AIMS Mathematics, 11(11): 39102–39147.
DOI: 10.3934/math.20261506
Received: 07 August 2026
Revised: 19 October 2026
Accepted: 02 November 2026
Published: 24 November 2026

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

Research article

Exogenous information in the precision submodel: A dual-covariate βARMA framework for bounded time series

Elsidieg I. Belhaj[1,2] **and Majid Khan Majahar Ali**[2,*]

[1] Department of Statistics, Misurata University, Libya
[2] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

* **Correspondence:** Email: majidkhanmajaharali@usm.my.

Abstract: Time series data constrained within the unit interval are commonly modeled using the beta autoregressive moving average (βARMA) class to accommodate skewness and heteroscedasticity. Existing dynamic precision extensions allow the precision parameter ϕ_t to vary only through its own lags, a lagged variance proxy, or seasonal terms, not through exogenous covariates. This paper proposes a dual-exogenous βARMA framework where independent covariate vectors enter both submodels through a recursive, observation-driven structure. We derive the conditional likelihood, a recursive score function, and the Fisher information matrix, and establish the asymptotic normality of the conditional maximum likelihood estimator, which yields Wald and likelihood ratio tests for the precision covariates. We show that the mean and precision parameter blocks are generally not orthogonal, and that misspecification of the precision submodel can make the mean submodel estimator inconsistent when it induces a nonzero bias in the mean score equations. An extensive Monte Carlo study confirms the finite-sample validity and directly quantifies the theoretical bias transmitted to the mean submodel coefficients under misspecification. Applications to the Malaysian unemployment rate and the US capacity utilization rate show that with the mean submodel held fixed, exogenous precision covariates substantially improve the fit relative to constant precision and lagged variance proxy specifications. Forecasting performance is evaluated with both realized and forecast covariates.

Keywords: Beta regression; time-varying precision; βARMA model; bounded time series; exogenous covariates; point and probabilistic forecasting
Mathematics Subject Classification: 62M10, 62F12, 62P20

1. Introduction

Accurate modeling and forecasting time series data that fall within the standard unit interval $(0, 1)$, such as rates or proportions, remains a practical challenge in applied research. Bounded variables such

as unemployment rates, mortality indices, and humidity levels, frequently occur across economics, epidemiology, and environmental science [1, 2]. For decades, such series were analyzed with the autoregressive integrated moving average (ARIMA) framework [3], whose Gaussian assumptions are poorly suited to bounded data, as rates and proportions typically display skewness, non-constant variance, and clustering near the boundaries that a symmetric normal law cannot capture [4, 5].

Early approaches addressed the boundary constraint with a logit transformation followed by standard Gaussian estimation, at the cost of interpretability in the original scale and of the heteroscedasticity inherent to beta-distributed data [5]. Ferrari and Cribari-Neto [5] instead reparameterized the beta distribution in terms of a conditional mean (μ) and a precision parameter (ϕ), and this framework was extended to time-dependent data through the βARMA model [4], which models the conditional mean through a link function while keeping the parameters interpretable.

Subsequent work extended the βARMA class to richer dynamics, including seasonal effects (βSARMA) [1] and improved the diagnostic and inferential tools [6, 7]. A further development concerns the precision parameter itself. Scher et al. [2] showed that letting ϕ_t evolve through an autoregressive mechanism or a lagged variance proxy improves both the point and probabilistic forecast accuracy, since the dispersion of a bounded process is rarely static.

However, current dynamic precision βARMA models fail to incorporate external predictors directly into the precision equation. In every existing formulation, the information used to drive ϕ_t is *internal* to the process itself: Its own lagged values, a deterministic seasonal indicator, or a lagged transformation of the response [1, 2]. No existing specification allows the conditional precision to respond to a genuinely exogenous covariate process $\boldsymbol{z}_t$ that is distinct from, and potentially uncorrelated with, the covariate vector $\boldsymbol{x}_t$ that drives the conditional mean. This omission has significant theoretical and practical consequences. Many bounded series of practical interest are subject to dispersion-inducing events whose timing and magnitude are observable to the analyst but are not captured by the series' own lagged dynamics. A national unemployment rate, for instance, may be driven in its *level* by the labor-market's fundamentals such as vacancy rates or wage growth, while its *volatility* in a given month is more directly related to a policy intervention, which leaves a footprint in auxiliary administrative data before it is reflected in the lagged variance of the series itself. Omitting relevant information misspecifies the precision submodel and may lead to poorly calibrated predictive distributions. More importantly, because the mean and precision blocks need not be information-orthogonal, such misspecification can also affect mean submodel estimation. As shown in Section 3, this occurs when the imposed precision restriction induces a nonzero bias in the population mean score equations. This paper makes three main contributions, developing new analytical tools that extend beyond the existing constant, seasonal, or autoregressive precision models. First, allowing $\boldsymbol{z}_t$ to enter the precision link breaks the recursive independence between the mean and precision score equations that autoregressive or seasonal precision models retain, so the score function, the recursive derivatives that make it computable in $O(n)$ time, and the Fisher information matrix all have to be rederived from first principles for the dual-exogenous case; we give this derivation in full, including the explicit recursive system for the mean submodel score and the analytic matrix form of the score used in estimation (Section 3). Second, and most consequentially, we show in Section 3 that the mean and precision blocks need not be information-orthogonal and we characterize sufficient conditions for non orthogonality. We further show that the precision submodel misspecification can render the mean submodel estimator inconsistent when it induces a nonzero mean score bias. Third, we develop

inferential machinery, namely Wald and likelihood ratio (LR) tests targeted specifically at $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ (Section 4), and an eight-experiment Monte Carlo design (Section 7) that isolates and quantifies this bias transmission channel directly, rather than only documenting an aggregate improvement in fit. These findings suggest that treating the precision submodel as a dynamic component driven by exogenous covariates is practically significant, rather than just a theoretical extension.

This paper proposes such a dual-exogenous βARMA architecture, in which external predictors enter both submodels independently within a recursive time-series structure. The conditional mean retains ARMA memory while the precision evolves jointly with its own exogenous covariates. Unlike existing dynamic precision specifications that rely only on lagged variance proxies [2] or seasonal terms [1], this approach introduces external predictors to both submodels. We illustrate the framework on two monthly bounded time series with contrasting dynamics (a rate and a utilization index) each paired with a small set of relevant exogenous covariates.

The proposed model consistently recovers the true structural configuration across submodels, with the recovery rates approaching one as the sample size grows, while constant precision benchmarks exhibit persistent estimation bias, confirming that neglecting exogenous precision drivers leads to misspecification and biased inference. Section 2 presents the model specification, while Section 3 details the estimation theory, analytic score equations, and proof of non-orthogonality. We cover hypothesis testing and confidence intervals in Section 4, model selection in Section 5, and our forecasting framework, including the Diebold–Mariano testing protocol, in Section 6. Finally, Sections 7 and 8 present the Monte Carlo simulations and empirical applications, before Section 9 offers concluding remarks.

2. Model specification

We consider a stochastic process $\{y_t\}_{t=1}^n$ where each observation is defined on the standard unit interval $(0, 1)$. Following the reparameterization proposed by [5], we assume that y_t is conditionally beta-distributed given the filtration $\mathcal{F}_{t-1}$, which represents the information set containing past realizations of the process and current and past values of exogenous covariates. The conditional density function is expressed as:

$$f(y_t \mid \mathcal{F}_{t-1}; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\,\Gamma((1-\mu_t)\phi_t)}\, y_t^{\mu_t\phi_t - 1}\,(1-y_t)^{(1-\mu_t)\phi_t - 1}, \tag{2.1}$$

where $0 < \mu_t < 1$ denotes the conditional mean and $\phi_t > 0$ serves as the precision parameter. Expressing the conditional variance as $\mathrm{Var}(y_t \mid \mathcal{F}_{t-1}) = \mu_t(1-\mu_t)/(1+\phi_t)$, clearly illustrates ϕ_t's role as inverse volatility: Higher precision concentrates probability mass around the mean, whereas lower precision allows greater dispersion in the tails [5, 7].

We propose a framework where dedicated exogenous covariates drive the simultaneous evolution of the conditional mean and precision. Specifically, the conditional mean μ_t is mapped to a real-valued linear predictor η_t via a strictly monotonic, twice-differentiable link function $g_1 : (0, 1) \to \mathbb{R}$, typically the logit link. The βARMA(p, q) structure with exogenous covariates is defined as follows [2, 4]:

$$g_1(\mu_t) = \eta_t = \alpha_1 + \boldsymbol{x}_t^\top\boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i[g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top\boldsymbol{\beta}] + \sum_{j=1}^{q} \theta_j r_{t-j}, \tag{2.2}$$

where $\boldsymbol{x}_t \in \mathbb{R}^k$ is a vector of exogenous covariates, $\boldsymbol{\beta}$ is the corresponding coefficient vector, and $r_t = g_1(y_t) - \eta_t$ denotes the innovation on the transformed scale. By centering the autoregressive terms as $[g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top \boldsymbol{\beta}]$, the model captures serial dependence relative to the exogenous structural trend, a refinement that maintains mathematical consistency with the generalized βARMA class [2, 4].

Parallel to the mean dynamics, we introduce a variable precision submodel to account for time-varying heteroscedasticity. Unlike standard βARMA formulations that assume a static ϕ, we relate the precision parameter to a second set of exogenous covariates through a link function $g_2 : (0, \infty) \to \mathbb{R}$. The necessity of this extension is empirically demonstrated in our simulation study (Section 7), which shows that models restricting ϕ_t to a constant fail to recover the true precision parameters and produce severely inflated mean-squared errors (MSE) (e.g., the MSE for $\boldsymbol{\gamma}$ remains an order of magnitude larger than in the proposed model). By allowing exogenous covariates to drive ϕ_t, the proposed framework correctly captures the dispersion dynamics of the conditional distribution, thereby mitigating the bias that arises from a misspecified precision structure. The precision linear predictor ζ_t is specified as:

$$g_2(\phi_t) = \zeta_t = \alpha_2 + \boldsymbol{z}_t^\top \boldsymbol{\gamma}, \tag{2.3}$$

where $\boldsymbol{z}_t \in \mathbb{R}^l$ is a vector of covariates that govern the conditional dispersion of the response variable, and $\boldsymbol{\gamma}$ is the associated parameter vector. This specification allows the model to differentiate between factors that affect the level of the response variable from those driving its conditional volatility. Under this unified framework, the joint parameter vector $\boldsymbol{\eta} = (\alpha_1, \alpha_2, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top, \boldsymbol{\gamma}^\top)^\top$ is estimated by conditional maximum likelihood. The log-likelihood function for a sample of n observations, conditioned on the first $m = \max(p, q)$ realizations, is:

$$\begin{aligned} \ell(\boldsymbol{\eta}) = \sum_{t=m+1}^{n} \Big\{ & \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) \\ & + (\mu_t\phi_t - 1)\log y_t + ((1-\mu_t)\phi_t - 1)\log(1-y_t) \Big\}. \end{aligned} \tag{2.4}$$

Optimization of $\ell(\boldsymbol{\eta})$ is achieved through quasi-Newton algorithms (L-BFGS-B) [8]. To establish confidence intervals and test hypotheses for the precision covariates, we derive the necessary asymptotic properties using the conditional score function and Fisher information matrix.

3. Inference and asymptotic theory

The estimation of the joint parameter vector $\boldsymbol{\eta} = (\boldsymbol{\eta}_\mu^\top, \boldsymbol{\eta}_\phi^\top)^\top$ is conducted via conditional maximum likelihood. Given the recursive structure of Eq (2.2), the log-likelihood cannot be evaluated for the initial $m = \max(p, q)$ observations. Consequently, we condition on the first m realizations, set the initial innovations $r_{t-j} = 0$ for $j \geq t$, and initialize the lagged transformed values $g_1(y_{t-i})$ at the sample mean of the transformed data, as suggested by [9].

Maximization of the conditional log-likelihood relies on the digamma function, $\psi(x) = \frac{d}{dx}\log\Gamma(x)$, and the trigamma function, $\psi'(x) = \frac{d^2}{dx^2}\log\Gamma(x)$, which arise naturally when differentiating $\ell(\boldsymbol{\eta})$ with respect to the mean and precision parameters. The conditional score vector $U(\boldsymbol{\eta}) = \partial\ell(\boldsymbol{\eta})/\partial\boldsymbol{\eta}$ is partitioned as $(U_{\boldsymbol{\eta}_\mu}(\boldsymbol{\eta})^\top, U_{\boldsymbol{\eta}_\phi}(\boldsymbol{\eta})^\top)^\top$. Defining $y_t^* = \log\{y_t/(1-y_t)\}$ and $y_t^\dagger = \log(1-y_t)$, with the conditional expectations $\mu_t^* = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$ and $\mu_t^\dagger = \psi((1-\mu_t)\phi_t) - \psi(\phi_t)$, the score components

are:

$$U_{\boldsymbol{\eta}_\mu}(\boldsymbol{\eta}) = \sum_{t=m+1}^{n} \phi_t \frac{d\mu_t}{d\eta_{1,t}}(y_t^* - \mu_t^*)\boldsymbol{\chi}_t, \tag{3.1}$$

$$U_{\boldsymbol{\eta}_\phi}(\boldsymbol{\eta}) = \sum_{t=m+1}^{n} \frac{d\phi_t}{d\eta_{2,t}}[\mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)]\,\boldsymbol{\zeta}_t, \tag{3.2}$$

where $\boldsymbol{\chi}_t = \partial\eta_{1,t}/\partial\boldsymbol{\eta}_\mu$ is computed recursively due to the βARMA dynamics (its explicit recursive form is derived in Section 3.1 below), and $\boldsymbol{\zeta}_t = (1, \boldsymbol{z}_t^\top)^\top$, which is available in closed form since the precision predictor (2.3) contains no recursive terms.

