financial support of this project.

Acknowledgments: The authors are thankful to the Deanship of Graduate Studies and Scientific Research at Najran University for funding this work under the Najran Research Funding Program grant code (NU/NRP/SERC/14/4440-2).

Data Availability Statement: The dataset used in this study is publicly accessible through the GitHub repository at https://github.com/aymanaamin/najran_solar_radiation/ (accessed on 15 September 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
2. de Castro, J.B.B.; Montini, A.d.Á. SARX model application for industrial power demand forecasting in Brazil. *Soc. Contab. Gest.* **2012**, *7*, 41–52.
3. Alharbi, F.R.; Csala, D. A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach. *Inventions* **2022**, *7*, 94.
4. Almuhaini, S.H.; Sultana, N. Forecasting long-term electricity consumption in Saudi Arabia based on statistical and machine learning algorithms to enhance electric power supply management. *Energies* **2023**, *16*, 2035.

5. Serrano, A.L.M.; Rodrigues, G.A.P.; Martins, P.H.d.S.; Saiki, G.M.; Rocha Filho, G.P.; Gonçalves, V.P.; Albuquerque, R.d.O. Statistical comparison of time series models for forecasting Brazilian monthly energy demand using economic, industrial, and climatic exogenous variables. *Appl. Sci.* **2024**, *14*, 5846.
6. West, M. Outlier models and prior distributions in Bayesian linear regression. *J. R. Stat. Soc. Ser. B (Methodol.)* **1984**, *46*, 431–439.
7. Zellner, A. Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *J. Am. Stat. Assoc.* **1976**, *71*, 400–405.
8. Lange, K.L.; Little, R.J.; Taylor, J.M. Robust statistical modeling using the *t* distribution. *J. Am. Stat. Assoc.* **1989**, *84*, 881–896.
9. Harvey, A.; Luati, A. Filtering with Heavy Tails. *J. Am. Stat. Assoc.* **2014**, *109*, 1112–1122.
10. Stock, J.H.; Watson, M.W. Why has US inflation become harder to forecast? *J. Money Credit. Bank.* **2007**, *39*, 3–33.
11. Stock, J.H.; Watson, M.W. *Modeling Inflation After the Crisis*; National Bureau of Economic Research: Cambridge, MA, USA, 2010.
12. Tsay, R.S. *Analysis of Financial Time Series*; John Wiley & Sons: Hoboken, NJ, USA, 2005; Volume 543.
13. Tanner, M.A.; Wong, W.H. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* **1987**, *82*, 528–540.
14. Gelfand, A.E.; Smith, A.F. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409.
15. Ismail, M.A. Bayesian Analysis of Seasonal Autoregressive Models. *J. Appl. Stat. Sci.* **2003**, *12*, 123–136.
16. Amin, A.A.; Emam, W.; Tashkandy, Y.; Chesneau, C. Bayesian subset selection of seasonal autoregressive models. *Mathematics* **2023**, *11*, 2878.
17. Amin, A.A. Gibbs sampler for Bayesian prediction of triple seasonal autoregressive processes. *Commun. Stat.-Theory Methods* **2025**, *54*, 1116–1134.
18. Peterka, V. Bayesian System Identification. *Automatica* **1981**, *17*, 41–53.
19. Pillonetto, G.; Ljung, L. Full Bayesian identification of linear dynamic systems using stable kernels. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2218197120. https://doi.org/10.1073/pnas.2218197120.
20. Cooper, A.; Simpson, D.; Kennedy, L.; Forbes, C.; Vehtari, A. Cross-validatory model selection for Bayesian autoregressions with exogenous regressors. *Bayesian Anal.* **2025**, *20*, 573–597.
21. Chen, X.; Zhao, S.; Liu, F.; Tao, C. Laplace Distribution Based Online Identification of Linear Systems with Robust Recursive Expectation–Maximization Algorithm. *IEEE Trans. Ind. Inform.* **2023**, *19*, 9028–9036.
22. Chen, X.; Zhang, J. Parameter Estimation of a Linear ARX Model With Skewed Heavy-Tailed Noise. *IEEE Access* **2026**, *14*, 13582–13596. https://doi.org/10.1109/ACCESS.2026.3653893.
23. Andrews, D.F.; Mallows, C.L. Scale mixtures of normal distributions. *J. R. Stat. Soc. Ser. B (Methodol.)* **1974**, *36*, 99–102.
24. West, M. On scale mixtures of normal distributions. *Biometrika* **1987**, *74*, 646–648.
25. Fernandez, C.; Steel, M.F. Bayesian regression analysis with scale mixtures of normals. *Econom. Theory* **2000**, *16*, 80–101.
26. Ferreira, G.; Castro, L.M.; Lachos, V.H.; Dias, R. Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors. *J. Appl. Stat.* **2013**, *40*, 1796–1816.
27. Maleki, M.; Nematollahi, A.R. Autoregressive models with mixture of scale mixtures of Gaussian innovations. *Iran. J. Sci. Technol. Trans. A Sci.* **2017**, *41*, 1099–1107. https://doi.org/10.1007/s40995-017-0237-6.
28. Mahmoudi, M.R.; Maleki, M.; Baleanu, D.; Nguyen, V.T.; Pho, K.H. A Bayesian approach to heavy-tailed finite mixture autoregressive models. *Symmetry* **2020**, *12*, 929.
29. Wong, C.S.; Chan, W.S.; Kam, P.L. A Student *t*-mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika* **2009**, *96*, 751–760.
30. Amin, A.A. Bayesian estimation of seasonal autoregressive models with scale-mixtures of normal errors. *Commun. Stat.-Theory Methods* **2026**, *55*, 659–675.
31. Amin, A.A. Full Bayesian analysis of seasonal autoregressive models under scale-mixtures of normal errors. *J. Stat. Comput. Simul.* **2026**, *96*, 184–204.
32. Amin, A.A.; Alghamdi, S.A. Bayesian Estimation of Autoregressive Models with Exogenous Variables Under Scale-Mixtures of Normal Errors. *Mathematics* **2026**, *14*, 2188.
33. Amin, A.A.; Alghamdi, S.A. Full Bayesian Analysis of ARX Models Under Scale-Mixtures of Normal Errors: An Application to Solar Radiation in Najran, Saudi Arabia. *Mathematics* **2026**, *14*, 3163.
34. Ljung, L. *System Identification: Theory for the User*; Prentice-Hall: Upper Saddle River, NJ, USA, 1987.
35. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994.
36. Zellner, A. On assessing prior distributions and Bayesian regression analysis with *g*-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*; Goel, P., Zellner, A., Eds.; North-Holland/Elsevier: Amsterdam, The Netherlands, 1986; pp. 233–243.
37. Fernandez, C.; Ley, E.; Steel, M.F. Benchmark priors for Bayesian model averaging. *J. Econom.* **2001**, *100*, 381–427.
38. Koop, G. *Bayesian Econometrics*; John Wiley & Sons: Chichester, UK, 2003.
39. Cabral, C.R.B.; Lachos, V.H.; Madruga, M.R. Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population. *J. Stat. Plan. Inference* **2012**, *142*, 181–200.

