AIMS Mathematics, 11(11): 36539–36570.
DOI:10.3934/math.20261413
Received: 02 June 2026
Revised: 03 October 2026
Accepted: 14 October 2026
Published: 05 November 2026

https://www.aimspress.com/journal/Math

*Research article*

# Bayesian forecasting of ARX models under scale-mixtures of normal disturbances: an application to macroeconomics

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

[1] Department of Mathematics, College of Arts and Sciences, Najran University, Najran, Saudi Arabia
[2] Science and Engineering Research Center, Najran University, Najran, Saudi Arabia
[3] Department of Mathematical Sciences, College of Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11564, Saudi Arabia

* **Correspondence:** Email: shaalghamdi@pnu.edu.sa.

**Abstract:** Generating reliable multi-step forecasts from autoregressive models with exogenous inputs (ARX) requires a principled treatment of predictive uncertainty together with a realistic specification of the disturbance distribution. This task becomes considerably more difficult when the innovations deviate from the Gaussian assumption that is typically imposed in the literature. We address this challenge by developing a Bayesian forecasting framework for ARX models whose disturbances are drawn from the scale-mixtures of normal (SMN) family, a broad class of symmetric, heavy-tailed distributions. Within an augmented-data setting, we derive closed-form full conditional distributions for the ARX coefficient vector and the disturbance scale under three prior specifications, conjugate normal-gamma, Zellner's $g$-prior, and Jeffreys' prior, yielding multivariate normal and inverse-gamma conditionals, respectively. Beyond point forecasts, we construct the complete conditional predictive distribution of an arbitrary $K$-step forecast vector. We show this conditional predictive distribution to be multivariate normal, with a mean vector that explicitly incorporates both the autoregressive recursion and the deterministic exogenous input contributions at future time points. For each SMN subfamily, we characterize the conditional distributions of the precision weights and shape parameters. When these distributions are non-standard, their updates are carried out via Metropolis–Hastings moves within the Gibbs sampling scheme. The resulting hybrid MCMC procedure is evaluated through Monte Carlo experiments spanning various ARX configurations and all non-Gaussian SMN cases, including a paired non-Bayesian benchmark, and subsequently applied to a macroeconomic forecasting problem involving conditional multi-step forecasts of US consumer price inflation.

**Keywords:** predictive distribution; augmented data; MCMC; Gibbs sampler; Metropolis-Hastings; heavy-tailed disturbances
**Mathematics Subject Classification:** 37M10, 62F15

## 1. Introduction

Reliable forecasting is a primary objective of time series modeling, and autoregressive models with exogenous inputs (ARX) rank among the most productive tools for achieving it. The ARX framework generalizes the standard autoregressive (AR) models by allowing observed external variables to enter as additional predictors. This structure enables the analyst to incorporate known external drivers of the response, such as macroeconomic indicators, policy instruments, meteorological variables, or engineering control inputs, alongside the internal momentum of the series itself [1, 2]. This balance between dynamic self-prediction and externally driven effects has established ARX models as a standard reference in applied fields spanning macroeconometrics, financial time series, environmental statistics, signal processing, and automatic control [3–5].

From a probabilistic forecasting perspective, the ARX model provides two distinct channels through which future observations are determined: the autoregressive channel, through which past realizations of the response propagate forward, and the exogenous channel, through which observed inputs at current and lagged time points shape the conditional mean. Constructing well-calibrated predictive distributions that quantify uncertainty across all future horizons requires integrating over both channels and over the posterior uncertainty in the coefficient vector, a task for which the Bayesian paradigm is uniquely suited. Several foundational contributions established the Bayesian treatment of linear models [6,7] and time series AR models [8–10] under Gaussian errors. These results were subsequently extended to more complex settings using MCMC methods for linear models [11, 12] and time series AR models [13, 14]. Within this tradition, the conjugate normal-gamma analysis yields closed-form predictive distributions that account coherently for parameter uncertainty, providing well-calibrated predictive intervals without requiring asymptotic approximations.

Despite its theoretical appeal, the Gaussian disturbance assumption that underlies classical Bayesian ARX analysis is frequently inappropriate in practice. Macroeconomic time series, financial returns, and numerous physical measurements are characterized by disturbance distributions with substantially heavier tails than the Gaussian benchmark, arising from structural breaks, rare large shocks, and measurement contamination [15–17]. Critically for forecasting, the consequences of misspecifying disturbances as Gaussian extend beyond parameter bias. Predictive intervals constructed under the Gaussian assumption are systematically too narrow for heavy-tailed data, which lead to intervals that under-cover future realizations and overstate forecasting confidence. Furthermore, in ARX models the autoregressive mechanism amplifies the influence of heavy-tailed disturbances across multiple horizons. A single outlying shock not merely contaminates the current forecast but propagates its effect forward through the AR recursion, corrupting the entire predictive trajectory. A Bayesian forecasting framework that correctly assigns probability to large disturbances and propagates that uncertainty through the predictive recursion is therefore essential for generating reliable multi-step forecasts from ARX models fit to real-world data.

The scale-mixtures of normal (SMN) family [18,19] provides a principled mechanism for achieving this goal. The SMN family is a broad class of symmetric, heavy-tailed distributions that subsumes the Student's $t$, slash, contaminated normal, and Gaussian distributions as indexed special cases. Every member of the SMN class is representable as a Gaussian variate whose variance has been randomly inflated by an observation-specific latent multiplier. By embedding these latent multipliers into the data-augmentation framework of [11], the model acquires a conditionally Gaussian complete-data

form. This preserves the conjugate analytical tools of Bayesian linear regression, even as the marginal distribution accommodates substantial tail thickness. In each SMN case, the posterior mean of mixing weights is inversely related to the squared standardized residual of the time series, so that outlying observations automatically receive reduced weight in the likelihood, delivering the robustness that Gaussian analysis cannot provide.

Bayesian inference and forecasting for AR models under SMN disturbances are by now well established [20, 21], including seasonal specifications [22, 23], and closed-form Bayesian analysis of ARX models under Gaussian errors is likewise well documented [24–26]. Nevertheless, an important practical gap remains. A complete Bayesian multi-step forecasting framework for ARX models under SMN disturbances, together with closed-form predictive distributions under several prior families, has not been available. The present paper develops exactly this combination and, in doing so, exposes issues that are specific to the ARX setting, most notably the role of future exogenous information in multi-step prediction, which we characterize explicitly.

Specifically, the contributions of this paper are as follows:

(i) We derive closed-form full conditional distributions for the coefficient vector and disturbance scale under three prior families: conjugate normal–gamma (including Minnesota- and Litterman-type shrinkage configurations), Zellner's $g$-prior, and Jeffreys' prior. We further obtain the full conditional distributions of the SMN shape parameters in a form marginalized over the latent mixing weights.

(ii) We derive the $K$-step-ahead conditional predictive distribution, which is multivariate normal with a mean that decomposes into an autoregressive recursion component and a deterministic exogenous contribution, and with a covariance that depends on the future precision weights. We then embed the predictive draw as a direct Gibbs sampler step that integrates over the future weights, yielding the marginal heavy-tailed predictive distribution. As a result, we obtain a unified MCMC procedure that simultaneously outputs posterior estimates of all ARX parameters and multi-step-ahead predictive draws from a single chain.

(iii) We characterize the issues specific to this extension, namely the role of future exogenous information in multi-step prediction, the construction of the exogenous contribution vector, and the distinction between conditional and unconditional forecasting. We then validate the framework through an extensive Monte Carlo study covering various ARX configurations, all non-Gaussian SMN cases, and a paired non-Bayesian benchmark, followed by a macroeconomic application involving conditional multi-step forecasting of US consumer price inflation.

The remainder of the paper is structured as follows. Section 2 discusses related work. Section 3 introduces the SMN family and specifies the ARX-SMN model. Section 4 develops the augmented Bayesian predictive framework, covering the complete-data likelihood, prior distributions, joint posterior, and multi-step predictive density. Section 5 derives all required conditional distributions. Section 6 assembles the unified MCMC-based forecasting algorithm. Section 7 presents the numerical studies and Section 8 concludes the work.

## 2. Related work

The Bayesian analysis of time series models has a long and rich history. Conjugate normal-gamma results for AR processes were established by [6] and [7]. [8] derived closed-form Bayesian predictive

distributions for univariate AR processes. Bayesian model comparison for AR processes using Kullback-Leibler criteria was treated by [27, 28]. The development of data-augmentation techniques by [11] and the popularization of the Gibbs sampler by [12] transformed Bayesian inference for models lacking conjugate posteriors. [10] applied Gibbs sampling to the Bayesian AR model, [9] extended MCMC to regression models with ARMA errors, [13] conducted full-likelihood Gibbs sampling for ARMA processes, and [29, 30] applied Gibbs sampling to multiplicative seasonal ARMA models. [31] developed Gibbs-sampler-based full Bayesian analysis including prediction for double seasonal AR processes.

Bayesian analysis of ARX specifications under Gaussian disturbances has been explored in several complementary studies. [24] derived the conjugate Bayesian posterior for system identification with Gaussian errors, encompassing ARX models as a particular instance. [25] presented a comprehensive kernel-based treatment that unifies regularized system identification within a Bayesian framework, with the ARX specification arising as a special case. Recently, [32] proposed a Bayesian procedure for linear system identification that places stable kernels in a probabilistic framework. Their approach searches over high-order ARX representations, formulates the kernel as a covariance matrix, and treats impulse responses as zero-mean Gaussian random vectors. [26] proposed a cross-validation (CV)-based model-selection procedure for Bayesian ARX models with Gaussian errors. They show that joint log-score criteria are markedly more efficient for models exhibiting serial dependence, and provide some of the first rigorous finite-sample results on the uncertainty of Bayesian CV in the ARX setting, arguing that joint approaches better capture serial dynamics than standard pointwise estimators.

An important limitation of the conjugate Bayesian ARX update is its poor robustness to anomalous observations. A single large error can skew the posterior and predictive results and contaminate all future inference and prediction. Therefore, the impact of heavy-tailed disturbances on ARX estimation and forecasting has motivated several robust modifications. [33] embedded a Student's $t$ noise model within a recursive expectation-maximization algorithm for ARX parameter estimation, retaining the recursive structure of least squares with improved outlier resistance. In addition, [34] further extended this work by deriving a Laplace-noise variant that promotes residual sparsity. Moreover, [35] proposed a generalized hyperbolic skewed Student's $t$ noise model within a state-space ARX formulation estimated by online variational Bayes. These contributions highlight the growing recognition that non-Gaussian disturbances are the rule rather than the exception in ARX applications.

The SMN family has increasingly been incorporated into AR modeling. [20] provided Bayesian methods for partial linear models with SMN AR(1) errors. [21] explored theoretical properties of AR processes with SMN disturbances and presented EM treatments of these models. Most recently, [22,23] introduced Bayesian inference and forecasting of seasonal AR models with SMN disturbances. Relative to this SMN-AR literature, extending the framework to ARX is not merely a matter of appending exogenous columns to the design matrix; two genuinely new elements arise. First, the $K$-step-ahead conditional mean depends on the future values of the exogenous inputs, which raises the question of how such paths are specified in practice. Second, the predictive mean has a decomposition in which the exogenous term propagates future inputs through the AR recursion, a structure with no analogue in pure AR models. Conversely, relative to Gaussian Bayesian ARX analysis, the SMN layer renders the likelihood precision observation-specific, so that the closed-form marginal predictive of the Gaussian case is lost and a hybrid Gibbs–Metropolis–Hastings scheme becomes necessary. The contribution of the present work is therefore an incremental but self-contained extension: the first

complete Bayesian framework combining SMN disturbances with ARX dynamics and closed-form multi-step predictive distributions under three prior families.

## 3. The ARX model with SMN disturbances

This section introduces the two building blocks of the proposed forecasting framework: the SMN distribution family and the ARX model specification with its matrix-vector notation.

### *3.1. The scale-mixtures of normal (SMN) family*

The SMN family comprises a broad set of symmetric heavy-tailed distributions, with the Student's $t$, slash, contaminated normal, and Gaussian distributions appearing as special instances [18, 19].

**Definition 3.1.** A random variable $\varepsilon$ is said to follow a scale-mixture of normal (SMN) distribution with location $\mu$, scale $\tau^2 > 0$, and shape parameter $\boldsymbol{\eta}$, written $\varepsilon \sim \text{SMN}(\mu, \tau^2; \boldsymbol{\eta})$, if it possesses the conditional hierarchical structure

$$\varepsilon \mid w \sim \mathcal{N}\left(\mu,\ w^{-1}\tau^2\right), \qquad w \sim h(w \mid \boldsymbol{\eta}), \quad w > 0, \tag{3.1}$$

where $w$ is a positive random variable, independent of the Gaussian component, governed by the mixing density $h(\cdot \mid \boldsymbol{\eta})$ indexed by $\boldsymbol{\eta}$.

The latent variable $w$ acts as an observation-specific precision multiplier. Realizations of $w$ near unity preserve near-Gaussian behavior, while small values of $w$ inflate the conditional variance $w^{-1}\tau^2$ and produce heavy-tailed marginal behavior after integrating out $w$. Conditioning on $w$ restores exact Gaussian structure in the conditional likelihood, a feature that is central to the tractability of the Bayesian inference developed below. The following non-degenerate members of the SMN family are studied throughout this paper.

- **SMN-$t$ (Student's $t$ distribution).** The mixing distribution is $w \sim \text{Gamma}(\eta/2, \eta/2)$ with shape and rate both equal to $\eta/2$, where $\eta > 2$ is the degrees-of-freedom parameter. The marginal distribution of $\varepsilon$ is a Student's $t$ with $\eta$ degrees of freedom, location $\mu$, and scale $\tau^2$. Decreasing $\eta$ thickens the tails; the Gaussian distribution is recovered in the limit $\eta \to \infty$.
- **SMN-sl (Slash distribution).** The mixing distribution is $w \sim \text{Beta}(\eta, 1)$ with $\eta > 0$, so that $h(w \mid \eta) = \eta w^{\eta-1}$ on $(0, 1)$. The resulting marginal is the slash distribution, which exhibits polynomial tail decay and is well suited to data containing extreme outliers [19].
- **SMN-cn (Contaminated normal distribution).** The mixing variable $w$ follows a two-point discrete distribution:

$$h(w \mid \eta, \kappa) = \eta\, \mathbb{I}_{\{\kappa\}}(w) + (1 - \eta)\, \mathbb{I}_{\{1\}}(w), \qquad \eta,\ \kappa \in (0, 1). \tag{3.2}$$

  Here, with probability $1 - \eta$ the observation originates from a standard Gaussian component ($w = 1$), while with probability $\eta$ it is drawn from a component with inflated variance ($w = \kappa < 1$), representing an outlier-generating subprocess.
- **SMN-$n$ (Standard Gaussian).** The degenerate case $h(w = 1) = 1$ recovers the Gaussian distribution, making it a nested special case.