3.1. Recursive computation of the mean submodel score

Because $\eta_{1,t}$ in Eq (2.2) depends on its own lagged values through the moving-average term, $\boldsymbol{\chi}_t$ cannot be written as a closed-form function of the data alone: It must itself be propagated recursively, alongside $\eta_{1,t}$, r_t, μ_t, and ϕ_t, in a single forward pass through the sample. Differentiating Equation (2.2) with respect to each element of $\boldsymbol{\eta}_\mu = (\alpha_1, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top)^\top$ in turn, and using $\partial r_{t-j}/\partial\boldsymbol{\eta}_\mu = -\partial\eta_{1,t-j}/\partial\boldsymbol{\eta}_\mu$ (since $r_{t-j} = g_1(y_{t-j}) - \eta_{1,t-j}$ and $g_1(y_{t-j})$ is data, not a function of $\boldsymbol{\eta}_\mu$), yields the following linear, order-q difference equations, in which each derivative at time t is expressed in terms of the corresponding derivative at times $t-1, \ldots, t-q$:

$$\frac{\partial\eta_{1,t}}{\partial\alpha_1} = 1 - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{1,t-j}}{\partial\alpha_1}, \tag{3.3}$$

$$\frac{\partial\eta_{1,t}}{\partial\beta_l} = x_{l,t} - \sum_{i=1}^{p} \varphi_i\, x_{l,t-i} - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{1,t-j}}{\partial\beta_l}, \qquad l = 1, \ldots, k, \tag{3.4}$$

$$\frac{\partial\eta_{1,t}}{\partial\varphi_{i_0}} = \tau_{t-i_0} - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{1,t-j}}{\partial\varphi_{i_0}}, \qquad i_0 = 1, \ldots, p, \tag{3.5}$$

$$\frac{\partial\eta_{1,t}}{\partial\theta_{j_0}} = r_{t-j_0} - \sum_{j=1}^{q} \theta_j \frac{\partial\eta_{1,t-j}}{\partial\theta_{j_0}}, \qquad j_0 = 1, \ldots, q, \tag{3.6}$$

where $\tau_{t-i} := g_1(y_{t-i}) - \boldsymbol{x}_{t-i}^\top\boldsymbol{\beta}$ is the exogenously detrended, link-scale lagged value entering the autoregressive part of Eq (2.2). Equation (3.5) isolates the single term τ_{t-i_0} because $\partial/\partial\varphi_{i_0}$ of $\sum_i \varphi_i\tau_{t-i}$ selects only the $i = i_0$ summand, while Eq (3.6) isolates r_{t-j_0} directly (an "own" effect) in addition to the indirect effect propagated through $\partial r_{t-j}/\partial\theta_{j_0} = -\partial\eta_{1,t-j}/\partial\theta_{j_0}$ for every $j = 1, \ldots, q$. Each of the four recursions (3.3)–(3.6) is driven by an exogenous forcing term, which are, respectively, the constant 1, the detrended covariate $x_{l,t} - \sum_i \varphi_i x_{l,t-i}$, the detrended lag τ_{t-i_0}, and the innovation r_{t-j_0}. Each recursion is initialized, consistently with the conditioning convention above, at $\partial\eta_{1,t}/\partial\boldsymbol{\eta}_\mu := \mathbf{0}$ for $t \le m$. Because the moving-average polynomial $\theta(u)$ is invertible under the causal invertibility conditions maintained throughout, each difference equation is stable, so $\boldsymbol{\chi}_t$ is computed forward in time in a single $O(n)$ pass – the recursion exploited directly by the L-BFGS-B maximization of $\ell(\boldsymbol{\eta})$. Setting $\boldsymbol{\beta} \equiv \mathbf{0}$ (no mean covariates) recovers exactly the recursive score derivatives of the constant precision βARMA model of [4]; retaining $\boldsymbol{\beta}$ while centering the autoregressive term as in Eq (2.2) introduces the additional forcing term $-\sum_i \varphi_i x_{l,t-i}$ in Eq (3.4), which is what allows $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varphi}}$ to be estimated without confounding the exogenous trend with the autoregressive dynamics.

3.2. Analytic matrix score equations

Collecting χ_t and ζ_t across $t = m+1, \ldots, n$ as the rows of the $(n-m) \times D_\mu$ matrix $\boldsymbol{X}$ and the $(n-m) \times D_\phi$ matrix $\boldsymbol{Z}$, the full-sample score vector admits the compact analytic matrix form

$$U(\boldsymbol{\eta}) = \begin{pmatrix} U_{\eta_\mu}(\boldsymbol{\eta}) \\ U_{\eta_\phi}(\boldsymbol{\eta}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{D}_1 \boldsymbol{\Phi}\, \boldsymbol{T} \\ \boldsymbol{Z}^\top \boldsymbol{D}_2 (\boldsymbol{M}\boldsymbol{T} + \boldsymbol{W}) \end{pmatrix}, \tag{3.7}$$

where $\boldsymbol{D}_1 = \mathrm{diag}\{1/g_1'(\mu_t)\}$, $\boldsymbol{D}_2 = \mathrm{diag}\{1/g_2'(\phi_t)\}$ are diagonal matrices of inverse link derivatives, $\boldsymbol{\Phi} = \mathrm{diag}(\phi_t)$ and $\boldsymbol{M} = \mathrm{diag}(\mu_t)$ are diagonal matrices of precision and mean parameters, and $\boldsymbol{T}$, $\boldsymbol{W}$ are the $(n-m)$ vectors with elements $(\boldsymbol{T})_t = y_t^* - \mu_t^*$ and $(\boldsymbol{W})_t = y_t^\dagger - \mu_t^\dagger$, respectively. Equation (3.7) is the exact analogue, for this observation-driven recursive model, of the iteratively reweighted least squares (IRLS) score representation familiar from ordinary GLM theory [10]: The mean submodel score is a weighted cross-product of the recursive design matrix $\boldsymbol{X}$ with the beta-family error term $\boldsymbol{T}$, and the precision submodel score is the analogous cross-product of $\boldsymbol{Z}$ with the composite error term $\boldsymbol{M}\boldsymbol{T} + \boldsymbol{W}$, which mixes both sufficient statistic residuals precisely because $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\phi$ are not information-orthogonal, as formalized next.

3.3. Fisher information matrix

The conditional Fisher information matrix $\mathcal{K}(\boldsymbol{\eta}) = \mathbb{E}[U(\boldsymbol{\eta})U(\boldsymbol{\eta})^\top \mid \mathcal{F}_{t-1}]$ is partitioned as:

$$\mathcal{K}(\boldsymbol{\eta}) = \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{\Omega}_{\mu\mu} \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{\Omega}_{\mu\phi} \boldsymbol{Z} \\ \boldsymbol{Z}^\top \boldsymbol{\Omega}_{\mu\phi} \boldsymbol{X} & \boldsymbol{Z}^\top \boldsymbol{\Omega}_{\phi\phi} \boldsymbol{Z} \end{pmatrix}, \tag{3.8}$$

with the diagonal weight matrices:

$$(\boldsymbol{\Omega}_{\mu\mu})_{tt} = \frac{\phi_t^2 \{\psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)\}}{\{g_1'(\mu_t)\}^2}, \tag{3.9}$$

$$(\boldsymbol{\Omega}_{\phi\phi})_{tt} = \frac{\mu_t^2 \psi'(\mu_t\phi_t) + (1-\mu_t)^2 \psi'((1-\mu_t)\phi_t) - \psi'(\phi_t)}{\{g_2'(\phi_t)\}^2}, \tag{3.10}$$

$$(\boldsymbol{\Omega}_{\mu\phi})_{tt} = \frac{\phi_t \{\mu_t \psi'(\mu_t\phi_t) - (1-\mu_t)\psi'((1-\mu_t)\phi_t)\}}{g_1'(\mu_t)\, g_2'(\phi_t)}. \tag{3.11}$$

3.4. Non-orthogonality and the consequences of precision submodel misspecification

In this dual submodel, the mean and precision parameters need not be information-orthogonal, which directly impacts how we approach inference and specification. In particular, estimation of the mean submodel parameters must account for the fact that the precision parameters are unknown, while an incorrectly restricted precision submodel may alter the population estimating equations for the mean parameters. The results below state these implications precisely.

Definition 1 (Conditional information orthogonality)**.** Let the conditional Fisher information matrix be partitioned conformably with $\boldsymbol{\eta} = (\boldsymbol{\eta}_\mu^\top, \boldsymbol{\eta}_\phi^\top)^\top$ as

$$\mathcal{K}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{K}_{\mu\mu}(\boldsymbol{\eta}) & \mathcal{K}_{\mu\phi}(\boldsymbol{\eta}) \\ \mathcal{K}_{\phi\mu}(\boldsymbol{\eta}) & \mathcal{K}_{\phi\phi}(\boldsymbol{\eta}) \end{pmatrix}. \tag{3.12}$$

The blocks $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\phi$ are said to be *conditionally information-orthogonal at* $\boldsymbol{\eta}_0$ if $\mathcal{K}_{\mu\phi}(\boldsymbol{\eta}_0) = \boldsymbol{0}$ almost surely. They are globally information-orthogonal if this equality holds for every admissible $\boldsymbol{\eta}$.

For the model considered here we have,

$$\mathcal{K}_{\mu\phi}(\boldsymbol{\eta}) = \boldsymbol{X}^\top \boldsymbol{\Omega}_{\mu\phi} \boldsymbol{Z}, \qquad (\boldsymbol{\Omega}_{\mu\phi})_{tt} = \frac{\phi_t \{\mu_t \psi'(\mu_t \phi_t) - (1-\mu_t)\psi'((1-\mu_t)\phi_t)\}}{g_1'(\mu_t) g_2'(\phi_t)}, \tag{3.13}$$

where ψ' denotes the trigamma function. The following proposition characterizes the observation-level cross-information weight and gives a transparent sufficient condition for sample-level non-orthogonality.

Proposition 2 (Observation-level and sample-level non-orthogonality)**.** *Assume that $\phi_t > 0$ and that the link functions are continuously differentiable with $g_1'(\mu_t) g_2'(\phi_t) \neq 0$. Then*

$$(\boldsymbol{\Omega}_{\mu\phi})_{tt} = 0 \quad \Longleftrightarrow \quad \mu_t = \frac{1}{2}. \tag{3.14}$$

Consequently, $\mu_t \equiv 1/2$ for all $t = m+1, \ldots, n$ is sufficient for $\mathcal{K}_{\mu\phi}(\boldsymbol{\eta}) = \boldsymbol{0}$.

The converse at the matrix level need not hold without conditions on the recursive design matrices, because weighted cross-products may vanish through cancellation or design orthogonality. A useful sufficient condition for non-orthogonality is the following: If $\boldsymbol{X}$ and $\boldsymbol{Z}$ both contain an intercept column, g_1' and g_2' have a constant sign, and either $\mu_t < 1/2$ for every t or $\mu_t > 1/2$ for every t, then $\mathcal{K}_{\mu\phi}(\boldsymbol{\eta}) \neq \boldsymbol{0}$.

Proof. Because $\phi_t > 0$ and the link derivatives are nonzero, the equality in (3.14) is equivalent to

$$\mu_t \psi'(\mu_t \phi_t) = (1-\mu_t)\psi'((1-\mu_t)\phi_t). \tag{3.15}$$

Let $q(x) = x\psi'(x)$ for $x > 0$. The function q is strictly decreasing on $(0, \infty)$; equivalently, $q'(x) = \psi'(x) + x\psi''(x) < 0$. Multiplying (3.15) by ϕ_t gives $q(\mu_t \phi_t) = q((1-\mu_t)\phi_t)$. The strict monotonicity of q therefore implies $\mu_t \phi_t = (1-\mu_t)\phi_t$, and hence $\mu_t = 1/2$. The reverse implication follows immediately by symmetry.

If $\mu_t \equiv 1/2$, every diagonal element of $\boldsymbol{\Omega}_{\mu\phi}$ is zero, so $\mathcal{K}_{\mu\phi} = \boldsymbol{X}^\top \boldsymbol{\Omega}_{\mu\phi} \boldsymbol{Z} = \boldsymbol{0}$. For the final assertion, note that the strict monotonicity of q implies that the numerator in (3.13) has the same nonzero sign for all t whenever all conditional means lie strictly on the same side of $1/2$. Since the link derivatives also have a constant sign, the intercept–intercept entry of $\mathcal{K}_{\mu\phi}$ is a sum of nonzero terms having the same sign. It cannot therefore vanish. □

Proposition 2 distinguishes two statements that are sometimes conflated. The beta likelihood has a nonzero observation-level mean–precision cross-information contribution whenever $\mu_t \neq 1/2$; however, whether the aggregated parameter blocks are orthogonal also depends on the design matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$. In the empirically relevant case in which both submodels contain intercepts and the fitted conditional means remain predominantly on one side of $1/2$, the cross-information block is generically nonzero.

Corollary 3 (Efficiency cost of estimating the precision submodel)**.** *Suppose that $\mathcal{K}(\boldsymbol{\eta}_0)$ is positive definite. The $\boldsymbol{\eta}_\mu$ block of the asymptotic covariance matrix of the joint conditional maximum likelihood estimator (CMLE) is*

$$\boldsymbol{V}_{\mu\mu} = \left[\mathcal{K}_{\mu\mu} - \mathcal{K}_{\mu\phi}\mathcal{K}_{\phi\phi}^{-1}\mathcal{K}_{\phi\mu}\right]^{-1}, \tag{3.16}$$

with all information blocks evaluated at $\boldsymbol{\eta}_0$*. Hence*

$$\boldsymbol{V}_{\mu\mu} \geq \mathcal{K}_{\mu\mu}^{-1}. \tag{3.17}$$

The inequality is strict in every direction $\boldsymbol{a}$ *for which* $\mathcal{K}_{\phi\mu}\mathcal{K}_{\mu\mu}^{-1}\boldsymbol{a} \neq \boldsymbol{0}$*. Thus, relative to the infeasible benchmark in which the precision parameters are known, estimating a non-orthogonal precision block weakly increases the asymptotic variance of the mean submodel estimator.*

Proof. The representation in (3.16) follows from the partitioned inverse formula [11, 12]. Since $\mathcal{K}_{\mu\phi}\mathcal{K}_{\phi\phi}^{-1}\mathcal{K}_{\phi\mu}$ is positive semidefinite,

$$\mathcal{K}_{\mu\mu} - \mathcal{K}_{\mu\phi}\mathcal{K}_{\phi\phi}^{-1}\mathcal{K}_{\phi\mu} \preceq \mathcal{K}_{\mu\mu}.$$

Inversion reverses the Loewner ordering for positive-definite matrices, which yields (3.17). The directional strictness condition follows from the corresponding quadratic form. □

The preceding corollary concerns estimation under a correctly specified joint model and compares it with the hypothetical case in which $\boldsymbol{\eta}_\phi$ is known. It does not, by itself, imply that standard errors from an arbitrarily restricted precision model are necessarily anti-conservative; under misspecification, the relevant covariance matrix is the sandwich covariance associated with the pseudo-true parameter.

To characterize the effect of precision misspecification on the probability limit of the mean estimator, let $\phi_{t,0}$ denote the true conditional precision and let ϕ_t^* denote the trajectory imposed by a restricted precision specification. Conditional on $\mathcal{F}_{t-1}$, the expected mean score contribution evaluated at the true mean parameter but at the restricted precision trajectory, it is

$$\begin{aligned}&\mathbb{E}_0\Big[U_{\mu,t}(\boldsymbol{\eta}_{\mu,0}, \boldsymbol{\eta}_\phi^*) \mid \mathcal{F}_{t-1}\Big] \\ &\quad = \frac{\phi_t^*}{g_1'(\mu_{t,0})}\chi_{t,0}\Big[\psi(\mu_{t,0}\phi_{t,0}) - \psi((1-\mu_{t,0})\phi_{t,0}) - \psi(\mu_{t,0}\phi_t^*) + \psi((1-\mu_{t,0})\phi_t^*)\Big],\end{aligned} \tag{3.18}$$

where $\chi_{t,0} = \partial\eta_{1,t}/\partial\boldsymbol{\eta}_\mu$ is evaluated at the true mean parameter. Equation (3.18) gives an exact, rather than merely local, criterion.