40. Robert, C.P.; Casella, G. The Metropolis–Hastings algorithm. In *Monte Carlo Statistical Methods*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 267–320.
41. Roberts, G.O.; Gelman, A.; Gilks, W.R. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **1997**, *7*, 110–120.
42. Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian Stat.* **1992**, *4*, 641–649.
43. Raftery, A.E.; Lewis, S.M. The number of iterations, convergence diagnostics and generic Metropolis algorithms. *Pract. Markov Chain Monte Carlo* **1995**, *7*, 763–773.
44. Link, W.A.; Eaton, M.J. On thinning of chains in MCMC. *Methods Ecol. Evol.* **2012**, *3*, 112–115.
45. LeSage, J.P. *Applied Econometrics Using MATLAB*; Technical Report; Department of Economics, University of Toronto: Toronto, ON, Canada, 1999.
46. Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76.
47. Ineichen, P. Comparison of eight clear sky broadband models against 16 independent data banks. *Sol. Energy* **2006**, *80*, 468–478.
48. Gueymard, C.A. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. *Sol. Energy* **2012**, *86*, 2145–2169.
49. Almasoud, A.H.; Gandayh, H.M. Future of solar energy in Saudi Arabia. *J. King Saud Univ.-Eng. Sci.* **2015**, *27*, 153–157.
50. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2002**, *64*, 583–639.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.