Computational efficiency in the Bayesian analysis stems from the hierarchical specification in (3.1). By conditioning on the latent vector $\mathbf{w}$, the complete-data model becomes Gaussian, ensuring that the location and scale parameters admit conjugate full conditional posteriors.

### *3.2. ARX model specification and notation*

Let $\{y_t\}_{t=1}^n$ be a univariate time series of interest and let $\{x_{k,t}\}_{t=1}^n$, $k = 1, \ldots, r$, be $r$ observed exogenous input sequences. A process $\{y_t\}$ is said to follow an ARX($p$; $d_1, \ldots, d_r$) model if it satisfies the stochastic difference equation

$$y_t = \theta_0 + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{k=1}^{r} \sum_{j=0}^{d_k} \psi_{kj} \, x_{k,t-j} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{3.3}$$

where $p \geq 1$ is the autoregressive order; $d_k \geq 0$ is the maximum distributed lag for exogenous input $k$; $\theta_0$ is the model intercept; $\phi_i$, $i = 1, \ldots, p$, are the autoregressive slope coefficients; $\psi_{kj}$, $j = 0, \ldots, d_k$, denote the lag-$j$ coefficients of the $k$th exogenous input; and $\varepsilon_t \overset{\text{iid}}{\sim} \text{SMN}(0, \tau^2; \boldsymbol{\eta})$ are the model disturbances.

Setting $r = 0$ recovers the standard autoregressive AR($p$) model; setting $p = 0$ yields a distributed-lag regression. The ARX formulation unifies both within a single linear framework [1]. A distinctive structural advantage of the ARX model is its global linearity in the parameter vector. This linearity allows the full conditional posterior of the coefficient vector to be obtained in closed form, and similarly allows the multi-step predictive distribution to be expressed in closed form.

Define the observation-$t$ regressor vector

$$\boldsymbol{z}_t = (1, \, y_{t-1}, \, \ldots, \, y_{t-p}, \, x_{1,t}, \, x_{1,t-1}, \, \ldots, \, x_{1,t-d_1}, \, \ldots, \, x_{r,t}, \, \ldots, \, x_{r,t-d_r})^\top \in \mathbb{R}^q \tag{3.4}$$

and the complete coefficient vector

$$\boldsymbol{\theta} = (\theta_0, \, \phi_1, \, \ldots, \, \phi_p, \, \psi_{10}, \, \psi_{11}, \, \ldots, \, \psi_{1d_1}, \, \ldots, \, \psi_{r0}, \, \ldots, \, \psi_{rd_r})^\top \in \mathbb{R}^q, \tag{3.5}$$

where $q = 1 + p + \sum_{k=1}^{r}(d_k + 1)$ counts all regression parameters. Model (3.3) then takes the compact scalar form

$$y_t = \boldsymbol{z}_t^\top \boldsymbol{\theta} + \varepsilon_t. \tag{3.6}$$

Stacking observations and treating the initial conditions $\{y_{1-p}, \ldots, y_0\}$ and $\{x_{k,1-d_k}, \ldots, x_{k,0}\}$ as fixed and known yields the matrix form

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{3.7}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, and $\boldsymbol{Z}$ is the $n \times q$ design matrix with $t$th row $\boldsymbol{z}_t^\top$. Stationarity in an ARX specification is ensured when the roots of the AR polynomial $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ fall outside the unit circle. Because this criterion pertains only to the AR dynamics, the inclusion of exogenous variables does not alter it [2]. As discussed earlier, the ARX specification augments the dynamics by incorporating exogenous regressors together with the AR lags in a single coefficient vector. This expansion changes the design matrix $\boldsymbol{Z}$ and increases the dimension of the coefficient space, requiring new closed-form posterior and predictive derivations under different prior families. These structural distinctions, along with the ability to capture external drivers, form the core contributions of this work beyond the existing AR literature.

## 4. Bayesian posterior and predictive framework

This section develops the complete Bayesian framework for joint posterior and predictive inference in the ARX model with SMN disturbances. It covers the augmented likelihood, three prior families, the joint augmented posterior, and the predictive density for future observations.

Under the SMN disturbance assumption, the marginal likelihood of $\boldsymbol{y}$ involves an intractable integral over the latent weights:

$$\mathcal{L}(\boldsymbol{\theta}, \tau^2, \boldsymbol{\eta} \mid \boldsymbol{y}) \propto \prod_{t=1}^{n} \int_0^{\infty} w_t^{1/2}\, h(w_t \mid \boldsymbol{\eta})\, (\tau^2)^{-1/2} \exp\Big\{-\frac{w_t}{2\tau^2}(y_t - \boldsymbol{z}_t^{\top}\boldsymbol{\theta})^2\Big\}\, dw_t, \tag{4.1}$$

where $\propto$ denotes equality up to a multiplicative constant that does not depend on the unknown quantities $(\boldsymbol{\theta}, \tau^2, \boldsymbol{\eta})$ over which the posterior is formed. Adopting the data-augmentation strategy of [11], we introduce $\boldsymbol{w} = (w_1, \ldots, w_n)^{\top}$ as latent auxiliary data. The resulting complete-data likelihood, which conditions on both $\boldsymbol{y}$ and $\boldsymbol{w}$, is

$$\mathcal{L}_c(\boldsymbol{\theta}, \tau^2, \boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{w}) \;\propto\; (\tau^2)^{-n/2} \exp\left\{-\frac{(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})^{\top}\boldsymbol{\Lambda}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})}{2\tau^2}\right\} \times \prod_{t=1}^{n} w_t^{1/2}\, h(w_t \mid \boldsymbol{\eta}), \tag{4.2}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(w_1, \ldots, w_n)$. The complete-data likelihood (4.2) has the form of a weighted Gaussian likelihood for $(\boldsymbol{\theta}, \tau^2)$ with precision weights $\{w_t\}$, which is the key structural property underlying the conjugate conditional posteriors derived in Section 5. In addition, the likelihood representation in (4.2) reveals that SMN errors introduce observation-level precision through the factors $w_t/\tau^2$. Observations with unusually large standardized residuals are, via the conditional distribution of $w_t$, assigned reduced precision in subsequent sampling iterations, yielding an adaptive and fully probabilistic form of outlier down-weighting. This mechanism is analogous to iteratively reweighted least squares, but embedded within a Bayesian framework that transmits all uncertainty, including that associated with the latent mixing weights, into the posterior distribution of $\boldsymbol{\theta}$.

In our prior setup, the structural component $(\boldsymbol{\theta}, \tau^2)$ is assumed to be independent of the SMN shape parameters $\boldsymbol{\eta}$:

$$\pi(\Xi \mid \boldsymbol{\Omega}) = \pi(\boldsymbol{\theta}, \tau^2 \mid \boldsymbol{\Omega}_1) \times \pi(\boldsymbol{\eta} \mid \boldsymbol{\Omega}_2), \tag{4.3}$$

where $\Xi = (\boldsymbol{\theta}, \tau^2, \boldsymbol{\eta})$ collects all unknowns and $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$ collects all hyperparameters. Three specifications for $\pi(\boldsymbol{\theta}, \tau^2 \mid \boldsymbol{\Omega}_1)$ are considered, reflecting different degrees of prior knowledge.

**(P1) Conjugate normal-gamma prior.** When substantive prior knowledge about $\boldsymbol{\theta}$ is available from theoretical considerations or previous studies, the conjugate normal–gamma specification is

$$\boldsymbol{\theta} \mid \tau^2 \;\sim\; \mathcal{N}_q\big(\boldsymbol{m}_0,\; \tau^2 \boldsymbol{S}_0\big), \qquad \tau^2 \;\sim\; \mathrm{IG}\Big(\tfrac{\alpha_0}{2}, \tfrac{\beta_0}{2}\Big), \tag{4.4}$$

with a joint density kernel given by

$$\pi(\boldsymbol{\theta}, \tau^2 \mid \boldsymbol{\Omega}_1) \propto (\tau^2)^{-(\alpha_0+q)/2-1} \exp\left\{-\frac{\beta_0 + (\boldsymbol{\theta} - \boldsymbol{m}_0)^{\top}\boldsymbol{S}_0^{-1}(\boldsymbol{\theta} - \boldsymbol{m}_0)}{2\tau^2}\right\}. \tag{4.5}$$

The set of hyperparameters $\{\boldsymbol{m}_0, \boldsymbol{S}_0, \alpha_0, \beta_0\}$ summarizes prior beliefs regarding the central tendency and variability of $\boldsymbol{\theta}$ as well as the scale $\tau^2$. When no trustworthy prior information is available, one typically employs diffuse specifications: $\boldsymbol{m}_0 = \boldsymbol{0}$, $\boldsymbol{S}_0 = c\boldsymbol{I}_q$ for large $c$, and small hyperparameters $\alpha_0$ and $\beta_0$.

**(P2) Zellner's $g$-prior.** To scale the prior precision with the information in the design matrix, the $g$-prior is often employed:

$$\pi(\boldsymbol{\theta}, \tau^2 \mid \boldsymbol{\Omega}_1) \propto (\tau^2)^{-q/2-1} \exp\left\{-\frac{g}{2\tau^2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^\top (\boldsymbol{Z}^\top \boldsymbol{Z})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})\right\}, \tag{4.6}$$

where $\bar{\boldsymbol{\theta}}$ is the prior mean and the hyperparameter $g > 0$ determines the degree of shrinkage, with $g = 1/n$ being a standard choice [28, 36]. This construction corresponds to placing the equivalent of $g^{-1}$ pseudo-observations at the prior center.

**(P3) Jeffreys' non-informative prior.** When no prior information is available, the standard non-informative reference prior is

$$\pi(\boldsymbol{\theta}, \tau^2) \propto (\tau^2)^{-1}, \qquad \tau^2 > 0, \tag{4.7}$$

which is improper but yields a proper posterior when $n > q$ [37].

The choice among (P1)–(P3) is guided by the sample size relative to the parameter dimension and by the availability of substantive prior information. (P1) is recommended when $n$ is small relative to $q$ or when trustworthy informative priors exist. In particular, setting $\boldsymbol{m}_0$ to a random-walk configuration with a tight prior variance on the first AR coefficient recovers Minnesota/Litterman-type shrinkage [38], so that (P1) nests the shrinkage priors commonly used in macroeconomic forecasting. (P2) scales the prior precision with the information in $\boldsymbol{Z}^\top \boldsymbol{Z}$, the choice $g = 1/n$ corresponding to unit information. The known limitations of fixed-$g$ choices [39] motivate hyper-$g$ extensions, which we do not pursue here. (P3) is the appropriate default when $n \gg q$: As the prior-sensitivity study in Section 7.2 demonstrates, the posterior is then essentially invariant across all four specifications.

**Priors for SMN shape parameters.** In line with [40] and [20], we adopt weakly informative conjugate priors for $\boldsymbol{\eta}$, defined on a case-by-case basis for each SMN distribution:

(i) **SMN-$t$:** The parameter $\eta$ is assigned a truncated exponential prior with rate $\rho > 0$, supported on $(\eta > 2)$, where the hyperparameter $\rho$ follows a Uniform$(c_0, d_0)$ distribution.

(ii) **SMN-sl:** The same truncated exponential structure is used, but with support restricted to $(\eta > 1)$, and with $\rho \sim$ Uniform$(c_0, d_0)$.

(iii) **SMN-cn:** Independent beta priors are adopted, with $\eta \sim$ Beta$(\alpha_\eta, \beta_\eta)$ and $\kappa \sim$ Beta$(\alpha_\kappa, \beta_\kappa)$, where all hyperparameters $\alpha_\eta, \beta_\eta, \alpha_\kappa, \beta_\kappa > 0$. For the contamination precision we use the uniform choice $\alpha_\kappa = \beta_\kappa = 1$. $\kappa$ is weakly identified when the expected number of contaminated observations is small, and our simulations indicate that the flat prior yields better-calibrated posterior inference in this regime.

Multiplying the complete-data likelihood (4.2) by the factored prior (4.3) and treating $\boldsymbol{w}$ as additional unknowns yields the joint augmented posterior:

$$\pi(\boldsymbol{\theta}, \tau^2, \boldsymbol{w}, \boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\Omega}) \;\propto\; \pi(\Xi \mid \boldsymbol{\Omega}) \times (\tau^2)^{-n/2} \exp\left\{-\frac{(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})^\top \boldsymbol{\Lambda} (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})}{2\tau^2}\right\}$$

$$\times \prod_{t=1}^{n} w_t^{1/2}\, h(w_t \mid \boldsymbol{\eta}). \tag{4.8}$$

The structure of (3.7) ensures global linearity in $\boldsymbol{\theta}$, so the joint posterior (4.8) becomes a quadratic exponential in this parameter. As a result, the full conditional distribution of $\boldsymbol{\theta}$ is available in closed form as a multivariate normal, as established in Section 5.

Let $\boldsymbol{Y}_f = (y_{n+1}, \ldots, y_{n+K})^\top$ denote the vector of $K$ future observations, which is the primary object of interest for forecasting. The predictive density of $\boldsymbol{Y}_f$ given the observed data $\boldsymbol{y}$ is obtained by integrating out all unknowns from the joint predictive-posterior:

$$\pi(\boldsymbol{Y}_f \mid \boldsymbol{y}, \boldsymbol{\Omega}) = \int \pi(\boldsymbol{Y}_f \mid \boldsymbol{\theta}, \tau^2, \boldsymbol{w}_f, \boldsymbol{\eta}, \boldsymbol{y})\; p(\boldsymbol{w}_f \mid \boldsymbol{\eta})\; \pi(\boldsymbol{\theta}, \tau^2, \boldsymbol{w}, \boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{\Omega})\; d\boldsymbol{w}_f\, d\boldsymbol{\theta}\, d\tau^2\, d\boldsymbol{w}\, d\boldsymbol{\eta}, \tag{4.9}$$

where $\boldsymbol{w}_f = (w_{n+1}, \ldots, w_{n+K})^\top$ collects the latent precision weights of the future disturbances and $p(\boldsymbol{w}_f \mid \boldsymbol{\eta}) = \prod_{l=1}^{K} h(w_{n+l} \mid \boldsymbol{\eta})$ is their mixing density given the shape parameters. The marginal predictive distribution is therefore a mixture over both the posterior parameter uncertainty and the future mixing weights. The multivariate normal distribution derived in Section 5.3 is the *conditional* predictive distribution given $\boldsymbol{w}_f$. The marginal object in (4.9) is obtained by additionally averaging over $\boldsymbol{w}_f \sim h(\cdot \mid \boldsymbol{\eta})$, which is precisely the operation performed by the forecasting step of the algorithm in Section 6. This integral is analytically intractable but can be approximated empirically by the MCMC strategy. At each retained iteration, a draw from the conditional predictive distribution provides a single sample from the joint predictive density, and the collection of such draws over all retained iterations constitutes a Monte Carlo approximation to $\pi(\boldsymbol{Y}_f \mid \boldsymbol{y}, \boldsymbol{\Omega})$.

## 5. Closed-form conditional distributions

We now derive the full set of conditional distributions needed to construct the MCMC-based forecasting algorithm for ARX models.