Proposition 4 (Mean parameter inconsistency under precision misspecification)**.** *Assume the usual regularity and identification conditions for a misspecified conditional maximum likelihood estimator. Let*

$$\boldsymbol{b}_\mu(\boldsymbol{\eta}_\phi^*) = \mathbb{E}_0\Big[U_\mu(\boldsymbol{\eta}_{\mu,0}, \boldsymbol{\eta}_\phi^*)\Big]. \tag{3.19}$$

If $\boldsymbol{b}_\mu(\boldsymbol{\eta}_\phi^*) \neq \boldsymbol{0}$*, then the restricted estimator of* $\boldsymbol{\eta}_\mu$ *is inconsistent for* $\boldsymbol{\eta}_{\mu,0}$*; instead, it converges to the pseudo-true value* $\boldsymbol{\eta}_\mu^*$ *satisfying*

$$\mathbb{E}_0\Big[U_\mu(\boldsymbol{\eta}_\mu^*, \boldsymbol{\eta}_\phi^*)\Big] = \boldsymbol{0}. \tag{3.20}$$

Moreover, for a local restriction $\boldsymbol{\eta}_\phi^* = \boldsymbol{\eta}_{\phi,0} + \boldsymbol{\delta}$*, we have*

$$\boldsymbol{b}_\mu(\boldsymbol{\eta}_\phi^*) = -\mathcal{K}_{\mu\phi}(\boldsymbol{\eta}_0)\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|). \tag{3.21}$$

Therefore, local precision misspecification displaces the population's mean score equations whenever $\mathcal{K}_{\mu\phi}(\boldsymbol{\eta}_0)\boldsymbol{\delta} \neq \boldsymbol{0}$*.*

Proof. Under the true beta distribution,

$$\mathbb{E}_0(y_t^* \mid \mathcal{F}_{t-1}) = \psi(\mu_{t,0}\phi_{t,0}) - \psi((1-\mu_{t,0})\phi_{t,0}),$$

which, upon substitution into the mean score expression, yields (3.18). If its unconditional sum is nonzero, $\boldsymbol{\eta}_{\mu,0}$ does not solve the population score equations of the restricted model. Identification of the pseudo-true value then implies $\boldsymbol{\eta}_\mu^* \neq \boldsymbol{\eta}_{\mu,0}$.

For the local result, a first-order expansion around $\boldsymbol{\eta}_0$ gives

$$\mathbb{E}_0\Big[U_\mu(\boldsymbol{\eta}_{\mu,0}, \boldsymbol{\eta}_{\phi,0} + \boldsymbol{\delta})\Big] = \mathbb{E}_0\Big[U_\mu(\boldsymbol{\eta}_0)\Big] + \mathbb{E}_0\left[\frac{\partial U_\mu(\boldsymbol{\eta}_0)}{\partial \boldsymbol{\eta}_\phi^\top}\right]\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|).$$

The first term is zero, while the information identity gives $\mathbb{E}_0[\partial U_\mu/\partial \boldsymbol{\eta}_\phi^\top] = -\mathcal{K}_{\mu\phi}$. This establishes (3.21). □

Proposition 4 supplies the precise scope of the contamination mechanism. Non-orthogonality makes the population mean score sensitive, to the first order, to perturbations of the precision parameters, but inconsistency requires that the imposed restriction have a nonzero projection through the cross-information block. Equivalently, non-orthogonality is a mechanism for transmission of the precision submodel misspecification, not a sufficient condition for bias under every possible restriction. In the present applications, this mechanism can be assessed empirically through the estimated cross-information block, the fitted precision trajectories, and the simulation designs that omit genuine exogenous precision drivers.

The information matrix defined above is a full-sample quantity. Each of its blocks is a sum of $n-m$ observation-level contributions and is therefore $O(n)$. To state the limiting distribution, we use the corresponding per-observation limit, and we write $\mathcal{K}_n(\boldsymbol{\eta})$ for the full-sample matrix and

$$\bar{\mathcal{K}}(\boldsymbol{\eta}) = \operatorname*{plim}_{n\to\infty} n^{-1}\mathcal{K}_n(\boldsymbol{\eta}) \tag{3.22}$$

as its per-observation limit, which is assumed to exist and be finite and positive definite at $\boldsymbol{\eta}_0$. Under the standard regularity conditions for observation-driven conditional maximum likelihood estimation [2, 4, 5], the conditional maximum likelihood estimator $\hat{\boldsymbol{\eta}}$ is consistent and asymptotically normally distributed as follows:

$$\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) \xrightarrow{d} N\Big(\mathbf{0},\, \bar{\mathcal{K}}(\boldsymbol{\eta}_0)^{-1}\Big). \tag{3.23}$$

In finite samples, $\mathcal{K}_n(\hat{\boldsymbol{\eta}})^{-1}$ provides the usual information-based estimate of the covariance matrix and is used for the standard errors and test statistics in Section 4 and throughout the empirical applications. This result provides the basis for inference on the exogenous precision parameters $\boldsymbol{\gamma}$.

4. Confidence intervals and hypothesis testing

Following the establishment of the large-sample properties of the CMLE, this section develops likelihood-based inference procedures for the proposed βARMA model with covariate-driven time-varying precision. The inferential framework is grounded in the asymptotic normality of the CMLE under a conditional likelihood approach, consistent with [13]. In beta-distributed time series, likelihood-based inference performs robustly even under complex dynamic specifications, provided that the conditional density is correctly specified and the observation-driven recursion is stable [1, 4].

4.1. Asymptotic confidence intervals

Let $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1, \ldots, \hat{\eta}_D)^\top$ denote the CMLE of dimension D. Under standard regularity conditions, the estimated standard error for the parameter $\hat{\eta}_j$ is $\widehat{\mathrm{se}}(\hat{\eta}_j) = \sqrt{[\mathcal{K}(\hat{\boldsymbol{\eta}})^{-1}]_{jj}}$, and an asymptotic $(1-\alpha)100\%$ confidence interval is:

$$\mathrm{CI}(\eta_j, 1-\alpha) = \left[\hat{\eta}_j - z_{1-\alpha/2}\,\widehat{\mathrm{se}}(\hat{\eta}_j),\ \hat{\eta}_j + z_{1-\alpha/2}\,\widehat{\mathrm{se}}(\hat{\eta}_j)\right], \tag{4.1}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution [14, 15]. These intervals facilitate direct inference on the persistence parameters (φ, θ) and the structural impacts of the exogenous covariates $(\boldsymbol{\beta}, \boldsymbol{\gamma})$. For nonlinear functions of the parameters, such as the predicted conditional mean $\mu_t = g_1^{-1}(\eta_{1,t})$, the delta method yields the associated uncertainty bounds [15, 16].

4.2. Large-sample hypothesis tests

We consider the general linear hypothesis $H_0 : \boldsymbol{C\eta} = \boldsymbol{\lambda}$ versus $H_1 : \boldsymbol{C\eta} \neq \boldsymbol{\lambda}$, where $\boldsymbol{C}$ is a $k \times D$ matrix of full row rank and $\boldsymbol{\lambda}$ is a $k \times 1$ vector of constants. The Wald statistic is:

$$W = (\boldsymbol{C\hat{\eta}} - \boldsymbol{\lambda})^\top [\boldsymbol{C}\mathcal{K}(\hat{\boldsymbol{\eta}})^{-1}\boldsymbol{C}^\top]^{-1}(\boldsymbol{C\hat{\eta}} - \boldsymbol{\lambda}). \tag{4.2}$$

Under H_0, $W \xrightarrow{d} \chi^2_k$ [1,15]. This test is particularly useful for verifying the necessity of the time-varying precision submodel by testing $H_0 : \boldsymbol{\gamma} = \mathbf{0}$; rejection provides formal evidence that the conditional dispersion of the series is driven by exogenous covariates rather than being static.

Alternatively, the LR test compares the maximized conditional log-likelihoods of the unrestricted ($\ell(\hat{\boldsymbol{\eta}})$) and restricted ($\ell(\tilde{\boldsymbol{\eta}})$) models:

$$\mathrm{LR} = 2\{\ell(\hat{\boldsymbol{\eta}}) - \ell(\tilde{\boldsymbol{\eta}})\}. \tag{4.3}$$

Under H_0, $\mathrm{LR} \xrightarrow{d} \chi^2_k$. Given the non-orthogonality of the parameter blocks, the LR test is often preferred in finite samples for testing restrictions on the precision submodel, as it accounts for the global curvature of the likelihood surface more comprehensively than the Wald test [4, 14, 16].

5. Model selection and information criteria

The generalized βARMA model requires simultaneous choices of (i) the autoregressive and moving average orders (p, q), (ii) the set of covariates entering the mean submodel, and (iii) the set of covariates (if any) entering the precision submodel. To balance the model's parsimony and predictive accuracy, we use a best-subset selection strategy based on information criteria, complemented by a diagnostic filter for residual autocorrelation.

We consider all candidate models with $p, q \in \{0, \ldots, P_{\max}\}$ and all possible covariate subsets for both submodels. Each candidate is estimated by conditional maximum likelihood using the L-BFGS-B algorithm, supplied with the analytic first derivatives derived in Section 3. For each estimated model, we compute the Akaike information criterion (AIC) [17] and the Bayesian information criterion (BIC) [18]:

$$\mathrm{AIC} = -2\ell(\widehat{\boldsymbol{\eta}}) + 2k, \qquad \mathrm{BIC} = -2\ell(\widehat{\boldsymbol{\eta}}) + k\log(n), \tag{5.1}$$

where $\ell(\widehat{\boldsymbol{\eta}})$ is the maximized conditional log-likelihood and k is the total number of estimated parameters (combining the mean and precision submodels). The AIC aims for efficient prediction, while the BIC is consistent for model identification as $n \to \infty$.

To ensure that a candidate model adequately captures serial dependence, we apply the Q_4 portmanteau test [6] to the link-scale residuals $r_t = g_1(y_t) - g_1(\mu_t)$. Models that reject the null hypothesis of no residual autocorrelation at the 5% significance level are discarded. From the remaining well-specified models, we retain two candidates for out-of-sample forecasting comparison: A fixed-precision βARMA model (Benchmark 1), and a model in which the precision submodel is restricted to the lagged proxy $z_{t-1} = y_{t-1}(1-y_{t-1})$ (Benchmark 2), adapting the autoregressive precision mechanism of [2] – originally proposed for a mean submodel without exogenous covariates – to the same covariate-augmented mean submodel used by the proposed model and Benchmark 1, so that all three specifications are compared on an equal covariate footing. This protocol ensures that final models are both statistically adequate and parsimonious.

6. Forecasting strategy and evaluation criteria

The proposed generalized βARMA model is designed to produce accurate point forecasts and well-calibrated predictive distributions for bounded time series on the unit interval. Let $\{y_t\}_{t=1}^n$ with $0 < y_t < 1$, and let $\mathcal{F}_n$ be the information set at the forecast origin n. For each horizon $h \geq 1$, the model yields a conditional predictive distribution for y_{n+h} given $\mathcal{F}_n$.

6.1. Recursive multi-step forecasting

Under the βARMA specification, $y_t \mid \mathcal{F}_{t-1} \sim \text{Beta}(\mu_t\phi_t, (1-\mu_t)\phi_t)$, the h-step-ahead conditional mean forecast is:

$$\hat{y}_{n+h|n} = \hat{\mu}_{n+h|n} = g_1^{-1}(\hat{\eta}_{1,n+h|n}), \qquad h \geq 1, \tag{6.1}$$

where $\hat{\eta}_{1,n+h|n}$ is computed recursively by a plug-in scheme as follows:

$$\hat{\eta}_{1,n+h|n} = \alpha_1 + \mathbf{x}_{n+h}^\top \boldsymbol{\beta} + \sum_{i=1}^{p} \varphi_i \hat{\tau}_{n+h-i|n} + \sum_{j=1}^{q} \theta_j \hat{r}_{n+h-j|n}, \tag{6.2}$$

with

$$\hat{\tau}_{n+k|n} = \begin{cases} g_1(y_{n+k}) - \mathbf{x}_{n+k}^\top \boldsymbol{\beta}, & k \leq 0, \\ \hat{\eta}_{1,n+k|n} - \mathbf{x}_{n+k}^\top \boldsymbol{\beta}, & k > 0, \end{cases} \qquad \hat{r}_{n+k|n} = \begin{cases} g_1(y_{n+k}) - \eta_{1,n+k}, & k \leq 0, \\ 0, & k > 0. \end{cases} \tag{6.3}$$

The precision submodel yields a forecast precision $\hat{\phi}_{n+h|n} = g_2^{-1}(\alpha_2 + \mathbf{z}_{n+h}^\top \boldsymbol{\gamma})$, so the plug-in predictive distribution at each horizon is:

$$y_{n+h} \mid \mathcal{F}_n \sim \text{Beta}\Big(\hat{\mu}_{n+h|n}\hat{\phi}_{n+h|n},\ (1-\hat{\mu}_{n+h|n})\hat{\phi}_{n+h|n}\Big). \tag{6.4}$$

For $h = 1$, this is the exact conditional law of y_{n+1} given $\mathcal{F}_n$. For $h \geq 2$, the intermediate observations are unobserved and enter the recursion, so the exact predictive distribution is generally a mixture of beta distributions over their possible future values. The expression above is therefore a plug-in approximation obtained by replacing these future observations with their forecasts, and formal calibration is assessed only at $h = 1$.

The forecast recursion requires the future covariate values $\mathbf{x}_{n+h}$ and $\mathbf{z}_{n+h}$, which are generally unknown at the forecast origin. We therefore evaluate forecasts under two regimes. In the *ex*

post (conditional) regime, the realized covariate values over the forecast window are used at the prediction step; they play no role in estimation or selection. This isolates the contribution of the model specification from the separate problem of forecasting the covariates [1, 19], but it does not correspond to an operational forecast. In the *ex ante* regime, every non-deterministic covariate is replaced by its forecast from an ARIMA model selected on the training window by the automatic algorithm of [20], in which the order of differencing is chosen by Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit-root tests and the autoregressive and moving-average orders by the corrected AIC. The break dummy enters these models as a regressor, since its future values are known, so that only information available at the forecast origin is used. Both regimes use the same fitted models, so differences between them reflect only the treatment of the covariates.

Deterministic covariates, such as the structural break indicators used in Section 8, are treated as known in both regimes. In both applications, the onset and resolution of the break fall within the initial training window, so the indicator is fully determined by information available at every forecast origin and introduces no look-ahead bias. For the other covariates, whether the indicator enters the precision submodel is decided by the subset selection procedure applied to the initial training window.

6.2. Point forecast evaluation

Let H denote the number of out-of-sample observations and $e_{n+h} = y_{n+h} - \hat{y}_{n+h|n}$ the forecast error at horizon h. We summarize point forecast accuracy using the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) [19, 21]:

$$\mathrm{RMSE} = \sqrt{\tfrac{1}{H}\sum_{h=1}^{H} e_{n+h}^2}, \qquad \mathrm{MAE} = \tfrac{1}{H}\sum_{h=1}^{H} |e_{n+h}|, \qquad \mathrm{MAPE} = \tfrac{100}{H}\sum_{h=1}^{H} |e_{n+h}/y_{n+h}|. \tag{6.5}$$

For all metrics smaller values indicate better accuracy.

6.3. Probabilistic forecast evaluation

Since the proposed model also produces a time-varying precision parameter, evaluation of the complete density forecast is essential. We use three complementary measures.

Continuous ranked probability score (CRPS). For the predictive cumulative distribution function (CDF) $F_{n+h|n}$ and the realized value y_{n+h}:

$$\mathrm{CRPS}(F_{n+h|n}, y_{n+h}) = \int_0^1 \left[F_{n+h|n}(u) - \mathbb{1}\{u \geq y_{n+h}\}\right]^2 du. \tag{6.6}$$

The CRPS is a strictly proper scoring rule that rewards both calibration and sharpness simultaneously; smaller values indicate better probabilistic forecasts [22].

Weighted Interval Score (WIS). Given K central $(1 - \alpha_k)$ prediction intervals with lower and upper predictive quantiles $\ell_{k,n+h|n}$ and $u_{k,n+h|n}$, the WIS is:

$$\mathrm{WIS}_{n+h|n} = \frac{1}{K + \frac{1}{2}} \left\{ \tfrac{1}{2}|y_{n+h} - m_{n+h|n}| + \sum_{k=1}^{K} \tfrac{\alpha_k}{2}\, \mathrm{IS}_{\alpha_k} \right\}, \tag{6.7}$$

where $m_{n+h|n} = F^{-1}_{n+h|n}(0.5)$ is the predictive median and IS_{α_k} penalizes the interval width and any coverage violations. Smaller WIS values indicate better probabilistic forecasts [23].

Predictive standard deviation (PredSD). Sharpness is quantified by the average predictive standard deviation [24]:

$$\text{PredSD} = \frac{1}{H}\sum_{h=1}^{H}\sqrt{\frac{\hat{\mu}_{n+h|n}(1-\hat{\mu}_{n+h|n})}{1+\hat{\phi}_{n+h|n}}}. \tag{6.8}$$

Smaller PredSD values correspond to sharper forecasts, although excessive sharpness without calibration is not desirable [25].

For a correctly specified continuous predictive distribution, the probability integral transform (PIT), $F_{n+h|n}(y_{n+h})$, is uniform on $(0, 1)$. Since the one-step-ahead predictive law is exactly a beta distribution, we assess calibration at $h = 1$ using PIT's uniformity and empirical coverage of the central prediction intervals. Because both empirical series are recorded on a discrete grid, we use the non-randomized PIT of [26], $\bar{F}(u)$, which averages the predictive distribution over the rounding interval. We summarize it by the maximum deviation $\sup_u |\bar{F}(u) - u|$ from the uniform distribution function and by the test of [27], which checks that Φ^{-1} of the mid-interval PIT has a zero mean, unit variance, and no first-order autocorrelation. Both p-values are obtained by a parametric bootstrap with 2000 draws [28], in which the observations are simulated from the predictive beta laws and rounded to the recording grid. Coverage of the central $(1 - \alpha)$ interval is computed as $\bar{F}(1 - \alpha/2) - \bar{F}(\alpha/2)$. Multi-step PIT values are not used for calibration because they are generally serially dependent particularly when forecast horizons overlap.

Together with the CRPS, WIS, and PredSD, these diagnostics can assess calibration, sharpness, and accuracy. A model is considered to be superior when it achieves lower point-forecast errors, lower CRPS and WIS values, and predictive intervals whose empirical coverage is close to nominal.

To confirm that the observed forecast improvements are not attributable to sampling variability, we apply the Diebold–Mariano (DM) test [19, 29] under squared error loss and CRPS. The test is implemented with a heteroscedasticity- and autocorrelation-consistent (HAC) variance estimator and the Harvey–Leybourne–Newbold (HLN) correction to account for finite-sample biases in overlapping multi-step forecast errors [30, 31].

7. Monte Carlo simulation study

We evaluate the finite-sample performance of the estimator, hypothesis tests, and model selection procedures through Monte Carlo experiments before presenting the empirical applications. We conduct eight complementary experiments examining distinct aspects of the proposed methodology: (S1) Finite-sample bias, efficiency, and the accuracy of asymptotic confidence intervals; (S2) the size and power of the large-sample tests developed in Section 4 for the necessity of the precision submodel; (S3) the accuracy of the joint order- and covariate selection procedure of Section 5, benchmarked against both a constant precision specification and an autoregressive-type precision specification driven by a lagged variance proxy [2]; (S4) robustness to precision link misspecification and the consequences of omitting a genuine precision covariate for inference on the mean submodel; (S5) robustness to non-Gaussian, heavy-tailed exogenous shocks; (S6) out-of-sample forecasting performance under a simulated structural break; (S7) out-of-sample forecasting performance when the model itself, not only its parameters, is selected from the data in every replication, using the joint order- and covariate-selection procedure of Section 5 rather than the true model order assumed in S1–S6; and (S8) a targeted quantification of the bias induced in the mean submodel slope when a genuine precision covariate

is omitted, across a factorial grid of precision covariate strength and mean/precision covariates' correlation. All experiments share a common data-generating process and estimation engine, which is described first.

7.1. Data-generating process

In every experiment, a bounded time series $\{y_t\}_{t=1}^{n}$ is generated according to the conditional beta distribution

$$y_t \mid \mathcal{F}_{t-1} \sim \text{Beta}(\mu_t\phi_t,\ (1-\mu_t)\phi_t), \tag{7.1}$$

with μ_t and ϕ_t governed by the dual exogenous structure of Eqs (2.2) and (2.3):

$$\text{logit}(\mu_t) = \alpha_1 + \mathbf{x}_t^\top\boldsymbol{\beta} + \sum_{i=1}^{p}\varphi_i[\text{logit}(y_{t-i}) - \mathbf{x}_{t-i}^\top\boldsymbol{\beta}] + \sum_{j=1}^{q}\theta_j r_{t-j}, \tag{7.2}$$

$$g_2(\phi_t) = \alpha_2 + \mathbf{z}_t^\top\boldsymbol{\gamma}, \tag{7.3}$$

where $r_t = \text{logit}(y_t) - \eta_t$ and g_2 is the precision link function, set to $\log(\cdot)$ unless explicitly varied (Section 7.5). The baseline data-generating configuration is a βARMA(1, 1) process with one relevant mean covariate, one irrelevant (noise) mean covariate, and one relevant precision covariate.

- *Mean submodel:* $\alpha_1 = -1.25$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (0.5,\ 0)^\top$, $\varphi_1 = 0.8$, $\theta_1 = 0.2$.
- *Precision submodel:* $\alpha_2 = \log(100)$, $\boldsymbol{\gamma} = (\gamma_1)^\top = (0.5)^\top$.

This baseline configuration is held fixed across S1, S3, S4, S5, and S7; it is perturbed along a single dimension at a time in S2 (the magnitude of γ_1) and S6 (a temporary shift in α_1 and α_2), as detailed below. Experiment S8 departs from this shared baseline more substantially, as a single mean covariate and a single, correlated precision covariate replace the two-covariate structure below, it described separately in its own subsection. Exogenous covariates for S1–S7 are generated as stationary, mean-zero AR(1) processes,

$$X_{1,t} = 0.45\,X_{1,t-1} + \epsilon_{1,t}, \quad X_{2,t} = 0.40\,X_{2,t-1} + \epsilon_{2,t}, \quad Z_{1,t} = 0.40\,Z_{1,t-1} + \epsilon_{3,t}, \tag{7.4}$$

with the innovations $\epsilon_{k,t} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_k^2)$, $(\sigma_1, \sigma_2, \sigma_3) = (0.80,\ 0.75,\ 0.65)$, in all experiments except S5, where Gaussian innovations are replaced by scaled Student-t_5 innovations to assess the robustness to heavy-tailed exogenous shocks while preserving the marginal variance. Each replicated series is generated with a burn-in of 150 observations (discarded prior to estimation) to remove dependence on the recursion's initial values, following the initialization convention of [9] described in Section 3.

Every candidate model in every replication is estimated by conditional maximum likelihood as described in Section 2. Optimization uses a box-constrained quasi-Newton algorithm (L-BFGS-B) supplied with the analytic score function of the conditional log-likelihood derived in Section 3. Starting values are obtained from a data-driven initialization of the mean and precision intercepts. Asymptotic standard errors are obtained from the inverse of the conditional Fisher information matrix $\mathcal{K}(\hat{\boldsymbol{\eta}})$ (Section 3), evaluated in closed form rather than by numerical differentiation.

7.2. Experiment S1: Finite-sample bias, efficiency, and confidence interval coverage

Experiment S1 isolates the finite-sample behavior of the conditional maximum likelihood estimator from the noise introduced by model selection: The true model order $(p, q) = (1, 1)$ and the true covariate sets are taken as known, and only the parameter estimation is evaluated. For sample sizes $n \in \{100, 250, 500, 1000\}$ and $R = 1000$ independent replications per sample size, we record, for every parameter in $\boldsymbol{\eta} = (\alpha_1, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \boldsymbol{\theta}^\top, \alpha_2, \boldsymbol{\gamma}^\top)^\top$, the empirical *bias*, $R^{-1} \sum_{r=1}^{R} (\hat{\eta}_r - \eta_0)$; the empirical RMSE $\sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\eta}_r - \eta_0)^2}$; and the empirical coverage rate of the nominal 95% asymptotic confidence interval (CI) defined in Section 4, i.e., the proportion of replications for which $\eta_0 \in \mathrm{CI}(\hat{\eta}_r, 0.95)$.

For direct comparison, the same statistics are computed for a constant precision benchmark (the proposed mean submodel paired with $\phi_t \equiv \phi$, i.e., $\boldsymbol{\gamma}$ is restricted to zero) fitted to the identical simulated data, isolating the estimation consequences of erroneously omitting genuine exogenous drivers of ϕ_t. As $n \to \infty$,the consistency and asymptotic normality of the CMLE (Section 3) imply that bias and RMSE should vanish and that empirical coverage should converge to the nominal level for the correctly specified (proposed) model, while no such guarantee holds for the misspecified benchmark.

7.3. Experiment S2: Size and power of tests for the precision submodel necessity

The second experiment evaluates the large-sample hypothesis tests of Section 4 for $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ versus $H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$, the test that operationalizes the paper's central methodological claim that precision can be driven by exogenous covariates. Data are generated under the baseline configuration of Section 7.1, with the true precision coefficient γ_1 varied over a grid $\gamma_1 \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, crossed with the sample sizes $n \in \{100, 250, 500\}$ and $R = 500$ independent replications per sample size. For each combination, both the Wald statistic W and the LR are computed for every replication, and the rejection rates at the nominal $\alpha = 0.05$ significance level are recorded. The case with $\gamma_1 = 0$ yields the empirical *size* of each test, which should track the nominal level under correct asymptotic calibration; the cases $\gamma_1 > 0$ trace out empirical *power curves* as a function of the effect size and sample size, allowing a direct comparison of the finite-sample power properties of the Wald and LR approaches discussed in Section 4.

7.4. Experiment S3: Structural identification accuracy

The third experiment extends the model selection assessment of the original submission along two dimensions: A wider sample-size grid, $n \in \{100, 250, 500\}$, $R = 500$ independent replications per sample size, and an additional benchmark specification. For each replication, an exhaustive search over the candidate space $p \in \{0, 1, 2\}$, $q \in \{0, 1, 2\}$, all subsets of the mean covariates $\{X_1, X_2\}$, and all subsets of the precision covariate $\{Z_1\}$ is conducted. To assess the sensitivity of the selection outcome to the choice of information criterion, the final model is selected using both the AIC and the BIC, consistent with Section 5. Three model classes are compared on the identical simulated data:

1. *Proposed:* The dual exogenous specification of Eqs (7.2)–(7.3), selected over the full candidate space described above;
2. *Benchmark 1 (constant precision):* The same mean submodel with $\phi_t \equiv \phi$, selected over (p, q) and mean-covariate subsets only;
3. *Benchmark 2 (lagged variance proxy):* The precision submodel restricted to a single

autoregressive-type covariate, $z_{t-1} = y_{t-1}(1 - y_{t-1})$, following [2], selected over (p, q) only.

For each selection criterion (AIC and BIC), model class, and sample size, we report: The recovery rate of the true mean covariate set, the true precision covariate set (where applicable), and the true ARMA order; the average false positive and false negative counts for covariate selection; and the average maximized log-likelihood, enabling a direct comparison of in-sample fit across the three model classes under identical data-generating conditions.

7.5. Experiment S4: Misspecification robustness

The fourth experiment examines two distinct forms of misspecification.

Precision-link misspecification. Data are generated under the baseline configuration with the canonical $\log(\cdot)$ precision link. For each replication, the dual exogenous model is estimated under three candidate precision links, $g_2 \in \{\log, \text{identity}, \sqrt{\cdot}\}$, and the BIC is compared across links to assess whether the model-selection procedure consistently favors the correctly specified link. Note that, unlike the logarithmic link, the identity and square root links require $\zeta_t > 0$ to ensure that the precision parameter remains in $(0, \infty)$. This restriction is not imposed in the general specification in Eq (2.3) and, as Table 16 shows, is binding in one of the two empirical applications.

Omitted-covariate contamination of the mean submodel. As discussed in Section 3, the non-orthogonality of $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\phi$ provides a channel through which the precision submodel misspecification can affect the mean submodel estimating equations. To assess this mechanism directly, for each replication we fit (i) the correctly specified dual exogenous model and (ii) a model that omits the true precision covariate Z_1 (constant precision), and compare the bias and RMSE of $\hat{\beta}_1$, a mean submodel parameter, across the two specifications. A systematic increase in bias under the misspecified specification provides empirical evidence that the omitted precision structure has a nonzero projection onto the mean score equations in that design.