### *5.1. Conditionals for $\boldsymbol{\theta}$ and $\tau^2$*

By extracting the $\boldsymbol{\theta}$-dependent components of the joint posterior (4.8) and completing the square, we obtain the full conditional:

$$\boldsymbol{\theta} \mid \tau^2, \boldsymbol{w}, \boldsymbol{\eta}, \boldsymbol{y} \sim \mathcal{N}_q(\boldsymbol{m}^*, \boldsymbol{C}^*),$$

with $\boldsymbol{m}^*$ and $\boldsymbol{C}^*$ taking forms that depend on the prior specification and are given as follows.

**(P1) Normal-gamma prior:**

$$\boldsymbol{C}^* = \tau^2\Big(\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{Z} + \boldsymbol{S}_0^{-1}\Big)^{-1}, \qquad \boldsymbol{m}^* = \boldsymbol{C}^* \tau^{-2}\Big(\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{y} + \boldsymbol{S}_0^{-1} \boldsymbol{m}_0\Big). \tag{5.1}$$

**(P2) $g$-prior:**

$$\boldsymbol{C}^* = \tau^2(\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{Z} + g\boldsymbol{Z}^\top \boldsymbol{Z})^{-1}, \qquad \boldsymbol{m}^* = \boldsymbol{C}^* \tau^{-2}\Big(\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{y} + g\boldsymbol{Z}^\top \boldsymbol{Z}\, \bar{\boldsymbol{\theta}}\Big). \tag{5.2}$$

**(P3) Jeffreys' prior:**

$$\boldsymbol{C}^* = \tau^2(\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{Z})^{-1}, \qquad \boldsymbol{m}^* = (\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{Z})^{-1}\, \boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{y}. \tag{5.3}$$

In all three cases the posterior mean $\boldsymbol{m}^*$ has the structure of a weighted least-squares (WLS) estimator combining data information through the precision-weighted cross-products $\boldsymbol{Z}^\top \boldsymbol{\Lambda} \boldsymbol{y}$ with prior information. Observations associated with small $w_t$, those with large standardized residuals under the current parameter values, contribute less to these cross-products, realizing the automatic outlier down-weighting inherent to the SMN framework without any manual intervention.

Extracting the $\tau^2$ terms from (4.8) reveals an inverse-gamma kernel, yielding

$$\tau^2 \mid \boldsymbol{\theta}, \boldsymbol{w}, \boldsymbol{\eta}, \boldsymbol{y} \sim \mathrm{IG}(a^*/2, b^*/2),$$

with $a^*$ and $b^*$ depending on the prior specification as follows.

**(P1) Normal-gamma prior:**

$$a^* = \alpha_0 + n + q, \qquad b^* = \beta_0 + (\boldsymbol{\theta} - \boldsymbol{m}_0)^\top \boldsymbol{S}_0^{-1} (\boldsymbol{\theta} - \boldsymbol{m}_0) + (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})^\top \boldsymbol{\Lambda} (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}). \tag{5.4}$$

**(P2) $g$-prior:**

$$a^* = n + q, \qquad b^* = g(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^\top (\boldsymbol{Z}^\top \boldsymbol{Z})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) + (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})^\top \boldsymbol{\Lambda} (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}). \tag{5.5}$$

**(P3) Jeffreys' prior:**

$$a^* = n, \qquad b^* = (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta})^\top \boldsymbol{\Lambda} (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}). \tag{5.6}$$

### *5.2. Conditional distributions for SMN mixing components*

The structure of the full conditional distributions for both the latent weights $\{w_t\}$ and the shape parameters $\boldsymbol{\eta}$ is dictated by the SMN subfamily employed. Let $\tilde{e}_t^2 = (y_t - \boldsymbol{z}_t^\top \boldsymbol{\theta})^2/\tau^2$ denote the squared standardized residual of observation $t$.

**SMN-$t$ subfamily.**

For each index $t$, the corresponding latent weight $w_t$ follows a gamma full conditional:

$$w_t \mid \cdot \;\sim\; \mathrm{Gamma}\left(\frac{\eta + 1}{2}, \;\; \frac{\tilde{e}_t^2 + \eta}{2}\right), \quad t = 1, \ldots, n. \tag{5.7}$$

An increase in $\tilde{e}_t^2$ results in a reduced posterior mean for $w_t$, illustrating how the model down-weights observations with unusually large residuals.

Rather than updating $\eta$ from its conditional distribution given $(\boldsymbol{w}, \boldsymbol{\theta}, \tau^2)$, which would couple the update to the latent weights, we integrate the $w_t$ out analytically and update $\eta$ from its marginalized full conditional

$$\begin{aligned} \log \pi(\eta \mid \boldsymbol{\theta}, \tau^2, \rho, \boldsymbol{y}) = C \;+\; & n\Big[\log \Gamma\Big(\tfrac{\eta+1}{2}\Big) - \log \Gamma\Big(\tfrac{\eta}{2}\Big) - \tfrac{1}{2}\log(\pi\tau^2) - \tfrac{1}{2}\log \eta\Big] \\ & - \frac{\eta + 1}{2} \sum_{t=1}^{n} \log\left(1 + \frac{\tilde{e}_t^2}{\eta}\right) \;-\; \rho\,(\eta - 2), \quad \eta > 2, \end{aligned} \tag{5.8}$$

where $C$ collects terms independent of $\eta$ and the final term is the truncated exponential prior with rate $\rho$. This kernel is the product of $n$ scaled Student-$t$ densities with $\eta$ degrees of freedom and

scale $\tau^2$. Samples are drawn via a random-walk Metropolis–Hastings update on $\log \eta$ (proposal $\eta^\dagger = \eta^{(\ell-1)} \exp(\sigma_\eta z)$, $z \sim \mathcal{N}(0,1)$, with Jacobian $\eta^\dagger/\eta^{(\ell-1)}$ included in the acceptance ratio), which we found to mix well in practice. The conditional-on-$\boldsymbol{w}$ alternative possesses a degenerate high-$\eta$ region for the slash model below, so the marginalized form is used uniformly. The rate hyperparameter $\rho$ follows a truncated gamma full conditional: $\rho \mid \eta \sim \text{TGamma}(2,\ \eta - 2,\ (c_0, d_0))$, which follows from the truncated exponential prior for $\eta$ with rate $\rho$ combined with $\rho \sim \text{Uniform}(c_0, d_0)$. We note that this conditional does not depend on the data, since the latent weights enter $\eta \mid \rho$ but not $\rho \mid \eta$.

**SMN-sl subfamily.**

The latent weight $w_t$ follows a truncated gamma full conditional:

$$w_t \mid \cdot \ \sim\ \text{TGamma}\Big(\eta + \tfrac{1}{2},\ \ \frac{\tilde{e}_t^2}{2},\ \ (0,1)\Big), \quad t = 1, \ldots, n, \tag{5.9}$$

restricted to the interval $(0, 1)$ so that it aligns with the support of the underlying beta mixing law. As for the SMN-$t$ case, $\eta$ is updated from its $w$-marginalized full conditional

$$\log \pi(\eta \mid \boldsymbol{\theta}, \tau^2, \rho, \boldsymbol{y}) = C \ +\ n \log \eta \ +\ \sum_{t=1}^{n} \log I\Big(\eta + \tfrac{1}{2},\ \tfrac{1}{2}\tilde{e}_t^2\Big) \ -\ \rho\,(\eta - 1), \quad \eta > 1, \tag{5.10}$$

where $I(s, a) = \int_0^1 w^{s-1} e^{-aw}\, dw$ denotes the raw integral underlying the slash marginal density (evaluated by a rapidly convergent power series for $a < 1$ and by the regularized incomplete gamma function otherwise), and the final term is the truncated exponential prior with rate $\rho$. The update uses the same log-scale random-walk Metropolis–Hastings scheme as above. Two features of this model deserve comment. First, the $w_t$-conditional (5.9) has a rate that does not involve $\eta$. Consequently, an update of $\eta$ conditional on $\boldsymbol{w}$ possesses a degenerate high-$\eta$ region. For large $\eta$ the latent weights concentrate at unity, the resulting conditional for $\eta$ no longer reflects the data, and the chain cannot return, which motivates the marginalized form (5.10). Second, the slash shape parameter is only weakly identified. The marginalized posterior of $\eta$ typically exhibits a heavy right tail, since for large $\eta$ the slash density approaches the Gaussian likelihood and the prior damping decays at the rate $\rho$. This behavior is reflected in the wide Monte Carlo posterior intervals reported in Section 7.1. The rate hyperparameter follows the corrected truncated gamma full conditional $\rho \mid \eta \sim \text{TGamma}(2,\ \eta - 1,\ (c_0, d_0))$, derived analogously to the SMN-$t$ case from the truncated exponential prior on $(\eta > 1)$ with rate $\rho$.

**SMN-cn subfamily.**

The full conditional of $w_t$ is a two-point discrete distribution:

$$w_t \mid \cdot \ =\ \begin{cases} \kappa & \text{with probability } \dfrac{\omega_{1t}}{\omega_{1t} + \omega_{2t}}, \\[2ex] 1 & \text{with probability } \dfrac{\omega_{2t}}{\omega_{1t} + \omega_{2t}}, \end{cases} \tag{5.11}$$

where the unnormalized component weights are

$$\omega_{1t} = \eta\, \kappa^{1/2} \exp\Big\{-\tfrac{1}{2}\kappa\, \tilde{e}_t^2\Big\}, \qquad \omega_{2t} = (1 - \eta) \exp\Big\{-\tfrac{1}{2}\tilde{e}_t^2\Big\}. \tag{5.12}$$

The ratio $\omega_{1t}/(\omega_{1t} + \omega_{2t})$ serves as a posterior outlier probability for observation $t$. It approaches one for observations with large $\tilde{e}_t^2$ and small contamination precision $\kappa$, labeling such observations as likely draws from the inflated-variance component.

The contamination proportion $\eta$ follows a beta full conditional:

$$\eta \mid \cdot \;\sim\; \text{Beta}\Big(\alpha_\eta + n_\kappa,\;\; \beta_\eta + n - n_\kappa\Big), \tag{5.13}$$

where $n_\kappa = \sum_{t=1}^{n} \mathbb{I}\{w_t = \kappa\}$ counts the observations assigned to the contaminated regime.

In contrast, the contamination precision $\kappa$ has a non-standard full conditional:

$$\pi(\kappa \mid \cdot) \;\propto\; \kappa^{\alpha_\kappa - 1}(1-\kappa)^{\beta_\kappa - 1} \prod_{t=1}^{n} [\omega_{1t} + \omega_{2t}], \quad \kappa \in (0,1), \tag{5.14}$$

and is updated via a Metropolis–Hastings move using a logit-normal proposal on the unit-interval support, with the logit Jacobian $\kappa(1-\kappa)$ included in the acceptance ratio, so that the target in (5.14) is sampled exactly.

### *5.3. Conditional predictive distribution of $\boldsymbol{Y}_f$*

To construct the conditional predictive distribution of $\boldsymbol{Y}_f = (y_{n+1}, \ldots, y_{n+K})^\top$, we write the $l$th future observation ($l = 1, \ldots, K$) as

$$y_{n+l} = \theta_0 + \sum_{i=1}^{p} \phi_i \, y_{n+l-i} + \boldsymbol{e}_{n+l}^\top \boldsymbol{\psi} + \varepsilon_{n+l}, \tag{5.15}$$

where $\varepsilon_{n+l} \overset{\text{iid}}{\sim} \text{SMN}(0, \tau^2; \boldsymbol{\eta})$, $y_{n+l-i}$ for $l - i \le 0$ is an observed value, and $y_{n+l-i}$ for $l - i > 0$ is a future value that is part of $\boldsymbol{Y}_f$. $\boldsymbol{e}_{n+l}^\top \boldsymbol{\psi}$ collects the exogenous contributions at future time $n + l$, and the exogenous inputs are assumed known or pre-specified at all future time points.

Stacking equations (5.15) for $l = 1, \ldots, K$, the system of future observations can be written in matrix form as

$$\boldsymbol{A}\, \boldsymbol{Y}_f = \boldsymbol{B}\, \boldsymbol{h} + \boldsymbol{u} + \boldsymbol{\varepsilon}_f, \tag{5.16}$$

where $\boldsymbol{h} = (y_n, y_{n-1}, \ldots, y_{n-p+1})^\top$, $\boldsymbol{\varepsilon}_f = (\varepsilon_{n+1}, \ldots, \varepsilon_{n+K})^\top$, and the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and vector $\boldsymbol{u}$ are defined as follows. The lower-triangular $K \times K$ autoregressive companion matrix is

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_1 & 1 & 0 & \cdots & 0 \\ -\phi_2 & -\phi_1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\phi^*_{K-1} & \cdots & -\phi_2 & -\phi_1 & 1 \end{pmatrix}_{K \times K}, \tag{5.17}$$

the $K \times p$ matrix of lagged-observation coefficients is

$$\boldsymbol{B} = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ \phi_2 & \phi_3 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \phi^*_K & \cdots & \cdots & 0 \end{pmatrix}_{K \times p}, \tag{5.18}$$

with $\phi_l^* = \phi_l$ for $l \le p$ and $\phi_l^* = 0$ for $l > p$, and $\boldsymbol{u} = (u_1, \ldots, u_K)^\top$ is the $K$-vector of deterministic exogenous input contributions at future time points, with the explicit expression

$$u_l \;=\; \boldsymbol{e}_{n+l}^\top \boldsymbol{\psi} \;=\; \sum_{k=1}^{r} \sum_{j=0}^{d_k} \psi_{kj}\, x_{k,\,n+l-j}, \qquad l = 1, \ldots, K, \tag{5.19}$$

which is fully determined by the pre-specified future paths of the exogenous inputs and the exogenous coefficients in $\boldsymbol{\theta}$.

Since $\boldsymbol{A}$ is lower triangular with unit diagonal, it is invertible, so $\boldsymbol{Y}_f = \boldsymbol{A}^{-1}(\boldsymbol{B}\boldsymbol{h} + \boldsymbol{u}) + \boldsymbol{A}^{-1}\boldsymbol{\varepsilon}_f$. Conditional on all model parameters and on the future latent precision weights $\boldsymbol{w}_f = (w_{n+1}, \ldots, w_{n+K})^\top$, the vector $\boldsymbol{A}^{-1}\boldsymbol{\varepsilon}_f$ is multivariate normal with mean zero and covariance $\tau^2 \boldsymbol{A}^{-1}\boldsymbol{\Lambda}_f^{-1}(\boldsymbol{A}^{-1})^\top$, where $\boldsymbol{\Lambda}_f = \mathrm{diag}(w_{n+1}, \ldots, w_{n+K})$. Therefore, the conditional predictive distribution of $\boldsymbol{Y}_f$ is multivariate normal:

$$\boldsymbol{Y}_f \mid \boldsymbol{\theta}, \tau^2, \boldsymbol{w}_f, \boldsymbol{y} \;\sim\; \mathcal{N}_K\big(\boldsymbol{\mu}_f,\; \boldsymbol{C}_f\big), \tag{5.20}$$

where the conditional predictive mean and covariance are

$$\boldsymbol{\mu}_f = \boldsymbol{A}^{-1}(\boldsymbol{B}\boldsymbol{h} + \boldsymbol{u}), \qquad \boldsymbol{C}_f = \tau^2 \boldsymbol{A}^{-1}\boldsymbol{\Lambda}_f^{-1}\big(\boldsymbol{A}^{-1}\big)^\top. \tag{5.21}$$

We emphasize that (5.20)–(5.21) define the *conditional* predictive distribution given $\boldsymbol{w}_f$, as reflected by the conditioning set. The marginal predictive distribution $\pi(\boldsymbol{Y}_f \mid \boldsymbol{y}, \boldsymbol{\Omega})$ of (4.9) is the scale mixture obtained by averaging (5.20) over $\boldsymbol{w}_f \sim h(\cdot \mid \boldsymbol{\eta})$ and over the posterior. This averaging is carried out within the MCMC scheme of Section 6.