7.6. Experiment S5: Robustness to heavy-tailed exogenous shocks

The fifth experiment assesses the sensitivity of the CMLE to departures from Gaussianity in the exogenous covariate processes. Using the fixed sample size $n = 300$ and $R = 500$, the AR(1) covariate-generating processes of Eq (7.4) are simulated twice: Once with Gaussian innovations $\epsilon_{k,t} \sim \mathcal{N}(0, \sigma_k^2)$, and once with scaled Student-t_5 innovations, $\epsilon_{k,t} = \sigma_k \sqrt{(\nu - 2)/\nu}\, T_{k,t}$ with $T_{k,t} \sim t_\nu$, $\nu = 5$, matched on marginal variance. The bias and RMSE of the full parameter vector $\hat{\boldsymbol{\eta}}$ are compared across the two innovation distributions to evaluate whether heavy-tailed exogenous shocks materially degrade estimation accuracy, a relevant robustness consideration, given that exogenous covariates in applied settings often exhibit occasional large excursions.

7.7. Experiment S6: Forecast performance under a simulated structural break

The sixth experiment evaluates the forecast performance under a structural break in a controlled simulation setting. Data are generated under the baseline configuration of Section 7.1, augmented with a temporary, exogenously timed perturbation. Over a short window of ℓ consecutive periods located at approximately 70% of the in-sample horizon, the mean intercept is shifted upward, $\alpha_1 \to \alpha_1 + \delta$ ($\delta > 0$, representing a sudden rise in the bounded outcome), and the precision intercept is simultaneously reduced, such that $\alpha_2 \to \alpha_2 + \log(\kappa)$ with $\kappa \in (0, 1)$ (representing a contemporaneous

spike in conditional volatility). Outside this window, the data-generating process reverts to the baseline parameters. For each replication, an in-sample segment of length n (which includes the perturbation window) is used to estimate the proposed model, Benchmark 1 (constant precision), and Benchmark 2 (lagged variance proxy). $H = 6$-step-ahead recursive forecasts are generated for each model immediately following the in-sample period using the plug-in scheme of Section 6. The Average absolute forecast error is computed by the horizon $h = 1, \ldots, 6$ across replications for each of the three models, providing a Monte Carlo analogue of the rolling-window forecast comparison reported in Section 8.3, but under data-generating conditions where the true magnitude and timing of the structural shock are known by construction.

7.8. Experiment S7: Out-of-sample forecasting under data-driven model selection

Experiments S1–S6 evaluate estimation and forecasting conditional on the true model order and covariate set being known. The seventh experiment removes this advantage. In every replication, the mean order p, the moving-average order q, and the mean and precision covariate subsets are themselves selected from the training data, exactly mirroring the procedure an analyst would apply to a real series. Data are generated under the baseline configuration of Section 7.1 for a combined sample of length $n + H$, with $n = 250$ observations used for estimation and model selection, and the final $H = 6$ observations withheld for out-of-sample evaluation. On the training segment, an exhaustive search over $p, q \in \{0, 1, 2\}$ and all subsets of the mean and precision covariates is conducted as in Section 7.4, yielding two data-driven specifications: The model minimizing the AIC and the model minimizing the BIC. These are compared against the same two benchmarks used in Experiment S3 – constant precision, and the lagged variance proxy precision specification of [2] – each independently re-selected over $p, q \in \{0, 1, 2\}$ on the training segment. For all four specifications, H-step-ahead recursive forecasts are generated using the plug-in scheme of Section 6, and two forecast accuracy measures are recorded by the horizon $h = 1, \ldots, 6$ across replications: The RMSE of the point forecast $\hat{\mu}_t$, and the CRPS of the full predictive beta distribution implied by $(\hat{\mu}_t, \hat{\phi}_t)$. The comparison isolates the practical cost – if any – of the additional uncertainty introduced by the data-driven order and covariate selection, relative to the idealized known-model setting of Experiment S6, and evaluates whether the proposed dual exogenous specification retains a forecasting advantage once this realistic source of uncertainty is included.

7.9. Experiment S8: Consequences of omitting the precision submodel

The eighth experiment isolates and quantifies, over a wider design than the single configuration considered in Experiment S4, the bias that omitting a genuine precision covariate induces in the mean submodel. The data-generating process is a dual exogenous βARMA$(1, 1)$ model with a single mean covariate x_t and a single precision covariate z_t,

$$\text{logit}(\mu_t) = \alpha_1 + \beta x_t + \varphi \tau_{t-1} + \theta r_{t-1}, \qquad \phi_t = \exp(\alpha_2 + \gamma z_t), \tag{7.5}$$

with (x_t, z_t) drawn as jointly normal, contemporaneously correlated pairs, $\text{Corr}(x_t, z_t) = \rho$ This is a deliberate departure from the independent AR(1) covariate processes of Eq (7.4) used elsewhere in this study, since the object of interest here is specifically how correlation between the omitted precision covariate and the retained mean covariate shapes the resulting bias. The remaining parameters are fixed at the study's baseline values, $\alpha_1 = -1.25, \beta = 0.5, \varphi = 0.8, \theta = 0.2$, and $\alpha_2 = \log(100)$. For each

replication, two estimators are compared on identical simulated data: The *full* model, which includes z_t in the precision submodel, and a *misspecified* model that assumes constant precision, omitting z_t entirely. The design crosses the true precision covariate effect $\gamma \in \{0.5, 0.75, 1.0, 1.25\}$, the covariate correlation $\rho \in \{0, 0.3, 0.6\}$, and the sample size $n \in \{250, 500\}$, with $R = 500$ independent replications per cell. For every cell, the bias and RMSE of the estimated mean submodel slope $\hat{\beta}$ are reported for both the full and misspecified models, together with the realized convergence rate of each. Because $\rho = 0$ reduces to a case in which the omitted covariate is uncorrelated with x_t, this experiment nests a partial replication of Experiment S4's omitted covariate finding while additionally tracing out how the resulting bias in $\hat{\beta}$ grows with the strength of the precision effect γ and the mean/precision covariate correlation ρ – the two mechanisms through which, as per the non-orthogonality of $\boldsymbol{\eta}_\mu$ and $\boldsymbol{\eta}_\phi$ discussed in Section 3, precision misspecification can contaminate the mean submodel inference.

7.10. Discussion of the simulation results

Across Experiments S1–S8, the finite-sample results align closely with the asymptotic theory of the CMLE, the reliability of the associated inferential and model-selection procedures, and the value of allowing the precision parameter to depend on exogenous covariates, interpreted against the asymptotic theory of Section 3.

Estimation accuracy and inference. Experiment S1 (Table 1) is consistent with the asymptotic theory of Section 3: Bias and RMSE decline monotonically as n increases from 100 to 1000, and empirical coverage of the nominal 95% CI, while remaining somewhat below target at $n = 100$ for several parameters (e.g., 0.874 for α_1, 0.892 for α_2), converges to within one to two percentage points of the nominal level by $n = 500$–1000 for both submodels. Experiment S2 (Table 2) shows that both the Wald and LR tests of $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ are mildly oversized in small samples ($n = 100$: empirical size 0.092 and 0.066, respectively, against a nominal 0.05) but approach the nominal level as n grows (at $n = 500$: 0.050 and 0.038), with power against $\gamma_1 > 0$ reaching near-unity by $n = 500$ for effects as small as $\gamma_1 = 0.4$. The LR test is consistently less liberal, and at small effect sizes, is marginally less powerful, than the Wald test, supporting the recommendation of Section 4 to prefer it in finite samples, given the non-orthogonality of the two submodels.

Model selection. Experiment S3 (Table 3) reveals a clear asymmetry between the AIC and BIC across submodels. BIC recovers the true mean-covariate set substantially more often than AIC at every n (e.g., 0.942 versus 0.819 at $n = 100$; 0.941 versus 0.826 at $n = 500$), consistent with its heavier parsimony penalty, but is markedly less reliable at recovering the true precision covariate in small samples (0.574 versus 0.797 at $n = 100$), since the precision submodel's information content per observation is intrinsically weaker. This asymmetry narrows sharply by $n = 500$ (BIC 0.939 vs. AIC 0.961 for precision-covariate recovery). The AIC also dominates ARMA-order recovery at every sample size, though the two criteria converge as n grows (0.612 vs. 0.386 at $n = 100$; 0.876 vs. 0.866 at $n = 500$). Because neither criterion dominates uniformly across all three selection targets, both AIC- and BIC-selected specifications are retained for the out-of-sample comparison in Experiment S7.

Misspecification and robustness. Experiment S4 (Tables 4 and 5) provides two robustness checks. Under link misspecification, BIC-based selection favors the true logarithmic precision link in the plurality of replications (0.676, versus 0.142 and 0.182 for the identity and square-root links), while a narrow spread of average BIC values across links (-1067.47 to -1065.71) shows that the misspecified links can still approximate the true dispersion structure reasonably well. Omitting the precision covariate increases the RMSE of $\hat{\beta}$ in every cell, and the increase grows with the strength of the omitted effect; at $\rho = 0.6$ and $n = 250$, for example, the RMSE rises from 0.01502 at $\gamma = 0.50$ to 0.01794 at $\gamma = 1.25$. The bias, however, depends on the covariates' correlation. When the omitted precision covariate is uncorrelated with the retained mean covariate ($\rho = 0$), the bias of the misspecified estimator remains small and shows no systematic sign, as expected when the omitted structure has no projection through the cross-information block. Once $\rho > 0$, the RMSE of the misspecified $\hat{\beta}_1$ increases steadily with the true precision effect γ, most clearly at $\rho = 0.6$, where it rises from 0.01502 at $\gamma = 0.50$ to 0.01794 at $\gamma = 1.25$ ($n = 250$). At this most adverse combination ($\rho = 0.6$, $\gamma = 1.25$, $n = 250$), the misspecified model's bias reaches 0.00526, roughly an order of magnitude larger than the 0.00049 bias of the correctly specified model in the same cell. The consequences of ignoring exogenous precision drivers therefore reach beyond the precision submodel itself: The omission transmits bias into the mean submodel estimates, and the size of that bias grows with both the strength of the omitted effect and its correlation with the retained mean covariate. Experiment S5 (Table 6) shows that replacing Gaussian covariate innovations with heavy-tailed, variance-matched Student-t_5 innovations leaves the overall estimation accuracy essentially unchanged: RMSE is marginally smaller under t_5 innovations for the mean dynamics parameters (α_1, φ_1, θ_1) and marginally larger for the precision submodel parameters (α_2, γ_1), a pattern indicating that the CMLE remains reliable under moderate departures from Gaussianity in the covariate process, with any loss of efficiency confined to, and small within, the precision submodel.

Forecasting. Experiments S6 and S7 evaluate the forecasting consequences of the dual exogenous specification. Under a simulated structural break (Experiment S6, Table 7), the proposed model achieves the lowest average absolute forecast error at three of the six horizons ($h = 1, 3, 4$), with Benchmark 2 marginally ahead at the remaining three ($h = 2, 5, 6$); in every case, the differences across the three models are small in absolute terms (on the order of 10^{-4}), suggesting that for a break of this magnitude, the precision specification has little effect on point-forecast accuracy. Experiment S7 (Tables 8 and 9) shows that a forecasting advantage is preserved once data-driven order and covariate selection are accounted for: Both the AIC- and BIC-selected specifications outperform the fixed-precision benchmarks at every horizon in terms of both RMSE and CRPS. The AIC-selected specification is consistently, though only marginally, ahead of the BIC-selected one at every horizon. The relative advantage of the data-driven specifications over the fixed-precision benchmarks narrows somewhat at longer horizons – for example, the RMSE reduction over Benchmark 1 falls from roughly 26% at $h = 1$ to 19% at $h = 6$ – even as the absolute error gap widens, consistent with the value of a precisely estimated precision submodel being greatest near the forecast origin and diminishing as all models converge toward the unconditional predictive distribution.

Overall, the simulations show that the proposed estimator behaves as the asymptotic theory predicts, and is robust to heavy-tailed covariate shocks. BIC favors the true precision link more often than the alternatives, although misspecified links can approximate the dispersion structure closely. When

the precision covariate is relevant, the proposed specification forecasts more accurately than the constant precision and the lagged proxy baselines (Experiment S7), while under the structural break of Experiment S6, the three models are practically indistinguishable. Experiments S4 and S8 (Tables 5 and 10) provide the most direct simulation evidence on the transmission mechanism of Section 3.4: Omitting a genuine precision covariate reduces the efficiency of the mean submodel estimates and, when the omitted covariate is correlated with the mean covariates, biases them. These findings support estimating the mean and precision submodels jointly.

Table 1. Experiment S1: Finite-sample bias, RMSE, and empirical coverage of the nominal 95% confidence interval for the CMLE, by parameter and sample size n.

n	Parameter	Bias	RMSE	Coverage
100	α_1	−0.01341	0.1565	0.874
	β_1	0.00003	0.0289	0.914
	β_2	−0.00031	0.0294	0.936
	φ_1	−0.03504	0.0833	0.929
	θ_1	0.01992	0.1161	0.914
	α_2	0.08254	0.1701	0.892
	γ_1	0.02834	0.2302	0.923
250	α_1	−0.00876	0.0920	0.918
	β_1	0.00110	0.0179	0.942
	β_2	−0.00024	0.0176	0.947
	φ_1	−0.01639	0.0441	0.948
	θ_1	0.01010	0.0692	0.940
	α_2	0.03537	0.0971	0.926
	γ_1	0.00654	0.1269	0.938
500	α_1	0.00229	0.0648	0.932
	β_1	−0.00053	0.0116	0.952
	β_2	0.00045	0.0119	0.945
	φ_1	−0.00980	0.0297	0.945
	θ_1	0.00266	0.0464	0.946
	α_2	0.01639	0.0666	0.929
	γ_1	0.00069	0.0911	0.944
1000	α_1	0.00331	0.0454	0.950
	β_1	−0.00015	0.0087	0.949
	β_2	0.00034	0.0087	0.950
	φ_1	−0.00597	0.0216	0.946
	θ_1	0.00006	0.0329	0.945
	α_2	0.01078	0.0465	0.936
	γ_1	0.00038	0.0640	0.945

Table 2. Experiment S2: Empirical rejection rates of the Wald and LR tests of $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, at the nominal level $\alpha = 0.05$, across sample sizes n and true precision coefficient values γ_1. The row $\gamma_1 = 0$ reports the empirical size; rows $\gamma_1 > 0$ report the empirical power.

n	γ_1	Wald Rate	LR Rate
100	0.0	0.092	0.066
	0.2	0.212	0.182
	0.4	0.564	0.518
	0.6	0.802	0.774
	0.8	0.966	0.958
	1.0	0.994	0.982
250	0.0	0.078	0.068
	0.2	0.328	0.326
	0.4	0.874	0.850
	0.6	0.996	0.994
	0.8	1.000	1.000
	1.0	1.000	1.000
500	0.0	0.050	0.038
	0.2	0.614	0.600
	0.4	0.994	0.994
	0.6	1.000	1.000
	0.8	1.000	1.000
	1.0	1.000	1.000

Table 3. Experiment S3: Structural identification accuracy and average maximized log-likelihood, by information criterion and sample size n. "Recovery" columns give the proportion of replications in which the true mean covariate set, precision covariate set, and ARMA order, respectively, were exactly recovered; "FP"/"FN" give the average false positive/false negative covariate counts; the final three columns give the average maximized log-likelihood for the proposed model, Benchmark 1 (constant precision), and Benchmark 2 (lagged variance proxy).

n	Criter.	Recovery			Mean covariates		Precision covariates		Average Log-Likelihood		
		Mean	Prec.	ARMA	FP	FN	FP	FN	Proposed	Bench.1	Bench.2
100	AIC	0.819	0.797	0.612	0.177	0.004	0.000	0.203	184.9	181.6	182.3
	BIC	0.942	0.574	0.386	0.054	0.004	0.000	0.426	183.8	181.6	182.3
250	AIC	0.821	0.961	0.814	0.161	0.023	0.000	0.039	458.7	454.1	454.8
	BIC	0.957	0.911	0.691	0.023	0.023	0.000	0.089	457.9	454.1	454.8
500	AIC	0.826	0.961	0.876	0.143	0.043	0.000	0.039	908.8	908.1	908.6
	BIC	0.941	0.939	0.866	0.017	0.043	0.000	0.061	908.4	908.1	908.6

Table 4. Experiment S4a: Sensitivity of the precision-link selection procedure to the true (log) link. "Mean Values" reports the average BIC across replications for each candidate link, and "Selection Rate" reports the proportion of replications in which each link was selected as minimizing the BIC.