The conditional predictive mean $\boldsymbol{\mu}_f$ decomposes into two interpretable components. First, $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{h}$ propagates the most recently observed values of the response forward through the AR recursion. Second, $\boldsymbol{A}^{-1}\boldsymbol{u}$ captures the cumulative effect of the exogenous inputs over the $K$-step forecast horizon. This additive decomposition is unique to the ARX model and is absent in the pure AR special case; it is what enables ARX models to produce forecasts that respond to known or projected future values of the exogenous drivers. Moreover, in contrast to the conditional predictive distribution under Gaussian disturbances (which does not depend on $\boldsymbol{w}_f$), the SMN-based predictive covariance $\boldsymbol{C}_f$ involves the future precision weights $\boldsymbol{\Lambda}_f$. By sampling $\boldsymbol{w}_f$ from its SMN distribution at each MCMC iteration and then drawing $\boldsymbol{Y}_f$ from $\mathcal{N}_K(\boldsymbol{\mu}_f, \boldsymbol{C}_f)$, the Gibbs sampler integrates over the predictive uncertainty induced by the heavy-tailed disturbance distribution. Thus, the resulting predictive intervals become wider than Gaussian intervals when the SMN assigns material probability to large disturbances.

## 6. Hybrid MCMC for forecasting ARX models under SMN disturbances

Using the conditional distributions derived in Section 5, we construct a coherent Gibbs-within-Metropolis–Hastings algorithm for joint inference and forecasting of ARX models under SMN disturbances, as laid out in Algorithm 1. The algorithm accepts as inputs the observed data, ARX orders, forecast horizon, and prior hyperparameters; it outputs posterior draws for all model parameters, multi-step predictive draws for $\boldsymbol{Y}_f$, and posterior summary statistics.

**Algorithm 1** Hybrid MCMC for forecasting ARX models under SMN disturbances.

**Inputs:** observations $\boldsymbol{y}$; exogenous inputs $\{x_{k,t}\}$; ARX orders $(p; d_1, \ldots, d_r)$; forecast horizon $K$; SMN subfamily; prior hyperparameters $\boldsymbol{\Omega}$; MH step size $\sigma_\eta$ for $\log \eta$.
Construct design matrix $\boldsymbol{Z}$ with $t$th row $\boldsymbol{z}_t^\top$ as in (3.4); form $\boldsymbol{h} = (y_n, \ldots, y_{n-p+1})^\top$ and $\boldsymbol{u}$ from known future exogenous values via (5.19); build matrices $\boldsymbol{A}$ (5.17) and $\boldsymbol{B}$ (5.18).
**Initialize:** set $\boldsymbol{\theta}^{(0)}$ and $(\tau^2)^{(0)}$ from OLS of model (3.7); set $w_t^{(0)} = 1$ for all $t$; set $\eta^{(0)} = 6$ (SMN-$t$), $\eta^{(0)} = 2$ (SMN-sl), and $(\eta^{(0)}, \kappa^{(0)}) = (0.5, 0.5)$ (SMN-cn).
Specify: total iterations $M_{\text{tot}}$; burn-in $M_{\text{b}}$; thinning interval $M_{\text{s}}$.
**for** $\ell = 1, 2, \ldots, M_{\text{tot}}$ **do**
    **[S1]** Draw $\boldsymbol{\theta}^{(\ell)} \sim \mathcal{N}_q(\boldsymbol{m}^*, \boldsymbol{C}^*)$ from Eqs (5.1)–(5.3).
    **[S2]** Draw $(\tau^2)^{(\ell)} \sim \text{IG}(a^*/2, b^*/2)$ from Eqs (5.4)–(5.6).
    **[S3]** Draw $\boldsymbol{Y}_f^{(\ell)} \sim \mathcal{N}_K(\boldsymbol{\mu}_f, \boldsymbol{C}_f)$ from Eqs (5.20)–(5.21), after first drawing $w_{n+l}^{(\ell)} \sim h(\cdot \mid \boldsymbol{\eta}^{(\ell-1)})$ for $l = 1, \ldots, K$ (this two-step draw realizes the Monte Carlo integration over the future weights in the marginal predictive density (4.9)).
    **[S4]** SMN-specific steps:
    *SMN-t:*
        Draw $w_t^{(\ell)} \sim \text{Gamma}\left(\frac{\eta^{(\ell-1)}+1}{2},\ \frac{\tilde{e}_t^2}{2} + \frac{\eta^{(\ell-1)}}{2}\right)$, $\forall t$, using $\boldsymbol{\theta}^{(\ell)}$ and $(\tau^2)^{(\ell)}$ for $\tilde{e}_t^2$.
        Propose $\eta^\dagger = \eta^{(\ell-1)} \exp(\sigma_\eta z)$, $z \sim \mathcal{N}(0,1)$; accept with probability $\min\{1, r_\eta\}$ where $r_\eta = \pi(\eta^\dagger \mid \cdot)\, \eta^\dagger / \pi(\eta^{(\ell-1)} \mid \cdot)\, \eta^{(\ell-1)}$, using the marginalized target (5.8).
        Draw $\rho^{(\ell)} \sim \text{TGamma}(2,\ \eta^{(\ell)} - 2,\ (c_0, d_0))$.
    *SMN-sl:*
        Draw $w_t^{(\ell)} \sim \text{TGamma}\left(\eta^{(\ell-1)} + \frac{1}{2},\ \frac{\tilde{e}_t^2}{2},\ (0,1)\right)$, $\forall t$.
        Propose $\eta^\dagger = \eta^{(\ell-1)} \exp(\sigma_\eta z)$; accept with probability $\min\{1, r_\eta\}$ as above, using the marginalized target (5.10).
        Draw $\rho^{(\ell)} \sim \text{TGamma}(2,\ \eta^{(\ell)} - 1,\ (c_0, d_0))$.
    *SMN-cn:*
        Draw $w_t^{(\ell)}$ from the two-point distribution (5.11)–(5.12), $\forall t$.
        Let $n_\kappa^{(\ell)} = \sum_t \mathbb{I}\{w_t^{(\ell)} = \kappa^{(\ell-1)}\}$; draw $\eta^{(\ell)} \sim \text{Beta}(\alpha_\eta + n_\kappa^{(\ell)},\ \beta_\eta + n - n_\kappa^{(\ell)})$.
        Propose $\kappa^\dagger \sim \text{LogitN}(\kappa^{(\ell-1)}, \tau_\kappa^2)$; accept with probability $\min\{1, r_\kappa\}$ where $r_\kappa = \pi(\kappa^\dagger \mid \cdot)/\pi(\kappa^{(\ell-1)} \mid \cdot) \cdot q(\kappa^{(\ell-1)})/q(\kappa^\dagger)$, using target (5.14) (the Hastings ratio includes the logit Jacobian $\kappa(1-\kappa)$).
**end for**
Discard iterations $1, \ldots, M_{\text{b}}$ (burn-in); retain every $M_{\text{s}}$-th draw thereafter.
Evaluate convergence using multiple convergence diagnostics.
Compute posterior summaries for $\boldsymbol{\theta}$, $\tau^2$, and $\boldsymbol{\eta}$: posterior mean, standard deviation, and 95% equal-tailed credible intervals.
Compute predictive summaries for $\boldsymbol{Y}_f$: predictive mean, standard deviation, and 95% predictive intervals.

Initializing the chain at the OLS estimate of $\boldsymbol{\theta}$, using the corresponding residual variance as $(\tau^2)^{(0)}$, and setting all mixing weights to one positions the sampler close to the likelihood mode under the Gaussian limit. From this starting point, the updates in step S4 enable the chain to learn the level of

tail inflation supported by the data. This starting scheme is reproducible and does not depend on the specific SMN subfamily.

Step S3 draws future precision weights $\{w_{n+l}^{(\ell)}\}$ from the current SMN mixing distribution $h(\cdot \mid \boldsymbol{\eta}^{(\ell-1)})$ and then draws $\boldsymbol{Y}_f^{(\ell)}$ from the conditional multivariate normal (5.20). This integrates over both the parameter posterior and the future disturbance uncertainty in a single coherent step. Because the weights are drawn before the forecast, the resulting $\boldsymbol{Y}_f^{(\ell)}$ are samples from the marginal predictive distribution (4.9), not merely from (5.20). The dominant cost at each iteration is the $q \times q$ matrix inversion required in step S1. For large $q$, a Cholesky factorization updated at each step reduces the per-iteration cost from $O(q^3)$ to $O(q^2)$.

Convergence is evaluated using multiple diagnostics. The **Geweke $z$-score** [41] compares early and late portions of the chain, with convergence indicated by $z$-scores in $(-1.96, 1.96)$. The **Raftery-Lewis diagnostic** [42] suggests appropriate burn-in and chain lengths for estimating posterior quantiles. We also inspect **autocorrelation functions** and thin the chain to reduce dependence. In addition, the effective sample size (ESS) and Monte Carlo standard errors of the posterior means are reported, and, for the application, split-$\hat{R}$ across independent chains is monitored. Additional implementation notes appear in [43] and [30]. As a stationarity safeguard, the maximum modulus of the AR companion-matrix eigenvalues of $\Phi(z)$ is monitored for all retained draws. In the designs of Section 7 no non-stationary draw was ever encountered (Section 7.1), so the unconstrained sampler of step S1 operates entirely within the stationary region at these sample sizes.

## 7. Numerical studies

The proposed MCMC algorithm is evaluated through two complementary numerical studies: a Monte Carlo simulation experiment to assess estimation and forecasting accuracy under controlled conditions, and a real-data illustration applying the framework to macroeconomic forecasting.

### *7.1. Monte Carlo simulation experiment*

Three ARX models are designed to cover a range of model structures and all three non-Gaussian SMN subfamilies; their specifications are summarized in Table 1. Model I is an ARX(1; 0) process with a single AR lag and one contemporaneous exogenous input, driven by SMN-$t$ disturbances with $\eta = 5$ degrees of freedom. Model II is an ARX(1; 1, 0) process with one AR lag and two exogenous inputs, namely, the first with one distributed lag and the second contemporaneous only, driven by SMN-cn disturbances with contamination parameters $\eta = 0.2$ and $\kappa = 0.1$. Model III is an ARX(2; 0, 1) process with two AR lags and two exogenous inputs, the first contemporaneous and the second with one lag, driven by SMN-sl disturbances with tail parameter $\eta = 2$. In all configurations, each exogenous series is generated independently as a stationary AR(1) process with coefficient 0.7 and standard Gaussian innovations, ensuring that the exogenous inputs carry predictive information while being exogenous to the ARX disturbances.

For each model, $R = 500$ independent realizations are generated and analyzed with Algorithm 1 under Jeffreys' prior and the corresponding SMN priors. The prior hyperparameters are as follows: for SMN-$t$, $(c_0, d_0) = (0.02, 0.5)$ with $\eta^{(0)} = 6$; for SMN-sl, $(c_0, d_0) = (0.01, 1.0)$ with $\eta^{(0)} = 2$; for SMN-cn, $\alpha_\eta = \beta_\eta = 2$ and $\alpha_\kappa = \beta_\kappa = 1$ (a uniform prior on $\kappa$), with $(\eta^{(0)}, \kappa^{(0)}) = (0.5, 0.5)$; all chains are initialized at $w_t^{(0)} = 1$. The MCMC settings are $M_{\text{tot}} = 25{,}000$, $M_{\text{b}} = 5000$, and $M_{\text{s}} = 20$, yielding 1000

retained draws per replication. Forecasts are produced in a fixed-origin manner: For each replication the model is estimated once on the first $n$ observations, and at every retained iteration the entire $K$-step path is drawn jointly from the conditional predictive distribution (5.20)–(5.21), with parameters never re-estimated along the horizon. The exogenous inputs over the forecast horizon, generated by the same AR(1) law, are treated as known (conditional) inputs by both the Bayesian procedure and the non-Bayesian benchmark, so that the comparison is not affected by asymmetric information.

**Table 1.** Design specifications for the three Monte Carlo ARX models. $K = 4$ forecast steps in all cases.

| Model | $\theta_0$ | $\phi_1$ | $\phi_2$ | $\psi_{10}$ | $\psi_{11}$ | $\psi_{20}$ | $\psi_{21}$ | $\tau^2$ | $n$ | Error |
|---|---|---|---|---|---|---|---|---|---|---|
| I | 1.5 | 0.6 | — | 0.7 | — | — | — | 1.0 | 150 | SMN-$t$ ($\eta = 5.0$) |
| II | 2.0 | 0.5 | — | 0.7 | 0.4 | 0.6 | — | 1.0 | 200 | SMN-cn ($\eta = 0.2, \kappa = 0.1$) |
| III | 1.8 | 0.4 | 0.3 | 0.5 | — | 0.6 | 0.4 | 1.0 | 150 | SMN-sl ($\eta = 2.0$) |

For each replication, posterior summary statistics for $\boldsymbol{\theta}$, $\tau^2$, and the SMN shape parameters are recorded: posterior mean $\hat{\mu}$, posterior standard deviation $\hat{s}$, and the lower and upper limits $(\hat{L}, \hat{U})$ of the 95% equal-tailed credible interval. These are averaged across the 500 replications to yield the reported Monte Carlo averages. Forecast accuracy is assessed via the root mean squared error (RMSE):

$$\mathrm{RMSE}_l = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(y_{n+l}^{(r)} - \hat{y}_{n+l}^{(r)}\right)^2}, \quad l = 1, \ldots, K, \tag{7.1}$$

where $\hat{y}_{n+l}^{(r)}$ is the posterior predictive mean for horizon $l$ in replication $r$. Coverage probability (CP) of 95% credible and predictive intervals is also reported. When the CP is close to 95%, the resulting posterior and predictive intervals exhibit proper uncertainty calibration.

Prior to summarizing performance over all replications, we begin with a representative example that highlights the dynamics of the proposed algorithm. We consider one simulated dataset drawn from Model I under SMN-$t$ disturbances and present the resulting Bayesian estimation and forecasting outputs in Table 2 and Figure 1, together with the complete set of convergence diagnostics described in Section 6 for this run in Table 3.

As evidenced by these outputs, the posterior distributions of the ARX coefficients are centered close to their true values and exhibit modest dispersion. Furthermore, the 95% credible intervals successfully encompass the true parameters. All four realized forecasts fall within the corresponding 95% predictive intervals, with predictive standard deviations between 1.50 and 2.33 that are commensurate with the distances between the predictive means and the realized values. Convergence is evaluated using the diagnostics described in Section 6, all of which indicate that the MCMC sampler has mixed adequately. Table 3 quantifies this assessment. All Geweke $z$-scores lie within $(-1.96, 1.96)$, with $p$-values of at least 0.79; and first-lag autocorrelations do not exceed 0.07 and decay to zero by lag 50. The effective sample sizes range from 826 to 888 out of the 1000 retained draws, with Raftery–Lewis dependence factors between 0.95 and 1.22, all close to unity. The Monte Carlo standard error of each posterior mean amounts to roughly 3.4% of the corresponding posterior standard deviation, so that Monte Carlo error is negligible relative to posterior uncertainty. Notably, the MH-updated shape parameter $\eta$ attains an effective sample size of 840, confirming that the hybrid Gibbs–Metropolis–Hastings scheme mixes

well even for the non-standard conditional blocks. This assessment is consistent with the behavior observed in the trace plots and marginal posteriors in Figure 1.

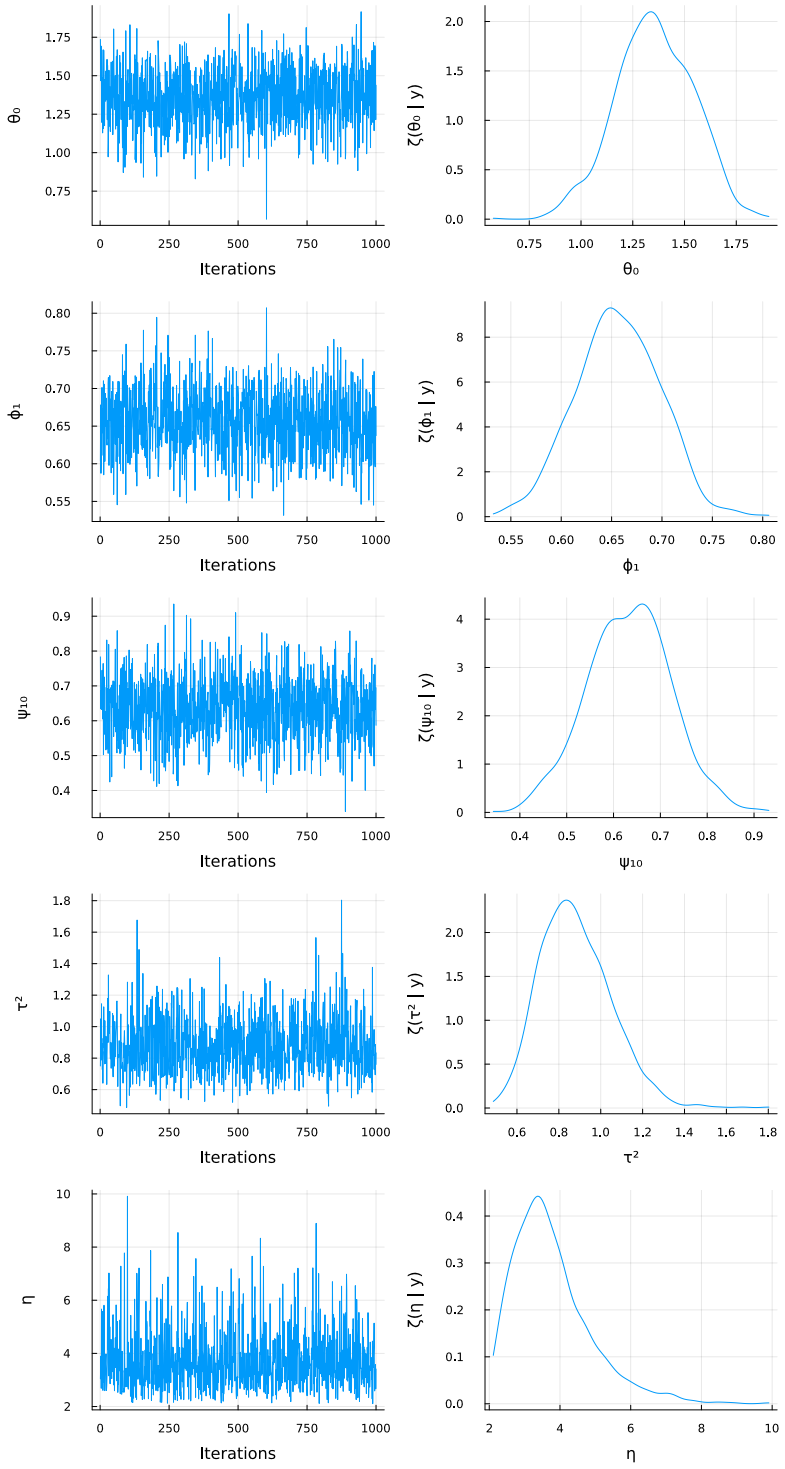



**Figure 1.** Trace plots and posterior densities of ARX parameters for a dataset simulated from Model I.

**Table 2.** Posterior and predictive summaries for one representative dataset from Model I (SMN-$t$, $\eta = 5$). $\hat{\mu}$: posterior/predictive mean; $\hat{s}$: posterior/predictive SD; $(\hat{L}, \hat{U})$: 95% credible/predictive interval.

| Parameter | True | $\hat{\mu}$ | $\hat{s}$ | $\hat{L}$ | $\hat{U}$ |
|---|---|---|---|---|---|
| | | | *Posterior estimates* | | |
| $\theta_0$ | 1.5 | 1.471 | 0.206 | 1.069 | 1.869 |
| $\phi_1$ | 0.6 | 0.575 | 0.048 | 0.486 | 0.674 |
| $\psi_{10}$ | 0.7 | 0.665 | 0.090 | 0.490 | 0.841 |
| $\tau^2$ | 1.0 | 0.998 | 0.218 | 0.644 | 1.509 |
| $\eta$ | 5.0 | 3.549 | 1.289 | 2.226 | 7.127 |
| | | | *Bayesian forecasts* | | |
| $y_{n+1}$ | 5.293 | 3.743 | 1.498 | 0.897 | 6.878 |
| $y_{n+2}$ | 1.399 | 2.928 | 1.652 | -0.418 | 6.119 |
| $y_{n+3}$ | 4.190 | 2.867 | 1.714 | -0.339 | 6.130 |
| $y_{n+4}$ | 5.155 | 3.704 | 2.332 | -0.027 | 7.411 |

**Table 3.** Convergence diagnostics for the representative dataset from Model I.

| | 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 (%) |
| $\theta_0$ | 1051 | 936 | 1.123 | 0.193 | 0.847 | 0.065 | −0.005 | −0.023 | −0.004 | 884 | 0.0069 | 3.36 |
| $\phi_1$ | 892 | 936 | 0.953 | 0.235 | 0.814 | 0.060 | 0.022 | −0.067 | 0.011 | 854 | 0.0016 | 3.42 |
| $\psi_{10}$ | 1142 | 936 | 1.220 | 0.155 | 0.877 | 0.022 | −0.038 | 0.012 | −0.020 | 888 | 0.0030 | 3.36 |
| $\tau^2$ | 892 | 936 | 0.953 | 0.240 | 0.811 | 0.026 | 0.000 | −0.011 | −0.016 | 826 | 0.0076 | 3.48 |
| $\eta$ | 1051 | 936 | 1.123 | −0.256 | 0.798 | 0.028 | 0.004 | −0.018 | −0.048 | 840 | 0.0445 | 3.45 |

We now present the Monte Carlo results aggregated across all replications and ARX model specifications, as reported in Tables 4–6. These results establish key properties of the proposed algorithm. For all three model specifications and each SMN variant, the Monte Carlo posterior means align closely with the corresponding true parameter values. The structural parameters $(\boldsymbol{\theta}, \tau^2)$ exhibit relatively small posterior standard deviations, confirming that the MCMC algorithm accurately identifies the posterior location. Empirical coverage probabilities also remain near the intended 95% level. Across the 21 parameter entries and 12 forecast entries of the three designs, the overall average coverage is 94.9%, with all entries lying between 92.4% and 98.2%.

Such behavior reflects well-calibrated uncertainty quantification that is free from systematic bias in interval coverage.

The average RMSE values for $K = 4$-step-ahead forecasts range from 1.37 to 2.09 and increase with horizon length, as expected from the accumulation of disturbance uncertainty over longer horizons. Moreover, at every horizon the average predictive standard deviation closely matches the corresponding empirical RMSE (e.g., 1.37 versus 1.37 for Model I at $l = 1$), indicating that the predictive variances are well calibrated to the realized forecast errors. The empirical coverage probabilities of the 95% predictive intervals are near nominal for all three SMN cases.

**Table 4.** Monte Carlo averages for Model I (SMN-$t$, $\eta = 5.0$). CP: empirical coverage probability of a 95% credible/predictive interval.

| Parameter | True | $\hat{\mu}$ | $\hat{s}$ | $\hat{L}$ | $\hat{U}$ | CP (%) |
|---|---|---|---|---|---|---|
| | | | *Posterior estimates* | | | |
| $\theta_0$ | 1.5 | 1.544 | 0.194 | 1.166 | 1.923 | 94.2 |
| $\phi_1$ | 0.6 | 0.587 | 0.045 | 0.500 | 0.674 | 94.6 |
| $\psi_{10}$ | 0.7 | 0.711 | 0.079 | 0.556 | 0.866 | 95.8 |
| $\tau^2$ | 1.0 | 1.031 | 0.200 | 0.686 | 1.462 | 96.6 |
| $\eta$ | 5.0 | 7.324 | 5.792 | 2.819 | 21.493 | 97.8 |
| | | | *Forecast RMSE and predictive coverage* | | | |
| $y_{n+1}$ | — | 1.37 | 1.37 | 1.14 | 6.37 | 94.6 |
| $y_{n+2}$ | — | 1.58 | 1.61 | 0.69 | 6.82 | 93.2 |
| $y_{n+3}$ | — | 1.64 | 1.69 | 0.57 | 7.00 | 94.0 |
| $y_{n+4}$ | — | 1.71 | 1.71 | 0.54 | 7.08 | 92.4 |

**Table 5.** Monte Carlo averages for Model II (SMN-cn, $\eta = 0.2$, $\kappa = 0.1$).

| Parameter | True | $\hat{\mu}$ | $\hat{s}$ | $\hat{L}$ | $\hat{U}$ | CP (%) |
|---|---|---|---|---|---|---|
| | | | *Posterior estimates* | | | |
| $\theta_0$ | 2.0 | 2.038 | 0.175 | 1.695 | 2.380 | 94.2 |
| $\phi_1$ | 0.5 | 0.493 | 0.036 | 0.422 | 0.564 | 94.2 |
| $\psi_{10}$ | 0.7 | 0.705 | 0.093 | 0.523 | 0.885 | 96.2 |
| $\psi_{11}$ | 0.4 | 0.407 | 0.107 | 0.197 | 0.616 | 94.8 |
| $\psi_{20}$ | 0.6 | 0.607 | 0.072 | 0.467 | 0.748 | 94.8 |
| $\tau^2$ | 1.0 | 0.983 | 0.204 | 0.621 | 1.419 | 97.4 |
| $\eta$ | 0.2 | 0.265 | 0.086 | 0.123 | 0.454 | 94.4 |
| $\kappa$ | 0.1 | 0.126 | 0.039 | 0.064 | 0.216 | 93.8 |
| | | | *Forecast RMSE and predictive coverage* | | | |
| $y_{n+1}$ | — | 1.59 | 1.71 | 0.45 | 7.71 | 96.0 |
| $y_{n+2}$ | — | 1.94 | 1.90 | 0.02 | 7.97 | 94.2 |
| $y_{n+3}$ | — | 2.07 | 1.95 | -0.10 | 7.97 | 93.6 |
| $y_{n+4}$ | — | 2.09 | 1.96 | -0.18 | 7.93 | 93.0 |

**Table 6.** Monte Carlo averages for Model III (SMN-sl, $\eta = 2.0$).

| Parameter | True | $\hat{\mu}$ | $\hat{s}$ | $\hat{L}$ | $\hat{U}$ | CP (%) |
|---|---|---|---|---|---|---|
| | | | *Posterior estimates* | | | |
| $\theta_0$ | 1.8 | 1.878 | 0.261 | 1.366 | 2.388 | 95.2 |
| $\phi_1$ | 0.4 | 0.389 | 0.070 | 0.251 | 0.526 | 93.8 |
| $\phi_2$ | 0.3 | 0.300 | 0.062 | 0.178 | 0.421 | 94.2 |
| $\psi_{10}$ | 0.5 | 0.507 | 0.086 | 0.339 | 0.675 | 95.0 |
| $\psi_{20}$ | 0.6 | 0.596 | 0.111 | 0.379 | 0.814 | 94.2 |
| $\psi_{21}$ | 0.4 | 0.415 | 0.127 | 0.167 | 0.663 | 94.4 |
| $\tau^2$ | 1.0 | 1.113 | 0.271 | 0.667 | 1.708 | 96.8 |
| $\eta$ | 2.0 | 4.286 | 7.111 | 1.286 | 18.547 | 98.2 |
| | | | *Forecast RMSE and predictive coverage* | | | |
| $y_{n+1}$ | — | 1.51 | 1.50 | 3.05 | 8.67 | 94.0 |
| $y_{n+2}$ | — | 1.42 | 1.63 | 2.82 | 8.93 | 96.4 |
| $y_{n+3}$ | — | 1.64 | 1.79 | 2.56 | 9.28 | 96.4 |
| $y_{n+4}$ | — | 1.79 | 1.88 | 2.47 | 9.47 | 93.8 |

To benchmark the proposed method against a frequentist alternative, the OLS estimator of the ARX model with classical Gaussian intervals is applied to the *same* 500 simulated datasets, providing a paired comparison in which sampling variability is common to both procedures. The results, reported in Table 7, are informative. The two approaches deliver nearly identical point-forecast accuracy. The RMSEs agree to within Monte Carlo error for every horizon and model, which reflects the fact that robust and Gaussian likelihoods yield similar conditional-mean estimates in these designs. The decisive difference lies in predictive calibration. The Gaussian intervals of the OLS benchmark systematically under-cover the actual observations, and the under-coverage *worsens with the forecast horizon*, for instance, from 93.4% to 86.0% across horizons 1–4 in Model I, because their variance cannot account for the heavy-tailed disturbance distribution, an effect that accumulates through the autoregressive recursion. By contrast, the SMN-based predictive intervals remain near nominal at all horizons in all three designs. This confirms the central motivation of the paper. The benefit of modeling heavy-tailed disturbances accrues not through sharper point predictions but through correctly calibrated predictive uncertainty in multi-step forecasting.