Link function	Average BIC	Selection rate
Log (true)	−1067.47	0.676
Identity	−1065.71	0.142
Square root	−1066.97	0.182

Table 5. Experiment S4b: Bias and RMSE of the mean submodel slope $\hat{\beta}_1$ under the correctly specified (full) model and under a misspecified model that omits the true precision covariate Z_1 (constant precision).

Metric	Correctly specified model	Precision omitted
Bias	−0.00096	−0.00126
RMSE	0.0153	0.0167

Table 6. Experiment S5: Bias and RMSE of the CMLE at $n = 300$ under Gaussian versus heavy-tailed (t_5) exogenous covariate innovations, matched on marginal variance.

Parameter	Normal		t_5	
	Bias	RMSE	Bias	RMSE
α_1	0.00284	0.0872	−0.00326	0.0837
β_1	0.00017	0.0153	0.00065	0.0155
β_2	−0.00078	0.0158	−0.00113	0.0160
φ_1	−0.01400	0.0404	−0.01401	0.0389
θ_1	0.00814	0.0663	0.00691	0.0628
α_2	0.02194	0.0863	0.02637	0.0866
γ_1	0.00044	0.1137	0.00355	0.1231

Table 7. Experiment S6: Average absolute out-of-sample forecast error by horizon h under a simulated structural break, for the proposed model, Benchmark 1 (constant precision), and Benchmark 2 (lagged variance proxy). $R = 300$.

Horizon	Proposed	Benchmark1	Benchmark2
h1	0.03079	0.03091	0.03085
h2	0.04685	0.04692	0.04681
h3	0.05818	0.05835	0.05827
h4	0.06133	0.06152	0.06146
h5	0.06572	0.06584	0.06562
h6	0.06532	0.06551	0.06526

Table 8. Experiment S7: Out-of-sample point-forecast RMSE by horizon h, comparing the AIC-selected and BIC-selected specifications against Benchmark 1 (constant precision) and Benchmark 2 (lagged variance proxy), all re-selected on the training segment. $R = 300$.

Horizon	AIC Selected	BIC Selected	Benchmark1	Benchmark2
h1	0.04129	0.04156	0.05570	0.06772
h2	0.06151	0.06168	0.09777	0.07986
h3	0.07283	0.07305	0.08784	0.09271
h4	0.08226	0.08247	0.10537	0.09816
h5	0.08679	0.08701	0.10877	0.10616
h6	0.09202	0.09256	0.11310	0.11203

Table 9. Experiment S7: Out-of-sample density-forecast accuracy, measured by the CRPS, by horizon h, comparing the AIC-selected and BIC-selected specifications against Benchmark 1 (constant precision) and Benchmark 2 (lagged variance proxy). $R = 300$

Horizon	AIC Selected	BIC Selected	Benchmark1	Benchmark2
h1	0.02316	0.02325	0.02672	0.02736
h2	0.03542	0.03555	0.04337	0.03921
h3	0.04299	0.04311	0.04718	0.04673
h4	0.04792	0.04803	0.05382	0.05122
h5	0.05103	0.05129	0.05772	0.05606
h6	0.05435	0.05472	0.06047	0.05945

Table 10. Experiment S8: Bias and RMSE of the mean submodel slope $\hat{\beta}$ under the full (correctly specified) model versus a misspecified constant precision model, across the true precision effect γ, the mean/precision covariate correlation ρ, and the sample size n ($R = 500$ replications per cell).

			Correctly specified			Misspecified		
n	ρ	γ	Bias	RMSE	Conv. Rate	Bias	RMSE	Conv. Rate
	0.0	0.50	0.000374	0.01329	100%	0.000312	0.01415	100%
	0.0	0.75	−0.000582	0.01373	100%	−0.000507	0.01458	100%
	0.0	1.00	0.000849	0.01328	100%	0.001249	0.01533	100%
	0.0	1.25	0.000013	0.01268	100%	−0.000424	0.01733	100%
250	0.3	0.50	0.000960	0.01377	100%	0.002368	0.01461	100%
	0.3	0.75	0.000174	0.01306	100%	0.002163	0.01484	100%
	0.3	1.00	−0.001349	0.01241	100%	0.001190	0.01606	100%
	0.3	1.25	−0.000170	0.01228	100%	0.003483	0.01704	100%
	0.6	0.50	0.000118	0.01385	100%	0.002311	0.01502	100%
	0.6	0.75	−0.000798	0.01267	100%	0.002970	0.01518	100%
	0.6	1.00	−0.000680	0.01215	100%	0.004413	0.01625	100%
	0.6	1.25	−0.000491	0.01203	100%	0.005263	0.01794	100%
	0.0	0.50	0.000500	0.00932	100%	0.000561	0.00993	100%
	0.0	0.75	−0.000059	0.00892	100%	0.000275	0.01013	100%
	0.0	1.00	−0.000634	0.00892	100%	−0.000843	0.01105	100%
	0.0	1.25	−0.000439	0.00892	100%	−0.000181	0.01150	100%
500	0.3	0.50	0.000434	0.00950	100%	0.001711	0.01027	100%
	0.3	0.75	−0.000282	0.00932	100%	0.001637	0.01025	100%
	0.3	1.00	−0.001074	0.00912	100%	0.000967	0.01148	100%
	0.6	0.50	0.000141	0.00934	100%	0.002455	0.00983	100%
	0.6	0.75	0.000021	0.00921	100%	0.003133	0.01140	100%
	0.6	1.00	0.000277	0.00872	100%	0.005310	0.01253	100%

8. Real data application

8.1. Malaysian unemployment rate

8.1.1. Data and preliminary analysis

The primary variable of interest is the monthly Malaysian unemployment rate (*Unemp*), covering January 2015 to November 2025 (131 observations), sourced from the Department of Statistics Malaysia (DOSM) and Bank Negara Malaysia (BNM) via the official open-data portal (`https://data.gov.my/data-catalogue`). Expressed as a proportion, the series is naturally bounded on $(0, 1)$ and is directly amenable to the βARMA framework. Descriptive statistics are reported in Table 11: The unemployment rate averaged 3.57%, with a right skew of 1.46 and excess kurtosis of 4.01, confirming the distributional inadequacy of Gaussian methods.

To capture both the level and volatility dynamics, we incorporate five monthly macroeconomic indicators as exogenous covariates:

- CPI: Consumer price index inflation year-on-year (%);
- IPI: Industrial production index (base year 2015=100);
- TB: Trade balance (RM million);
- LR: Average lending rate (%);
- LFR: Labour force participation rate (%).

All continuous regressors are standardized prior to estimation. The correlation matrix (Figure 3) shows pairwise correlations ranging from −0.33 to 0.52, and variance inflation factors (VIFs) are all below 2 (CPI: 1.16; IPI: 1.77; TB: 1.89; LR: 1.45; LFR: 1.14), ruling out multicollinearity as a concern.

Table 11. Descriptive statistics for the monthly Malaysian unemployment rate and macroeconomic covariates (January 2015–November 2025; $n = 131$).

Variable	Min	Max	Mean	Standard deviation	Skew	Kurt
Unemp	0.029	0.053	0.0357	0.0054	1.4611	4.0094
CPI	−2.70	1.30	0.145	0.4581	−1.9653	14.507
IPI	77.052	143.99	117.175	12.526	−0.0810	2.7683
TB	−4.465	31.839	13.408	7.021	0.4719	2.8006
LR	3.277	5.491	4.548	0.626	−0.7335	2.5916
LFR	−0.014	0.008	0.0015	0.0025	−2.2704	14.869

Visual inspection of the unemployment series reveals a pronounced, temporary disturbance coincident with the onset of the COVID-19 pandemic and Malaysia's first movement control order (18 March 2020) [32]: The rate rises sharply from its pre-pandemic average of 3.3%–3.4% to a sample maximum of 5.3%, before declining gradually back toward its pre-pandemic level over the following two years. Because this disturbance spans a substantial fraction of the 131-observation sample rather than resolving within a handful of periods, its start and end dates are identified formally using the multiple structural change test of [33,34], as implemented in the `strucchange` package [35]. With the standard 15% trimming parameter, the minimum detectable segment length for $n = 131$ is approximately 20 observations; since the disturbance in question persists for 24 observations, it comfortably satisfies the asymptotic requirements of the test, and the Bai–Perron procedure identifies two breaks in the mean of the series: The first at $t = 63$ (March 2020), coincident with the initial nationwide restrictions, and the second at $t = 87$ (March 2022), coincident with Malaysia's gradual transition toward the endemic phase and the associated relaxation of movement restrictions [36].

To assess whether the Bai–Perron end date corresponds to a genuine resolution of the shock rather than an artifact, we cross-validate it against a temporary change (TC) intervention model [37, 38], in which the effect on the mean decays geometrically at a rate λ. Minimizing the BIC over a grid of candidate values yields $\hat{\lambda} \approx 0.95$, implying that approximately 31% of the initial shock remains at the March 2022 boundary. This gradual decay pattern supports the use of a finite-duration structural break rather than a single-period outlier. The resulting temporary COVID-19 dummy,

$$D_t^{\mathrm{COVID}} = \mathbb{1}\{63 \le t \le 87\}, \tag{8.1}$$

is equal to one from March 2020 through to March 2022 and zero elsewhere (Figure 1), which is included as an additional covariate in the mean submodel to prevent the autoregressive and moving-average components from absorbing an atypical, exogenously timed event.

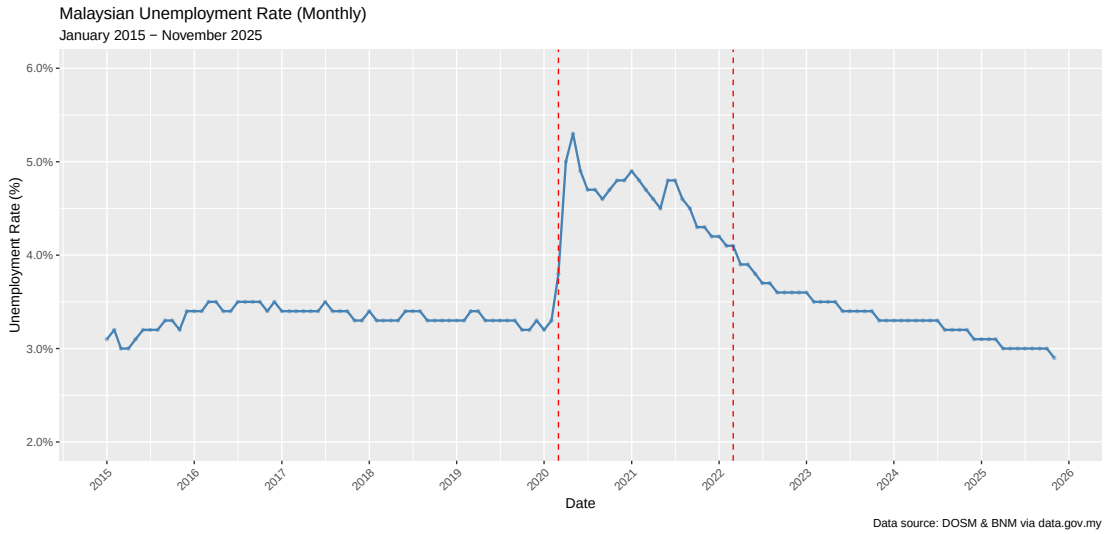

Figure 1. Monthly Malaysian unemployment rate (January 2015–November 2025) with dashed lines indicating the start and end of the break window.

Serial dependence. Figure 2 presents the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the unemployment rate. The very slow hyperbolic decay in the ACF is characteristic of a highly persistent process, while the PACF cuts off sharply after lag 1, suggesting that a low-order autoregressive structure is appropriate.

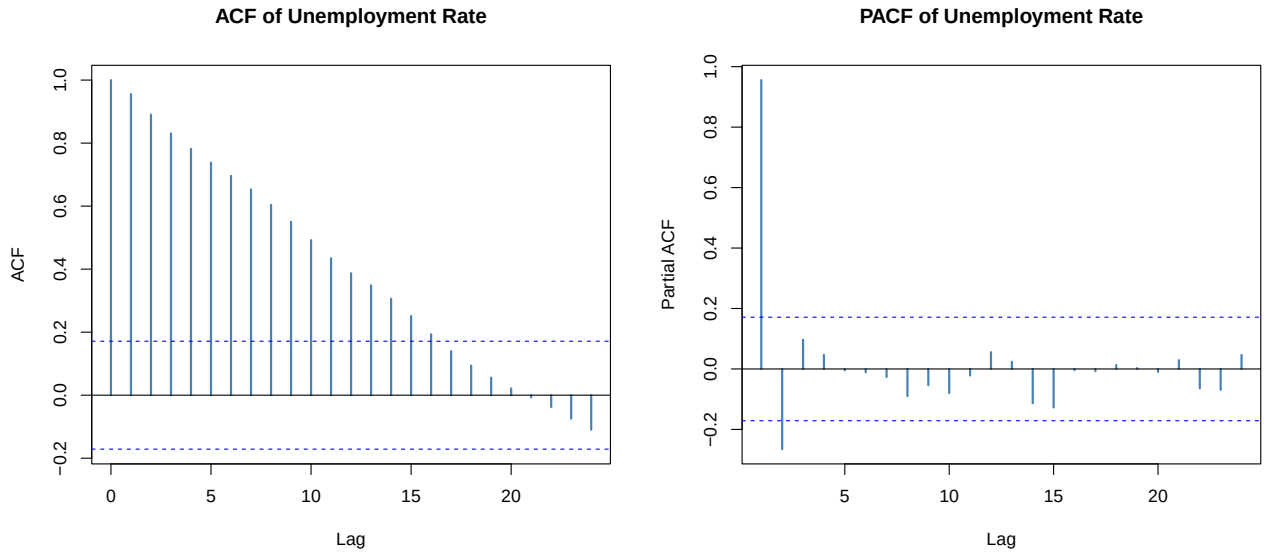

Figure 2. Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the monthly Malaysian unemployment rate. The slow decay in the ACF confirms strong serial dependence; the PACF suggests a low-order autoregressive structure.