A further concern is whether the unconstrained Gibbs updates of Section 5 can produce non-stationary parameter draws, since the AR-polynomial stationarity condition is not imposed during sampling. Step S1 of Algorithm 1 draws $\boldsymbol{\theta}$ from an unrestricted multivariate normal. To assess this directly, we monitored the maximum modulus of the companion-matrix eigenvalues of $\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ for every retained posterior draw across 100 replications of each design. The fraction of non-stationary draws was 0.0000% in all three models (300,000 draws examined in total), and the 95th percentiles of the maximum root were 0.68, 0.58, and 0.84 for Models I–III, respectively, well inside the unit circle and close to the roots implied by the true generating processes (0.60, 0.50, and 0.68). The unconstrained sampler therefore operates entirely within the stationary region at these sample sizes, and posterior stationarity is verified rather than assumed. A constrained (e.g., rejection-based) variant of Step S1 would only be warranted for very short samples or heavily persistent designs,

and is left as an implementation option.

All these Monte Carlo results confirm that the MCMC sampler accurately estimates ARX parameters and generates reliable forecasts under various SMN error laws, with accurate posterior estimation, near-nominal predictive coverage, in contrast to the under-covering Gaussian benchmark of Table 7, and satisfactory convergence diagnostics.

**Table 7.** Non-Bayesian benchmark: fixed-origin OLS-ARX point forecasts and Gaussian predictive intervals evaluated on the same 500 simulated datasets as the Bayesian analysis.

| | OLS-ARX (Gaussian) | | Bayesian ARX-SMN | |
|---|---|---|---|---|
| Horizon | RMSE | CP (%) | RMSE | CP (%) |
| *Model I (SMN-t), cf. Table 4* | | | | |
| $y_{n+1}$ | 1.37 | 93.4 | 1.37 | 94.6 |
| $y_{n+2}$ | 1.58 | 90.6 | 1.58 | 93.2 |
| $y_{n+3}$ | 1.65 | 88.0 | 1.64 | 94.0 |
| $y_{n+4}$ | 1.72 | 86.0 | 1.71 | 92.4 |
| *Model II (SMN-cn), cf. Table 5* | | | | |
| $y_{n+1}$ | 1.59 | 94.4 | 1.59 | 96.0 |
| $y_{n+2}$ | 1.97 | 91.4 | 1.94 | 94.2 |
| $y_{n+3}$ | 2.10 | 88.6 | 2.07 | 93.6 |
| $y_{n+4}$ | 2.11 | 90.2 | 2.09 | 93.0 |
| *Model III (SMN-sl), cf. Table 6* | | | | |
| $y_{n+1}$ | 1.52 | 94.4 | 1.51 | 94.0 |
| $y_{n+2}$ | 1.41 | 95.2 | 1.42 | 96.4 |
| $y_{n+3}$ | 1.65 | 92.0 | 1.64 | 96.4 |
| $y_{n+4}$ | 1.79 | 88.2 | 1.79 | 93.8 |

### *7.2. An application to macroeconomics: forecasting US CPI inflation*

To demonstrate the practical forecasting value of the proposed framework, we apply it to the monthly annualized US Consumer Price Index (CPI) inflation rate, defined as $y_t = 1200 \ln(P_t/P_{t-1})$, where $P_t$ is the seasonally adjusted CPI index. This application is motivated by three considerations. First, AR-based models are the standard benchmark for inflation forecasting [4, 44], and the ARX extension allows well-established macroeconomic predictors to be incorporated formally. Second, the empirical literature consistently documents non-Gaussian, heavy-tailed CPI innovation distributions arising from oil price shocks, financial crises, and policy discontinuities [3, 16], making the SMN framework directly relevant. Third, the multi-step forecasting horizon is the operationally relevant output: Central banks and researchers require forecast paths, not just one-step predictions.

The ARX specification incorporates two exogenous predictors. $x_{1,t}$ captures monthly movements in the 10-year U.S. Treasury yield (in basis points, serving as an indicator of monetary policy tightening). $x_{2,t}$ measures annualized monthly growth in real industrial production (in percent, capturing real activity and demand pressure conditions). These series, together with the dependent variable, are drawn from the FRED database* and cover the period January 1986–December 2024, yielding 468

*https://fred.stlouisfed.org/

monthly observations. Figure 2 illustrates their temporal evolution. Stationarity is supported for all variables by augmented Dickey–Fuller tests at the 5% level.

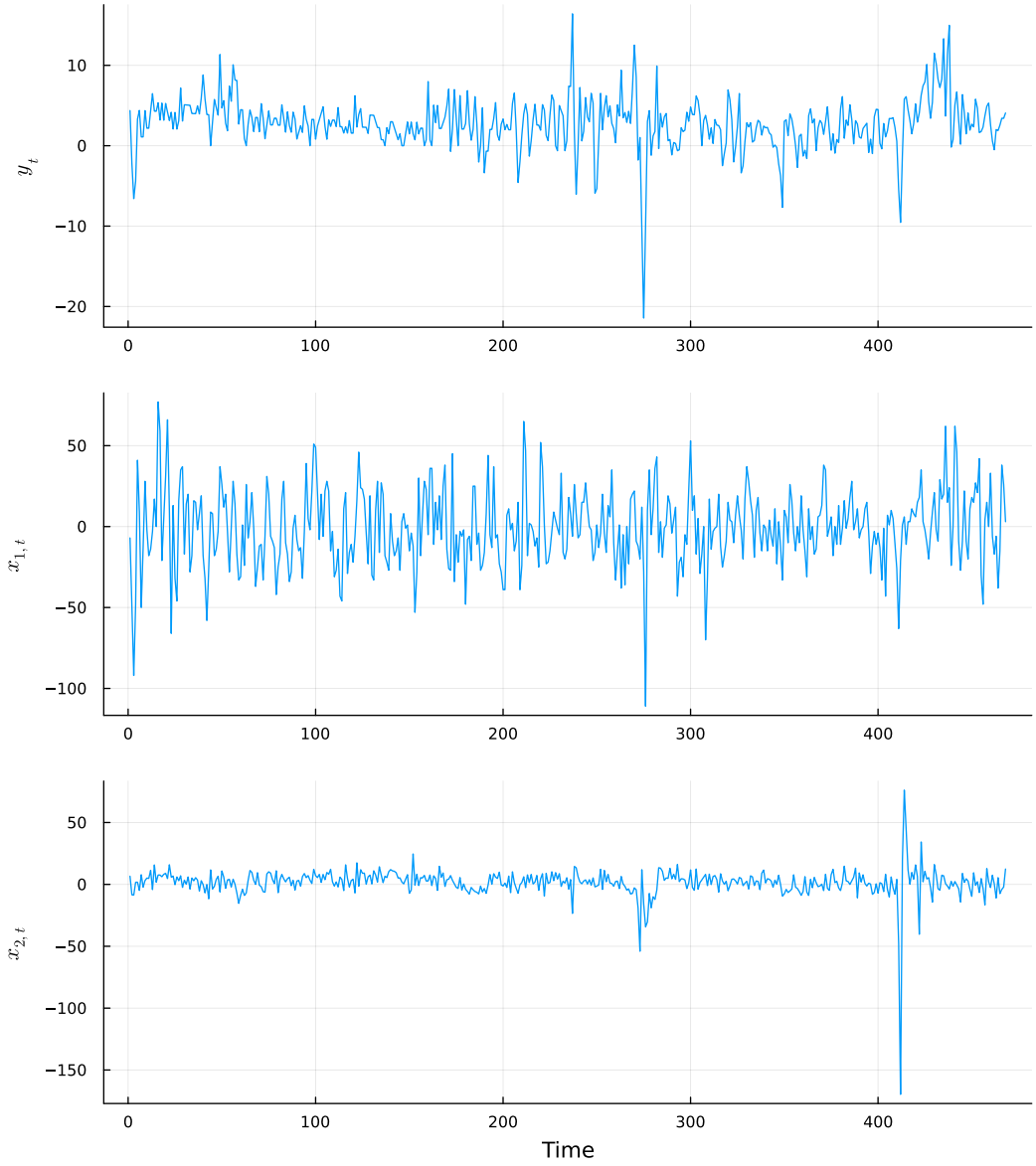



**Figure 2.** Monthly series used in the application: top, annualized CPI inflation $y_t$; middle, monthly change in the 10-year Treasury yield $x_{1,t}$; bottom, annualized industrial production growth $x_{2,t}$.

ARX orders are selected by minimizing the Bayesian information criterion (BIC) over the grid $p \in \{1, \ldots, 5\}$ and $d_k \in \{0, \ldots, 5\}$ for $k = 1, 2$. The BIC-optimal specification is ARX(3; 0, 0): three autoregressive lags, with both exogenous predictors entering contemporaneously only, giving $q = 6$ regression parameters.

Before applying the SMN-MCMC algorithm, we assess the suitability of heavy-tailed disturbances

by fitting the ARX(3; 0, 0) model by OLS and examining the residuals. The OLS residuals exhibit sample kurtosis of approximately 12.0 (compared with the Gaussian value of 3.0), and a normal quantile-quantile plot shows clear departure from Gaussianity in both tails. These diagnostics strongly motivate the use of SMN-$t$ disturbances.

The full sample of 468 observations is divided into 444 months for model estimation and 24 months reserved for forecast assessment. After the three autoregressive lags, the effective conditioning sample comprises $n_{\text{obs}} = 441$ observations. Using the training data, we fit an ARX(3; 0, 0) specification with SMN-$t$ disturbances, adopting Jeffreys' prior for $(\boldsymbol{\theta}, \tau^2)$ and a truncated exponential prior for $\eta$ with $(c_0, d_0) = (0.02, 0.5)$. The MCMC settings mirror those of the simulation study: $M_{\text{tot}} = 25{,}000$, $M_{\text{b}} = 5000$, $M_{\text{s}} = 20$, yielding 1000 retained post-burn-in draws. A $K = 24$ months forecast is generated simultaneously. Tables 8 and 9 contain the posterior and predictive summaries, Figure 3 illustrates the trace behavior and marginal posterior shapes, and Figure 4 displays the actual and predicted annualized CPI inflation rates over a 24-month horizon.

To verify convergence quantitatively, the model was refitted with four independent chains (1000 retained draws each). All split-$\hat{R}$ values are at most 1.001 and effective sample sizes range from 3554 to 3959 out of the 4000 pooled draws, with Metropolis–Hastings acceptance rates for $\eta$ between 33% and 36%; these complement the diagnostics reported in Table 10 for a single chain. Moreover, none of the 4000 retained draws corresponds to a non-stationary AR polynomial (maximum-modulus companion root: mean 0.52, 95th percentile 0.64), confirming that the unconstrained sampler operates well within the stationary region for this application. This assessment is consistent with the patterns visible in Figure 3.

The posterior estimates in Table 8 are economically interpretable, although the contemporaneous specification warrants a careful reading. Three statistically significant AR lags ($\phi_1 \approx 0.46$, $\phi_2 \approx -0.14$, and $\phi_3 \approx 0.10$) document the well-known inertia of price dynamics, with the first lag accounting for the dominant share of autoregressive momentum [4]. The contemporaneous coefficient on Treasury-yield changes ($\hat{\psi}_{10} \approx 0.014$) is positive, with a 95% credible interval excluding zero. We deliberately refrain from a structural interpretation of this coefficient. In a single-equation model with a contemporaneous regressor, the positive association is equally consistent with a Fisher-effect channel, in which nominal yields rise because markets price higher expected inflation, as with a pass-through of monetary conditions, which would typically operate with lags and, on impact, in the opposite direction. More broadly, yields, output, and inflation are jointly determined, so $\hat{\psi}_{10}$ should be read as a conditional association rather than a causal elasticity. Similarly, the positive coefficient on industrial-production growth ($\hat{\psi}_{20} \approx 0.032$) reflects demand-side co-movement with inflation. The estimated degrees-of-freedom parameter $\hat{\eta} \approx 3.3$, with a 95% credible interval of $[2.3, 4.6]$, provides unambiguous evidence of heavy-tailed CPI innovation shocks, validating the choice of SMN-$t$ disturbances over the Gaussian limit.

The predictive results in Table 9 indicate strong multi-step forecast performance. The predictive mean tracks the actual out-of-sample values closely over the full 24-step horizon, with |Bias| below 4.3 points at all horizons. The largest deviation, 4.22 points, occurs at the first step, where realized inflation spiked to 6.36%. All 24 held-out observations fall within the reported 95% predictive intervals, as shown in Figure 4. This is reassuring for this particular forecast path, but a single 24-step trajectory is not sufficient to establish predictive calibration; a rolling-origin or repeated forecast evaluation would be required for a formal assessment of empirical coverage.

**Table 8.** Posterior summaries for the ARX(3; 0, 0) model fitted to monthly US CPI inflation under SMN-$t$ disturbances.

| Parameter | $\hat{\mu}$ | $\hat{s}$ | $\hat{L}$ | $\hat{U}$ |
|---|---|---|---|---|
| $\theta_0$ | 1.574 | 0.188 | 1.201 | 1.926 |
| $\phi_1$ | 0.457 | 0.043 | 0.369 | 0.542 |
| $\phi_2$ | −0.141 | 0.047 | −0.230 | −0.048 |
| $\phi_3$ | 0.096 | 0.038 | 0.016 | 0.168 |
| $\psi_{10}$ | 0.014 | 0.005 | 0.005 | 0.023 |
| $\psi_{20}$ | 0.032 | 0.012 | 0.005 | 0.052 |
| $\tau^2$ | 3.326 | 0.428 | 2.580 | 4.178 |
| $\eta$ | 3.309 | 0.595 | 2.348 | 4.634 |

**Table 9.** Bayesian multi-step-ahead predictive summaries for monthly US CPI inflation under the ARX(3; 0, 0) SMN-$t$ model. $\hat{\mu}_f$: predictive mean; $\hat{s}_f$: predictive SD; $(\hat{L}_f, \hat{U}_f)$: 95% predictive interval; Bias = actual − predictive mean.

| | Actual | $\hat{\mu}_f$ | $\hat{s}_f$ | $\hat{L}_f$ | $\hat{U}_f$ | Bias |
|---|---|---|---|---|---|---|
| $y_{n+1}$ | 6.360 | 2.140 | 2.791 | −3.033 | 7.985 | 4.220 |
| $y_{n+2}$ | 4.107 | 3.258 | 3.060 | −3.054 | 8.908 | 0.849 |
| $y_{n+3}$ | 1.476 | 2.865 | 3.091 | −2.991 | 9.278 | −1.389 |
| $y_{n+4}$ | 4.064 | 2.436 | 2.947 | −3.686 | 8.160 | 1.628 |
| $y_{n+5}$ | 1.936 | 2.678 | 3.068 | −2.876 | 7.782 | −0.742 |
| $y_{n+6}$ | 2.687 | 2.555 | 3.432 | −3.422 | 8.641 | 0.132 |
| $y_{n+7}$ | 2.346 | 3.017 | 3.202 | −2.946 | 9.059 | −0.671 |
| $y_{n+8}$ | 5.789 | 3.210 | 3.257 | −2.960 | 9.354 | 2.579 |
| $y_{n+9}$ | 4.672 | 3.245 | 3.057 | −2.636 | 8.678 | 1.427 |
| $y_{n+10}$ | 1.639 | 3.359 | 3.157 | −3.155 | 9.500 | −1.720 |
| $y_{n+11}$ | 1.761 | 2.537 | 3.331 | −3.508 | 8.490 | −0.776 |
| $y_{n+12}$ | 2.307 | 1.668 | 3.008 | −4.417 | 7.534 | 0.639 |
| $y_{n+13}$ | 3.714 | 1.806 | 3.400 | −4.380 | 8.068 | 1.908 |
| $y_{n+14}$ | 4.907 | 3.055 | 3.451 | −3.304 | 10.069 | 1.852 |
| $y_{n+15}$ | 5.306 | 2.876 | 2.926 | −3.393 | 8.738 | 2.430 |
| $y_{n+16}$ | 2.602 | 3.030 | 3.422 | −2.858 | 10.029 | −0.428 |
| $y_{n+17}$ | 0.583 | 2.970 | 3.282 | −3.529 | 9.426 | −2.387 |
| $y_{n+18}$ | −0.502 | 2.539 | 3.497 | −3.485 | 9.117 | −3.041 |
| $y_{n+19}$ | 2.011 | 2.217 | 3.970 | −4.256 | 8.580 | −0.206 |
| $y_{n+20}$ | 1.885 | 2.144 | 3.511 | −3.979 | 8.198 | −0.259 |
| $y_{n+21}$ | 2.557 | 2.064 | 3.976 | −3.840 | 7.895 | 0.493 |
| $y_{n+22}$ | 3.423 | 2.760 | 3.464 | −3.355 | 9.384 | 0.663 |
| $y_{n+23}$ | 3.405 | 3.034 | 3.172 | −3.136 | 9.174 | 0.371 |
| $y_{n+24}$ | 4.072 | 3.261 | 3.308 | −3.270 | 9.279 | 0.811 |

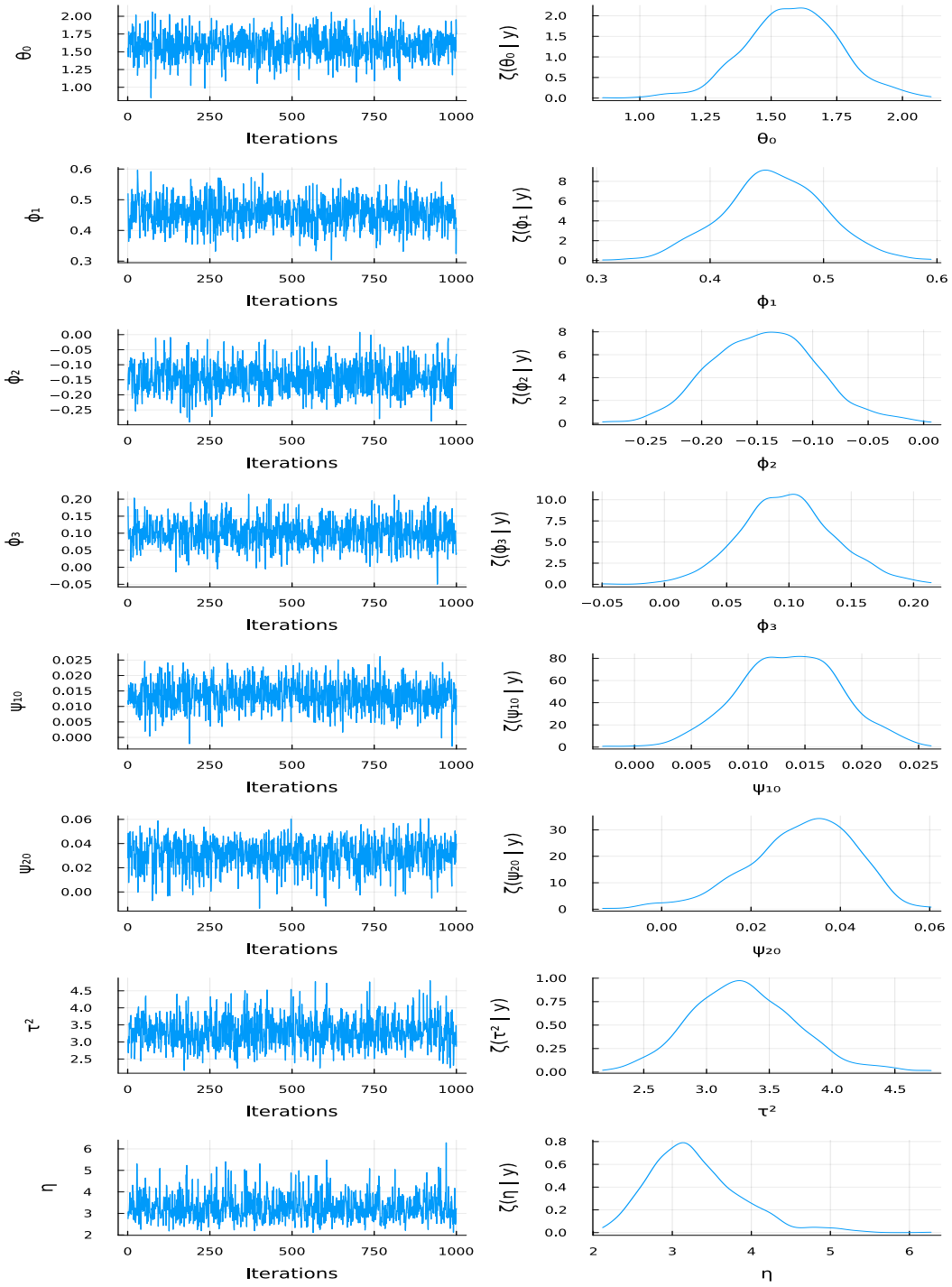



**Figure 3.** Trace plots and marginal posterior densities for the ARX(3; 0, 0) model estimated on monthly US CPI inflation under SMN-*t* disturbances.

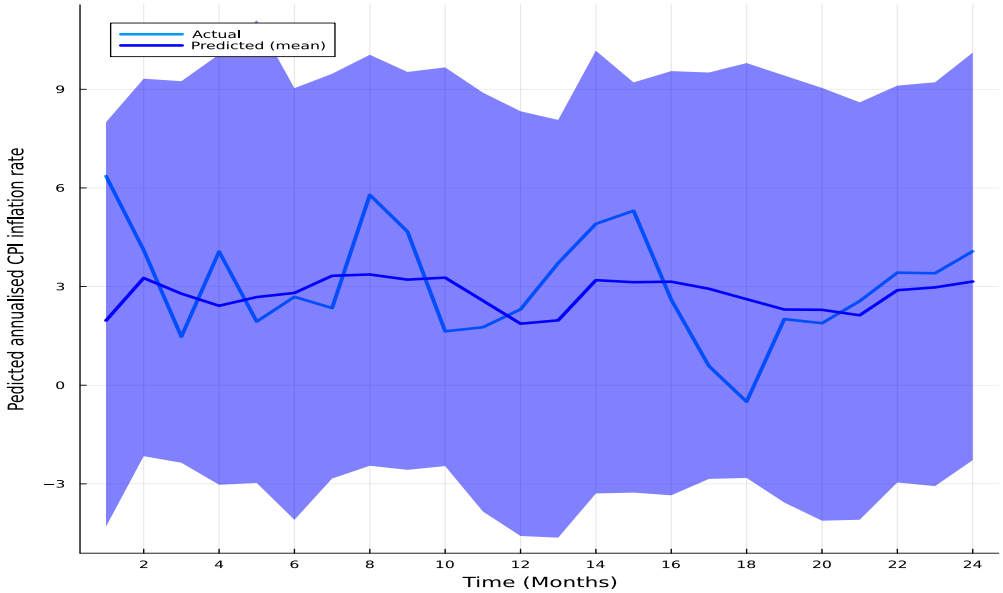



**Figure 4.** Predictions of annualized CPI inflation rates: The blue line represents actual values, the dark blue line represents the predictive mean, and the light blue shaded region indicates the 95% credible interval.

**Table 10.** Convergence diagnostics for the ARX(3; 0, 0) model fitted to monthly US CPI inflation under SMN-$t$ disturbances.

| | 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 (%) |
| $\theta_0$ | 968 | 936 | 1.033 | 0.203 | 0.839 | −0.005 | 0.012 | −0.005 | −0.004 | 1010 | 0.0060 | 3.15 |
| $\phi_1$ | 892 | 936 | 0.953 | 1.118 | 0.264 | −0.014 | −0.067 | −0.009 | 0.007 | 934 | 0.0014 | 3.27 |
| $\phi_2$ | 892 | 936 | 0.953 | −0.589 | 0.556 | −0.100 | 0.006 | −0.032 | −0.084 | 1251 | 0.0013 | 2.83 |
| $\phi_3$ | 1051 | 936 | 1.123 | −0.304 | 0.761 | −0.008 | 0.005 | −0.030 | −0.017 | 1014 | 0.0012 | 3.14 |
| $\psi_{10}$ | 968 | 936 | 1.033 | −0.844 | 0.398 | −0.013 | −0.057 | −0.013 | 0.022 | 1027 | 0.0001 | 3.12 |
| $\psi_{20}$ | 968 | 936 | 1.033 | 0.531 | 0.596 | 0.008 | −0.008 | −0.034 | −0.031 | 927 | 0.0004 | 3.28 |
| $\tau^2$ | 968 | 936 | 1.033 | 0.026 | 0.979 | −0.010 | 0.001 | −0.018 | 0.035 | 927 | 0.0141 | 3.29 |
| $\eta$ | 968 | 936 | 1.033 | −1.292 | 0.196 | 0.034 | −0.018 | 0.023 | 0.056 | 698 | 0.0225 | 3.79 |

Consistent with the derivation of Section 5.3, the reported forecasts are *conditional* on the future paths of the exogenous variables. The hold-out values of $x_{1,t}$ and $x_{2,t}$ over the 24-month window are treated as known at the forecast origin, so the exercise should be interpreted as a conditional ex-post forecast evaluation rather than a genuine real-time forecasting experiment. To quantify the cost of this assumption, we repeated the exercise under an unconditional information set in which no future exogenous information is used. At every MCMC iteration, the future paths are simulated from AR(1) models fitted to the historical exogenous series. Table 11 reports both variants alongside the OLS benchmark of Table 7. The unconditional SMN-$t$ forecasts suffer only a modest degradation (RMSE 1.801 versus 1.664) and remain well calibrated (24/24 coverage), so the qualitative conclusions are robust to the information set. Mirroring the simulation study, the point forecasts of all three procedures are of comparable accuracy; the distinctive contribution of the heavy-tailed Bayesian treatment lies in fit and calibrated predictive uncertainty rather than in point accuracy.

**Table 11.** Forecast evaluation under alternative information sets and benchmarks (24-month hold-out). Bias = actual − predictive mean; coverage is over 24 binary outcomes.

| Procedure | Bias | MAE | RMSE | Inside 95% PI |
|---|---|---|---|---|
| ARX-SMN-*t*, conditional exogenous | 0.349 | 1.318 | 1.664 | 24/24 |
| ARX-SMN-*t*, unconditional exogenous | 0.466 | 1.382 | 1.801 | 24/24 |
| OLS-ARX (Gaussian), conditional | 0.284 | 1.317 | 1.683 | 24/24 |

As a further diagnostic, Engle's ARCH-LM test with four lags applied to the posterior-mean residuals strongly rejects conditional homoskedasticity (LM = 58.8, $p < 0.001$). This finding is informative in two respects. First, it reinforces the inadequacy of the iid Gaussian benchmark. Second, it clarifies the scope of the SMN framework. The scale-mixture specification captures *unconditional* heavy tails through observation-specific precision weights, not the temporal clustering of volatility. A joint SMN–GARCH or stochastic-volatility extension is the natural next step. The error-law comparisons reported below remain valid as comparative statements among the fitted distributions.

To assess the gains delivered by the heavy-tailed SMN-*t* specification over the Gaussian ARX benchmark (SMN-*n*) and other SMN variants, we compute the deviance information criterion (DIC) [45], as reported in Table 12. The deviance is based on the marginal likelihood with the latent weights integrated out analytically, and $\bar{D}$ averages the deviance over the 1000 retained posterior draws. The effective number of parameters is $p_D = \bar{D} - D(\bar{\boldsymbol{\theta}})$, evaluated at the posterior means, and DIC $= \bar{D} + p_D$. The resulting $p_D$ values (7.0–8.4) closely match the nominal parameter counts (7–9), indicating a well-behaved complexity penalty. Forecasting accuracy is evaluated using the Bias measure together with the mean absolute error (MAE) and RMSE, where smaller values indicate superior predictive accuracy. As shown in Table 12, the SMN-*t* model attains the smallest DIC, whereas the SMN-sl model yields the smallest |Bias|, MAE, and RMSE. The differences among the three heavy-tailed SMN specifications are relatively small ($\Delta$DIC $\leq 3.4$; MAE within 0.03), while all three improve substantially over the Gaussian benchmark ($\Delta$DIC = 87.0, far exceeding the conventional threshold of 10 [45]). The Gaussian specification performs worst in fit and RMSE, confirming the importance of modeling heavy-tailed disturbances in CPI inflation. Within the SMN family, no single heavy-tailed law is decisively favored, in line with empirical studies documenting heavy-tailed macroeconomic shocks [46].

**Table 12.** DIC and forecasting metrics for the ARX$(3; 0, 0)$ CPI inflation model under different error distributions. $\Delta$DIC is relative to SMN-*t*; Bias = actual − predictive mean.

| Error Distribution | DIC | $\Delta$DIC | $p_D$ | Bias | MAE | RMSE |
|---|---|---|---|---|---|---|
| SMN-*t* | 2075.5 | — | 8.03 | 0.349 | 1.318 | 1.664 |
| SMN-sl | 2077.0 | 1.5 | 7.47 | 0.282 | 1.292 | 1.647 |
| SMN-cn | 2078.9 | 3.4 | 8.43 | 0.385 | 1.324 | 1.684 |
| SMN-*n* | 2162.5 | 87.0 | 7.03 | 0.360 | 1.340 | 1.703 |

To assess the sensitivity of posterior inference and forecasting performance to prior specification, the SMN-*t* ARX(3;0,0) model was refitted under four alternative priors: Jeffreys (P3), diffuse normal-gamma (P1; $\boldsymbol{m}_0 = \mathbf{0}$, $\boldsymbol{S}_0 = 100\boldsymbol{I}$), Minnesota-style (P1; first AR lag centered at 1 with prior variance

0.04, remaining coefficients shrunk to zero), and Zellner $g$-prior (P2; $g = 1/n$, $\bar{\boldsymbol{\theta}} = \mathbf{0}$). Table 13 shows that the posterior and forecasts are essentially invariant to the prior on $(\boldsymbol{\theta}, \tau^2)$. Across the four specifications P1–P3, $\hat{\phi}_1$ ranges over $[0.448, 0.456]$, $\hat{\psi}_{10}$ over $[0.0312, 0.0324]$, and $\hat{\eta}$ over $[3.18, 3.28]$, with a DIC spread of 1.4 points and an RMSE spread of 0.05. With $n_{\text{obs}} = 441$ and $q = 6$ the likelihood dominates, so the Jeffreys default is inconsequential for the conclusions, while informative or shrinkage-type priors of the P1 family become relevant for short samples or heavily parameterized designs (Section 4).