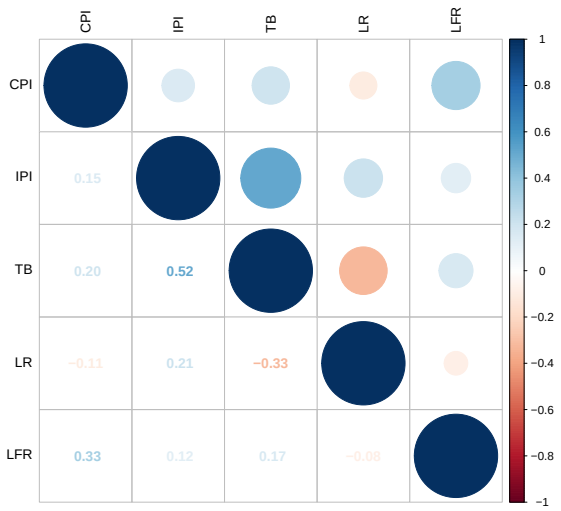

Figure 3. Correlation matrix of the macroeconomic covariates and the unemployment rate. Pairwise correlations range from −0.33 to 0.52; all VIF values are below 2, confirming the absence of multicollinearity.

8.2. Model estimation and in-sample fit

We compare three model formulations, each sharing the logit link for the mean submodel and the log link for the precision submodel (the superiority of the log precision link over the identity and square-root alternatives is confirmed in Table 16). Covariate selection for all models was performed via the AIC best-subset search:

- *Proposed*: βARMA$(2, 1)$ with exogenous covariates in both the mean and precision submodels;
- *Benchmark 1*: βARMA$(1, 1)$ with constant precision ($\phi_t \equiv \phi$);
- *Benchmark 2*: βARMA$(1, 0)$ with precision driven solely by the lagged variance proxy $y_{t-1}(1 - y_{t-1})$, as in [2].

Goodness of fit. Table 12 compares the three specifications, each selected independently. The proposed model achieves the highest log-likelihood (765.30, against 709.96 and 723.46) and the lowest AIC and BIC, by margins of roughly 74 to 107 points. Because the three models differ in their ARMA orders and mean covariates as well as in their precision submodels, this comparison reflects the combined effect of both submodels. Table 13 isolates the precision submodel by fixing the mean submodel at the proposed specification, a βARMA$(2, 1)$ with CPI, LFR, and the COVID-19 dummy, and varying only the precision specification. Under this common mean structure, the lagged variance proxy provides no improvement over constant precision (LR ≈ 0), whereas the three exogenous precision covariates raise the log-likelihood by about 70 points (LR = 141.04 on 3 degrees of freedom, $p < 0.001$). The gain in fit can therefore be attributed to the exogenous precision covariates rather than to differences in the mean dynamics.

Table 12. Goodness-of-fit comparison between the proposed model and two benchmarks for the monthly Malaysian unemployment rate data.

Model	AR(p)	MA(q)	Parameters	logLik	Avg. logLik	AIC	BIC	Δ AIC	Δ BIC
Proposed	2	1	11	765.30	5.84	−1508.60	−1476.97		
Benchmark1	1	1	9	709.96	5.42	−1401.93	−1376.05	106.67	100.92
Benchmark2	1	0	9	723.46	5.52	−1428.93	−1403.05	79.67	73.92

Table 13. Controlled comparison of precision specifications for the Malaysian unemployment rate: The mean submodel is held fixed and only the precision submodel varies.

						LR test against Benchmark 1		
	Precision submodel	k	logLik	AIC	BIC	LR	df	p-value
Benchmark 1	Constant precision	8	694.78	−1373.56	−1350.56	–	–	–
Benchmark 2	Lagged variance proxy	9	694.75	−1371.50	−1345.63	−0.06	1	1.000
Proposed	Exogenous covariates	11	765.30	−1508.60	−1476.97	141.04	3	<0.001

Note: All specifications share the mean submodel βARMA(2, 1) with CPI, LFR, and Dum; z_t = (IPI, LFR, Dum). LR tests $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ against Benchmark 1.

Table 14 reports the parameter estimates. Best-subset search selects an AR(2) mean dynamic for the proposed model, against AR(1) for both benchmarks, and a precision submodel containing IPI, LFR, and the COVID-19 dummy, each estimated at more than three times its standard error (s.e.). The dummy enters the precision submodel with a large negative coefficient (−1.5288, s.e. = 0.3162), implying that conditional precision fell by a factor of about $e^{1.53} \approx 4.6$ during the episode, whereas its effect on the mean is modest (0.0429 on the logit scale). The pandemic period therefore appears primarily as a dispersion event, which a constant precision specification cannot represent. The mean submodels also differ across specifications. IPI and the lending rate, which enter the mean submodel of both benchmarks, are not selected for the proposed model's mean, and IPI appears instead in its precision submodel; the CPI's coefficient is also attenuated in the proposed model (−0.0047, versus −0.0068 and −0.0121). This reallocation is consistent with the non-orthogonality mechanism of Section 3.4: When the precision submodel omits the relevant dispersion covariates, part of their influence can be absorbed by the mean equation. Because the three models also differ in their ARMA orders, however, these differences cannot be attributed to the precision specification alone. The standard errors in Table 14, and the Wald statistics below, are computed after best-subset selection on the same data and do not account for selection uncertainty; they are best read as descriptive measures of the strength of the selected effects. The fitted autoregressive and moving-average polynomials of the proposed model have smallest root moduli 1.045 and 1.041, respectively, so the fitted process is stationary and invertible, although the moving-average root lies close to the unit circle.

Table 14. Parameter estimates and standard errors for the proposed model and two benchmark models applied to the monthly Malaysian unemployment rate data.

		Proposed		Benchmark 1		Benchmark 2	
Submodel	Parameter	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Mean	Intercept	−3.4688	0.0571	−3.3068	0.0278	−3.3470	0.0281
	CPI	−0.0047	0.0017	−0.0068	0.0023	−0.0121	0.0023
	IPI			−0.0190	0.0047	−0.0223	0.0048
	TB						
	LR			−0.0816	0.0189	−0.0438	0.0209
	LFR	−0.0024	0.0011	−0.0035	0.0017	−0.0027	0.0015
	Dum	0.0429	0.0116	0.0593	0.0212	0.0911	0.0207
	AR(1)	0.1464	0.0305	0.8726	0.0454	0.9181	0.0344
	AR(2)	0.7765	0.0254				
	MA(1)	0.9606	0.0185	0.3241	0.0918		
Precision	Intercept	11.615	0.1390	10.382	0.1240	16.003	0.870
	IPI	0.4245	0.1290				
	LFR	0.5523	0.1321				
	Dum	−1.5288	0.3162				
	$y_{t-1}(1-y_{t-1})$					−157.09	24.98

Note: Empty cells indicate that the variable was not selected by the best-subset search. Standard errors are based on the inverse Fisher information matrix.

The joint Wald tests of Table 15 corroborate the parameter-level evidence above. For the proposed model, the block test of $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ against the three precision covariates yields $W = 60.52$ on 3 degrees of freedom ($p < 0.001$), rejecting the constant precision null on real data and mirroring the high power documented for this test in Experiment S2 (Table 2). Benchmark 2's single-covariate precision block is also significant ($W = 39.56$) but smaller, consistent with the fit comparison of Table 12. The mean covariate and AR/MA blocks are significant in all three models, the latter confirming the strong persistence documented in Figure 2.

Table 15. Joint Wald tests for coefficient blocks in the proposed model and two benchmarks, when applied to the Malaysian unemployment rate data. All tests use a χ^2 reference distribution.

Model	Block	Statistic	df	p-value
Proposed	Mean covariates (excl. intercept)	33.12	3	0.000
	AR/MA dynamics	4775.0	3	0.000
	Precision covariates (excl. intercept)	60.52	3	0.000
	All precision parameters	8340.1	4	0.000
Benchmark 1	Mean covariates (excl. intercept)	65.86	5	0.000
	AR/MA dynamics	474.33	2	0.000
	All precision parameters	7005.0	1	0.000
Benchmark 2	Mean covariates (excl. intercept)	108.39	5	0.000
	AR/MA dynamics	710.43	1	0.000
	Precision covariates (excl. intercept)	39.56	1	0.000
	All precision parameters	7325.9	2	0.000

Table 16. Sensitivity of model fit to the choice of the precision link function for the proposed βARMA model (logit mean link) applied to the monthly Malaysian unemployment rate data.

Precision link	logLik	AIC	BIC	Converged
Identity	530.8	−1039.6	−1007.9	Yes
Square root	712.2	−1402.4	−1370.8	Yes
Log	765.3	−1508.6	−1477.0	Yes

Unlike the narrow spread across precision links seen in the simulation study (Experiment S4), the choice is decisive here: The log link improves the log-likelihood by over 230 points relative to the identity link and over 50 points relative to the square-root link, consistent with the log link's multiplicative structure better accommodating the wide range spanned by the estimated ϕ_t over the sample. All three links converge without difficulty, so the preference is a matter of fit rather than estimability.

Model diagnostics. Figure 4 presents four diagnostic plots for the proposed βARMA model. The quantile residuals versus time panel shows no systematic patterns or outliers [39]. The observed versus fitted values scatter closely around the 45-degree line, indicating accurate in-sample prediction. The normal quantile–quantile (Q–Q) plot of quantile residuals is nearly linear, and the "ACF of residuals" plot reveals that all autocorrelation coefficients fall entirely within the statistical significance bands.

In addition to the graphical checks, we assess residual serial dependence using formal portmanteau tests. For the link-scale residuals, the Q_4 statistic is 14.9 ($p = 0.092$), indicating no evidence of remaining autocorrelation. Applying the Box–Ljung test [40] to the quantile residuals yields $\chi^2 = 10.295$ with 12 degrees of freedom ($p = 0.590$). Both tests fail to reject the null hypothesis of no residual dependence, confirming that the fitted βARMA structure captures the serial dynamics satisfactorily.

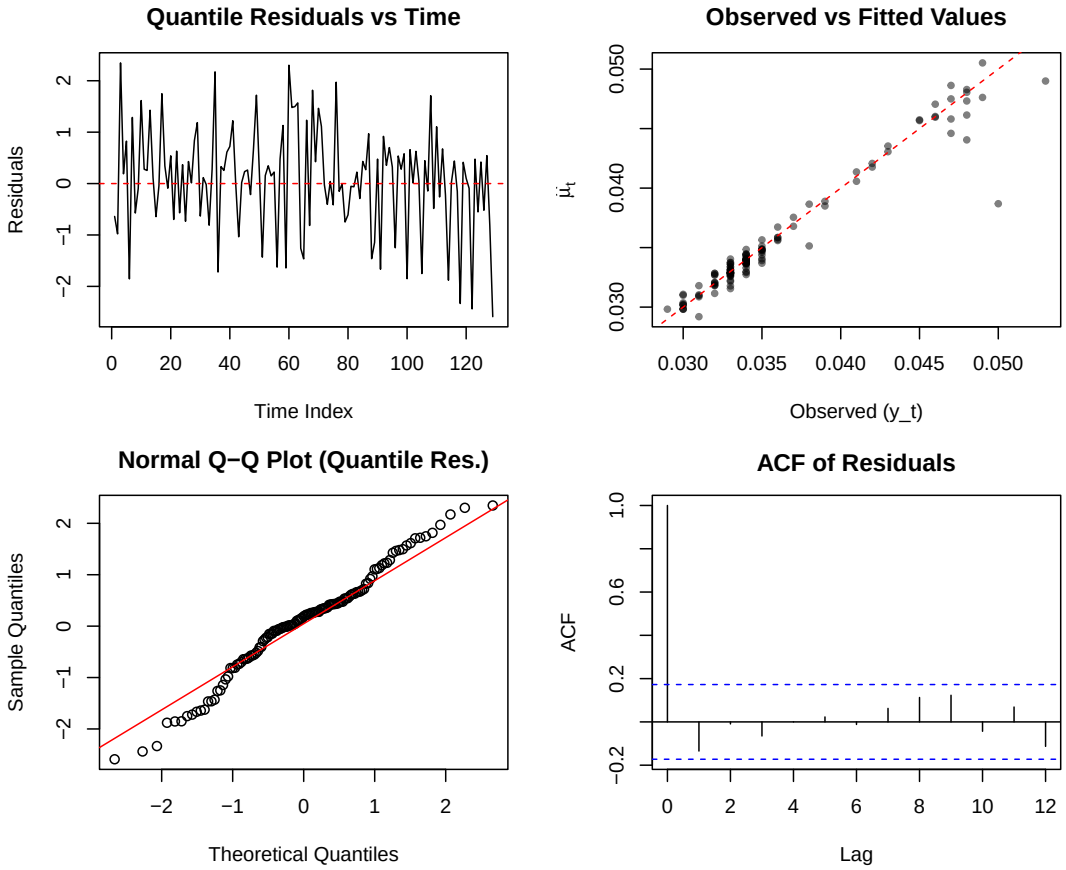

Figure 4. Diagnostic plots for the proposed βARMA model. Clockwise from top left: quantile residuals versus time index; observed versus fitted conditional means; ACF of the residuals; normal Q–Q plot of quantile residuals.

8.3. Out-of-sample forecasting evaluation

We assess predictive performance using a recursive rolling-window design producing h-step-ahead forecasts for $h \in \{1, \ldots, 6\}$ months, under three expanding training window sizes, $n_{\text{train}} \in \{105, 111, 118\}$ ($\approx$80%, 85%, and 90% of the full sample), following standard practice [19]. At each origin the model (and any covariate/order selection step) is re-estimated using only observations up to that point, and the standardization parameters are likewise computed within the training window and applied out of sample. Consistent with the convention set out in Section 6, the realized covariate values over the forecast window are used at the prediction step, and they play no role in estimation or selection at any origin.