**Table 13.** Prior sensitivity of the SMN-$t$ ARX(3; 0, 0) posterior and forecasts.

| Prior | DIC | RMSE | $\hat{\phi}_1$ | $\hat{\psi}_{10}$ | $\hat{\eta}$ |
|---|---|---|---|---|---|
| P3: Jeffreys | 2075.9 | 1.688 | 0.456 | 0.0312 | 3.280 |
| P1: diffuse | 2074.8 | 1.642 | 0.456 | 0.0317 | 3.182 |
| P1: Minnesota-style | 2074.7 | 1.661 | 0.448 | 0.0324 | 3.186 |
| P2: $g$-prior $(1/n)$ | 2074.5 | 1.672 | 0.453 | 0.0316 | 3.252 |

## 8. Concluding remarks

This paper has developed an MCMC-based multi-step forecasting framework for ARX models under SMN disturbances. The contribution is an incremental but self-contained extension of the Bayesian SMN-error machinery established for AR and seasonal AR models [22, 23] to the ARX case. The global linearity of the ARX model in the coefficient vector, combined with the Gaussian complete-data structure restored by conditioning on the latent precision weights, enables closed-form full conditional distributions throughout under three prior families, conjugate normal-gamma (including Minnesota- and Litterman-type shrinkage configurations), Zellner's $g$-prior, and Jeffreys' prior. The posterior of the coefficient vector is multivariate normal and the posterior of the scale parameter is inverse-gamma. Crucially, the conditional predictive distribution of the $K$-step forecast vector is multivariate normal with a mean that decomposes cleanly into an autoregressive recursion component and a deterministic exogenous contribution. For each of the three non-Gaussian SMN subfamilies (Student's $t$, slash, and contaminated normal), the conditional distributions of the latent mixing weights and shape parameters were derived, with $w$-marginalized out for the shape parameters. The resulting hybrid Gibbs–Metropolis–Hastings sampler simultaneously outputs posterior parameter draws and multi-step predictive draws from a single integrated chain.

Monte Carlo validation across three ARX configurations and all non-Gaussian SMN subfamilies confirmed the accuracy of posterior means, near-nominal empirical coverage of 95% credible and predictive intervals, and small forecast RMSE values at all prediction horizons. A paired comparison with an OLS-ARX Gaussian benchmark on the same simulated datasets showed that point-forecast accuracy is essentially equivalent across methods, while the Gaussian intervals systematically under-cover with increasing horizon and the SMN intervals remain near nominal. Thus, the benefit of the heavy-tailed framework lies in calibrated predictive uncertainty rather than sharper point predictions.

The macroeconomic application to monthly US CPI inflation forecasting recovered interpretable macroeconomic coefficients (read as conditional associations rather than causal effects), yielded strong evidence for heavy-tailed disturbances, and produced accurate and well-calibrated predictive intervals. The exercise was framed as a conditional ex-post forecast evaluation, with future exogenous paths

treated as known at the forecast origin. An unconditional-information-set variant, in which the future paths are simulated within the sampler, degraded RMSE only modestly and preserved coverage, indicating robustness to the information set. The posterior and forecasts proved essentially invariant to the prior across four specifications spanning Jeffreys', diffuse and Minnesota-style normal-gamma, and Zellner's $g$-prior. ARCH-LM testing revealed significant conditional heteroskedasticity in the residuals, reinforcing the rejection of the iid Gaussian benchmark while delimiting the scope of the present framework. The SMN class captures unconditional heavy tails through observation-specific precision weights, not volatility clustering.

Several extensions follow naturally. A joint SMN–GARCH or stochastic-volatility specification would capture volatility clustering alongside unconditional heavy tails. A rolling-origin or repeated forecast evaluation, in which parameters are re-estimated as the origin advances and no future information is used, would enable a formal assessment of empirical coverage in real time. Hyper-$g$ priors would address the known limitations of fixed $g$; and state-space representations or locally adaptive shrinkage priors would allow the ARX parameter vector to change over time while remaining robust to non-Gaussian disturbances.

## Author contributions

Ayman A. Amin: Conceptualization, Writing–original draft, Methodology, Software, Validation; Shuhrah A. Alghamdi: Validation, Writing–review & editing, Project administration, Funding acquisition. All authors have read and approved the final version of the manuscript for publication.

## Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## 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 Consortium Funding Program grant code (NU/CPL/SERC/14/4440-7). Moreover, the authors sincerely thank the handling editor and the anonymous reviewers for their constructive feedback, which significantly enhanced the clarity and quality of this manuscript.

## Conflict of interest

The authors declare no conflicts of interest.

## References


1. L. Ljung, *System identification: theory for the User*, Englewood Cliffs, NJ: Prentice-Hall, 1987.
2. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time series analysis: forecasting and control*, 5 Eds., John Wiley & Sons, 2015.
3. R. S. Tsay, *Analysis of financial time series*, 3 Eds., Hoboken, NJ,: John Wiley & Sons, 2010. http://doi.org/10.1002/9780470644560
4. J. H. Stock, M. W. Watson, Why has U.S. inflation become harder to forecast?, *J. Money Credit Bank.*, **39** (2007), 3–33. http://doi.org/10.1111/j.1538-4616.2007.00014.x
5. R. J. Hyndman, G. Athanasopoulos, *Forecasting: principles and practice*, 3 Eds., Melbourne, Australia: OTexts, 2021.
6. G. E. P. Box, G. Tiao, *Bayesian inference in statistical analysis*, Reading, MA: Addison-Wesley, 1973.
7. L. D. Broemeling, *Bayesian analysis of linear models*, New York: Marcel Dekker, 1985.
8. L. Broemeling, M. Land, On forecasting with univariate autoregressive processes: a Bayesian approach, *Commun. Stat.-Theor. Meth.*, **13** (1984), 1305–1320. http://doi.org/10.1080/03610928408828758
9. S. Chib, E. Greenberg, Bayes inference in regression models with arma (p, q) errors, *J. Econometrics*, **64** (1994), 183–206. http://doi.org/10.1016/0304-4076(94)90063-9
10. R. E. McCulloch, R. S. Tsay, Bayesian analysis of autoregressive time series via the Gibbs sampler, *J. Time Ser. Anal.*, **15** (1994), 235–250. http://doi.org/10.1111/j.1467-9892.1994.tb00188.x
11. M. A. Tanner, W. H. Wong, The calculation of posterior distributions by data augmentation, *J. Amer. Stat. Assoc.*, **82** (1987), 528–540. http://doi.org/10.1080/01621459.1987.10478458
12. A. E. Gelfand, A. F. Smith, Sampling-based approaches to calculating marginal densities, *J. Amer. Stat. Assoc.*, **85** (1990), 398–409. http://doi.org/10.1080/01621459.1990.10476213
13. J. Marriott, N. Ravishanker, A. E. Gelfand, Bayesian analysis of arma processes: complete sampling-based inferences under full likelihood, Department of Statistics, Stanford University, Technical Report No.471, 1993.
14. G. Barnett, R. Kohn, S. Sheather, Bayesian estimation of an autoregressive model using Markov chain Monte Carlo, *J. Econometrics*, **74** (1996), 237–254. http://doi.org/10.1016/0304-4076(95)01744-5
15. A. Zellner, Bayesian and non-Bayesian analysis of the regression model with multivariate student-*t* error terms, *J. Amer. Stat. Assoc.*, **71** (1976), 400–405. http://doi.org/10.1080/01621459.1976.10480357
16. A. Harvey, A. Luati, Filtering with heavy tails, *J. Amer. Stat. Assoc.*, **109** (2014), 1112–1122. http://doi.org/10.1080/01621459.2014.887011
17. K. L. Lange, R. J. Little, J. M. G. Taylor, Robust statistical modeling using the *t* distribution, *J. Amer. Stat. Assoc.*, **84** (1989), 881–896. http://doi.org/10.1080/01621459.1989.10478852

18. D. F. Andrews, C. L. Mallows, Scale mixtures of normal distributions, *J. Royal Stat. Soc. B*, **36** (1974), 99–102. http://doi.org/10.1111/j.2517-6161.1974.tb00989.x
19. M. West, On scale mixtures of normal distributions, *Biometrika*, **74** (1987), 646–648. http://doi.org/10.1093/biomet/74.3.646
20. G. Ferreira, L. M. Castro, V. H. Lachos, R. Dias, Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors, *J. Appl. Stat.*, **40** (2013), 1796–1816. http://doi.org/10.1080/02664763.2013.796349
21. M. Maleki, A. R. Nematollahi, Autoregressive models with mixture of scale mixtures of Gaussian innovations, *Iran. J. Sci. Technol. Trans. Sci.*, **41** (2017), 1099–1107. http://doi.org/10.1007/s40995-017-0237-6
22. A. A. Amin, Bayesian modeling and forecasting of seasonal autoregressive models with scale-mixtures of normal errors, *Comput. Stat.*, **40** (2025), 3453–3475. http://doi.org/10.1007/s00180-025-01617-2
23. A. A. Amin, Full Bayesian analysis of seasonal autoregressive models under scale-mixtures of normal errors, *J. Stat. Comput. Sim.*, **96** (2026), 184–204. http://doi.org/10.1080/00949655.2025.2556269
24. V. Peterka, Bayesian system identification, *Automatica*, **17** (1981), 41–53. http://doi.org/10.1016/0005-1098(81)90083-2
25. G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, L. Ljung, Kernel methods in system identification, machine learning and function estimation: a survey, *Automatica*, **50** (2014), 657–682. http://doi.org/10.1016/j.automatica.2014.01.001
26. A. Cooper, D. Simpson, L. Kennedy, C. Forbes, A. Vehtari, Cross-validatory model selection for Bayesian autoregressions with exogenous regressors, *Bayesian Anal.*, **20** (2025), 573–597. http://doi.org/10.1214/23-BA1409
27. A. A. Amin, A Kullback-Leibler divergence based comparison of approximate Bayesian estimations of arma models, *Commun. Stat. Appl. Meth.*, **29** (2022), 471–486. http://doi.org/10.29220/CSAM.2022.29.4.471
28. A. A. Amin, Kullback-Leibler divergence to evaluate posterior sensitivity to different priors for autoregressive time series models, *Commun. Stat.-Simul. Comput.*, **48** (2019), 1277–1291. http://doi.org/10.1080/03610918.2017.1410709
29. M. A. Ismail, A. A. Amin, Gibbs sampling for SARMA models, *Pakistan Journal of Statistics*, **30** (2014), 153–168.
30. A. A. Amin, Gibbs sampling for Bayesian prediction of sarma processes, *Pak. J. Stat. Oper. Res.*, **15** (2019), 397–418. http://doi.org/10.18187/pjsor.v15i2.2174
31. A. A. Amin, Full Bayesian analysis of double seasonal autoregressive models with real applications, *J. Appl. Stat.*, **51** (2024), 1524–1544. http://doi.org/10.1080/02664763.2023.2211754
32. G. Pillonetto, L. Ljung, Full Bayesian identification of linear dynamic systems using stable kernels, *Proceedings of the National Academy of Sciences*, **120** (2023), e2218197120. http://doi.org/10.1073/pnas.2218197120

33. X. Chen, S. Zhao, F. Liu, Robust identification of linear ARX models with recursive EM algorithm based on student's *t*-distribution, *J. Franklin I.*, **358** (2021), 1103–1121. http://doi.org/10.1016/j.jfranklin.2020.06.003
34. X. Chen, S. Zhao, F. Liu, C. Tao, Laplace distribution based online identification of linear systems with robust recursive expectation–maximization algorithm, *IEEE Trans. Ind. Inform.*, **19** (2023), 9028–9036. http://doi.org/10.1109/TII.2022.3225026
35. X. Chen, J. Zhang, Parameter estimation of a linear ARX model with skewed heavy-tailed noise, *IEEE Access*, **14** (2026), 13582–13596. http://doi.org/10.1109/ACCESS.2026.3653893
36. C. Fernandez, E. Ley, M. F. J. Steel, Benchmark priors for Bayesian model averaging, *J. Econometrics*, **100** (2001), 381–427. http://doi.org/10.1016/S0304-4076(00)00076-2
37. G. Koop, *Bayesian econometrics*, Chichester, UK: John Wiley & Sons, 2003.
38. R. B. Litterman, Forecasting with Bayesian vector autoregressions—five years of experience, *J. Bus. Econ. Stat.*, **4** (1986), 25–38. http://doi.org/10.1080/07350015.1986.10509491
39. F. Liang, R. Paulo, G. Molina, M. A. Clyde, J. O. Berger, Mixtures of g priors for Bayesian variable selection, *J. Amer. Stat. Assoc.*, **103** (2008), 410–423. http://doi.org/10.1198/016214507000001337
40. C. R. B. Cabral, V. H. Lachos, M. R. Madruga, Bayesian analysis of skew-normal independent linear mixed models with heterogeneity in the random-effects population, *J. Stat. Plan. Infer.*, **142** (2012), 181–200. http://doi.org/10.1016/j.jspi.2011.07.007
41. J. Geweke, Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments, In: *Bayesian statistics 4: Proceedings of the fourth Valencia international meeting, dedicated to the memory of Morris H. DeGroot, 1931–1989*, Oxford University Press, 1992, 169-193. http://doi.org/10.1093/oso/9780198522669.003.0010
42. A. E. Raftery, S. M. Lewis, The number of iterations, convergence diagnostics and generic metropolis algorithms, In: *Practical Markov Chain Monte Carlo*, London, U.K.: Chapman and Hall, 1995, 763–773.
43. J. P. LeSage, *Applied econometrics using MATLAB*, Department of Economics, University of Toledo, 1999.
44. J. H. Stock, M. W. Watson, Modeling inflation after the crisis, In: *NBER Working paper series*, National Bureau of Economic Research, Working paper 16488. http://doi.org/10.3386/w16488
45. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit, *J. Royal Stat. Soc. B*, **64** (2002), 583–639. http://doi.org/10.1111/1467-9868.00353
46. V. Curdia, A. Ferrero, G. C. Ng, A. Tambalotti, Has U.S. monetary policy tracked the efficient interest rate?, *J. Monetary Econ.*, **70** (2015), 72–83. http://doi.org/10.1016/j.jmoneco.2014.09.004




AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)