Across all training window sizes and both information criteria, the proposed model dominates both benchmarks at every horizon on every point-forecast metric: RMSE , MAE, and MAPE (Tables 17, 31, and 30), with the margin widening as the training window lengthens: At $h = 1$, the RMSE reduction over Benchmark 1 grows from 18% at $n_{\text{train}} = 105$ to 43% at $n_{\text{train}} = 118$, and the reduction persists, and often strengthens, at longer horizons. This pattern, in which the gain does not shrink with the forecast horizon, contrasts with the narrowing gap documented under the simulated structural break of Experiment S6 (Table 7). The CRPS (Table 18) is also lower for the proposed model in every case,

although at $n_{\text{train}} = 105$ under AIC, its margin over Benchmark 2 is negligible at the longest horizons. The WIS (Table 32) gives a less uniform picture: The proposed model has the lowest WIS at $n_{\text{train}} = 118$ and, in most cases, at $n_{\text{train}} = 111$, but at $n_{\text{train}} = 105$, it is outperformed at several horizons, notably by both benchmarks under the AIC for $h \geq 2$. Its average predictive standard deviation (Table 33) is roughly one-third to three-fifths of the benchmarks', so its predictive distributions are considerably sharper. DM tests [29] under squared-error loss (Table 19) reject equal accuracy at the 5% level in 32 of the 36 comparisons. Under CRPS loss (Table 34), the differences are significant at every horizon for $n_{\text{train}} = 118$ and at short horizons for the two smaller windows, but mostly not beyond $h = 3$. These comparisons involve fully specified models, which differ in their mean as well as their precision submodels, so they measure the joint benefit of the selected specification rather than that of the precision submodel alone.

Table 17. RMSE comparison across models and forecast horizons for three training window sizes, evaluated under AIC- and BIC-selected models, for monthly Malaysian unemployment rate data.

Train		AIC			BIC		
Size	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
105	1	0.000578	0.000706	0.000659	0.000471	0.000766	0.000654
	2	0.000794	0.001029	0.000901	0.000670	0.000970	0.000882
	3	0.000958	0.001326	0.001095	0.000750	0.001190	0.001077
	4	0.001122	0.001608	0.001322	0.000832	0.001406	0.001318
	5	0.001264	0.001884	0.001535	0.000947	0.001613	0.001579
	6	0.001499	0.002112	0.001760	0.001162	0.001797	0.001785
111	1	0.000504	0.000732	0.000686	0.000486	0.000729	0.000686
	2	0.000736	0.001135	0.000983	0.000694	0.001159	0.000983
	3	0.000853	0.001483	0.001222	0.000810	0.001524	0.001222
	4	0.001025	0.001816	0.001498	0.000925	0.001870	0.001498
	5	0.001290	0.002170	0.001831	0.001073	0.002235	0.001831
	6	0.001544	0.002463	0.002093	0.001306	0.002540	0.002093
118	1	0.000489	0.000863	0.000829	0.000489	0.000845	0.000787
	2	0.000740	0.001356	0.001307	0.000740	0.001380	0.001271
	3	0.000873	0.001790	0.001712	0.000873	0.001831	0.001537
	4	0.001049	0.002229	0.002105	0.001049	0.002283	0.002006
	5	0.001331	0.002645	0.002452	0.001331	0.002711	0.002296
	6	0.001613	0.003037	0.002841	0.001613	0.003115	0.002452

Table 18. CRPS comparison across models and forecast horizons for three training window sizes, for monthly Malaysian unemployment rate data. Lower values indicate better probabilistic calibration.

Train		AIC			BIC		
Size	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
105	1	0.000320	0.000406	0.000379	0.000253	0.000437	0.000372
	2	0.000498	0.000586	0.000517	0.000407	0.000554	0.000507
	3	0.000631	0.000787	0.000669	0.000486	0.000683	0.000651
	4	0.000792	0.000996	0.000853	0.000580	0.000837	0.000837
	5	0.000963	0.001191	0.001010	0.000672	0.000971	0.001024
	6	0.001178	0.001371	0.001180	0.000876	0.001111	0.001187
111	1	0.000259	0.000414	0.000382	0.000249	0.000416	0.000382
	2	0.000439	0.000656	0.000579	0.000414	0.000673	0.000579
	3	0.000575	0.000916	0.000778	0.000553	0.000949	0.000778
	4	0.000769	0.001188	0.001019	0.000701	0.001234	0.001019
	5	0.000997	0.001485	0.001303	0.000816	0.001544	0.001303
	6	0.001208	0.001756	0.001547	0.001040	0.001827	0.001547
118	1	0.000258	0.000478	0.000471	0.000258	0.000471	0.000447
	2	0.000448	0.000827	0.000819	0.000448	0.000843	0.000770
	3	0.000586	0.001206	0.001160	0.000586	0.001241	0.001019
	4	0.000781	0.001630	0.001521	0.000781	0.001678	0.001374
	5	0.001043	0.002018	0.001864	0.001043	0.002081	0.001610
	6	0.001329	0.002424	0.002234	0.001329	0.002497	0.001829

Table 19. DM tests (squared error loss) of the proposed model against Benchmark 1 and Benchmark 2, for monthly Malaysian unemployment rate data. Negative mean differences indicate that the proposed model has lower squared forecast errors.

Train		Benchmark 1			Benchmark 2		
Size	Horizon	Mean Diff.	DM Stat.	p-value	Mean Diff.	DM Stat.	p-value
105	1	-3.65×10^{-7}	−2.36	0.009	-2.06×10^{-7}	−1.76	0.039
	2	-5.18×10^{-7}	−2.45	0.007	-3.54×10^{-7}	−1.81	0.036
	3	-8.52×10^{-7}	−2.25	0.012	-5.95×10^{-7}	−1.93	0.027
	4	-1.28×10^{-6}	−2.10	0.018	-1.04×10^{-6}	−1.97	0.024
	5	-1.70×10^{-6}	−1.77	0.039	-1.59×10^{-6}	−1.78	0.038
	6	-1.88×10^{-6}	−1.40	0.081	-1.84×10^{-6}	−1.44	0.075
111	1	-2.96×10^{-7}	−1.96	0.025	-2.34×10^{-7}	−1.57	0.059
	2	-8.96×10^{-7}	−2.60	0.005	-5.17×10^{-7}	−2.20	0.014
	3	-1.67×10^{-6}	−2.59	0.005	-8.34×10^{-7}	−2.28	0.011
	4	-2.64×10^{-6}	−2.19	0.014	-1.39×10^{-6}	−2.16	0.015
	5	-3.84×10^{-6}	−1.92	0.028	-2.20×10^{-6}	−2.20	0.014
	6	-4.75×10^{-6}	−1.71	0.044	-2.68×10^{-6}	−2.03	0.021
118	1	-4.75×10^{-7}	−2.35	0.009	-4.03×10^{-7}	−2.73	0.003
	2	-1.36×10^{-6}	−3.59	0.000	-9.61×10^{-7}	−2.61	0.004
	3	-2.59×10^{-6}	−5.10	0.000	-1.25×10^{-6}	−3.04	0.001
	4	-4.11×10^{-6}	−5.11	0.000	-2.31×10^{-6}	−2.39	0.008
	5	-5.58×10^{-6}	−3.43	0.000	-2.04×10^{-6}	−2.06	0.020
	6	-7.11×10^{-6}	−3.13	0.001	-2.76×10^{-6}	−1.17	0.121

Table 20 examines whether this advantage depends on access to the realized covariates. It reports the ex ante RMSE of the AIC-selected specifications for the expanding window from $n_{\text{train}} = 105$. With the covariates replaced by their forecasts, the proposed model remains the most accurate specification

at every horizon, with RMSE reductions of about 12–21% relative to Benchmark 2 and 25–37% relative to Benchmark 1. For this series, the forecasting gain therefore does not rely on knowledge of the future covariate values.

Table 21 assesses the calibration of the one-step-ahead predictive distributions. The Berkowitz test does not reject calibration for the proposed model ($p = 0.775$ under the AIC and 0.748 under the BIC), whereas it rejects for Benchmark 1 ($p \leq 0.006$) and is borderline for Benchmark 2 ($p = 0.089$ and 0.061). The maximum deviation of the mean PIT from uniformity is also smallest for the proposed model (0.132 and 0.138, against 0.18–0.20 for the benchmarks), although the associated bootstrap test is significant at the 5% level for all three models. Interval coverage reveals a complementary pattern. Benchmark 1's central intervals are too wide at every level, Benchmark 2's are close to nominal at the 90% and 95% levels, and the proposed model's are too narrow, covering 71–74%, 80–82%, and 86–88% of outcomes at nominal levels of 80%, 90%, and 95%. The sharper predictive distributions of the proposed model thus come at the cost of some under-coverage, which matters when the prediction intervals themselves are of primary interest. With 26 one-step origins, these diagnostics have limited power, and the coverage rates carry considerable sampling uncertainty.

Table 20. Ex ante out-of-sample point-forecast RMSE ($\times 10^3$) by horizon for the Malaysian unemployment rate, with the covariates replaced by their forecasts.

h	Proposed	Benchmark 1	Benchmark 2
1	0.502	0.693	0.606
2	0.699	1.014	0.843
3	0.825	1.302	1.045
4	0.969	1.446	1.183
5	1.166	1.643	1.380
6	1.381	1.830	1.562

Note: AIC-selected specifications; expanding window from $n_{\text{train}} = 105$. Non-deterministic covariates are replaced by automatic ARIMA forecasts [20].

Table 21. Calibration of one-step-ahead predictive distributions for the Malaysian unemployment rate.

		Probability integral transform				Coverage		
Criteria	Model	Mean	$\sup_u \lvert\bar{F}(u) - u\rvert$	p (sup)	p (Berkowitz)	80%	90%	95%
AIC	Benchmark 1	0.414	0.202	0.033	0.005	93%	95%	98%
	Benchmark 2	0.392	0.182	0.031	0.089	87%	91%	94%
	Proposed	0.406	0.132	0.035	0.775	71%	80%	86%
BIC	Benchmark 1	0.417	0.195	0.034	0.006	91%	95%	98%
	Benchmark 2	0.389	0.195	0.022	0.061	87%	91%	94%
	Proposed	0.411	0.138	0.018	0.748	74%	82%	88%

Note: One-step-ahead ex post forecasts from $n_{\text{train}} = 105$. $\bar{F}$: Non-randomized PIT [26]; p-values from a parametric bootstrap with 2000 draws.

8.4. Industry capacity utilization rate

As a second real-world test, we analyze the US total industry capacity utilization rate (CU), a bounded index measuring the share of installed productive capacity in use, spanning January 1990 to May 2026 ($n = 437$), retrieved from the Federal Reserve Economic Data (FRED) database (`https://fred.stlouisfed.org`). As covariates, we use the IPI, the effective federal funds rate (FFR), the CPI, the monthly growth rate of industrial production (IPG), and the monthly IR.

Table 22 shows a mean utilization rate of 78.44% with a standard deviation of 3.61%; the negative skew (−0.64) and excess kurtosis (4.04) again indicate departures from normality that motivate the beta-based framework. Unlike the sustained unemployment disturbance, CU exhibits an abrupt, short-lived collapse: It falls to 64.08% in March 2020 and recovers within five months. As this disturbance spans under 1% of the sample, well below the minimum segment length required by the Bai–Perron test ($\approx$ 65 observations), we instead identify it via the additive outlier procedure of intervention analysis [37, 38], which locates the onset at March 2020 ($t = 363$) and the resolution at July 2020 ($t = 367$), yielding the dummy $D_t^{\mathrm{COVID}} = \mathbb{1}\{363 \le t \le 367\}$, included as a mean submodel covariate.

Table 22. Descriptive statistics for the monthly total industry capacity utilization rate and macroeconomic covariates (January 1990–May 2026; $n = 437$).

Variable	Min	Max	Mean	Standard deviation	Skew	Kurt
CU	0.6408	0.8499	0.7844	0.0361	−0.6448	4.0353
IPI	60.3346	104.1004	90.4597	12.8690	−1.1128	2.9279
FFR	0.0500	8.2900	2.8879	2.3311	0.2394	1.8003
CPI	127.50	333.979	211.6318	53.022	0.4191	2.3387
IPG	−0.1319	0.0655	0.0012	0.0100	−5.2496	80.2342
IR	−0.0177	0.0138	0.0022	0.0027	−0.7864	11.9889

The capacity utilization application ($n = 437$) yields model rankings consistent with those observed in the unemployment rate application Table 12. As shown in Table 23, the proposed specification improves the average log-likelihood per observation from 5.1 to 5.64 relative to both benchmarks and reduces the AIC by 467.9/465 points and the BIC by 455.7/461 points. The larger absolute AIC/BIC gaps here reflect the larger sample size, but the gain in the (sample-size-invariant) average log-likelihood confirms a comparably strong improvement in fit across both series despite their different dynamic character. Notably, Benchmark 2's lagged variance proxy offers almost no improvement over Benchmark 1 in this application (a log-likelihood gain of only 3.4 points, against 13.5 points in the first application), reinforcing, on a second dataset, the simulation-based conclusion of Experiment S3 that a single autoregressive-type precision proxy captures only part of the information available in genuine exogenous precision covariates. The controlled comparison of Table 24, which fixes the mean submodel at the proposed βARMA$(3, 2)$ specification with IPI, CPI, IPG, and the COVID-19 dummy, shows that this gain comes from the precision submodel: The exogenous precision covariates raise the log-likelihood by about 233 points relative to constant precision (LR = 465.28 on 3 degrees of freedom, $p < 0.001$), whereas the lagged variance proxy raises it by only 1.4 points (LR = 2.87, $p = 0.091$). In both applications, therefore, the lagged variance proxy captures little of the variation in

precision once the mean submodel is fixed, while the exogenous covariates capture a substantial part of it.

Table 23. Goodness-of-fit comparison between the proposed model and two benchmarks for the capacity utilization rate data.

Model	AR(p)	MA(q)	Parameters	logLik	Average logLik	AIC	BIC	Δ AIC	Δ BIC
Proposed	3	2	14	2464.1	5.64	−4900.1	−4843		
Benchmark1	3	3	11	2227.1	5.1	−4432.2	−4387.3	467.9	455.7
Benchmark2	3	3	13	2230.5	5.1	−4435.1	−4382	465	461

Table 24. Controlled comparison of precision specifications for the US capacity utilization rate: The mean submodel is held fixed and only the precision submodel varies.

						LR test against Benchmark 1		
	Precision submodel	k	logLik	AIC	BIC	LR	df	p-value
Benchmark 1	Constant precision	11	2231.46	−4440.91	−4396.03	–	–	–
Benchmark 2	Lagged variance proxy	12	2232.89	−4441.78	−4392.82	2.87	1	0.091
Proposed	Exogenous covariates	14	2464.1	−4900.1	−4843	465.28	3	<0.001

Note: All specifications share the mean submodel βARMA(3, 2) with IPI, CPI, IPG and Dum; z_t = (CPI, IR, Dum). LR tests $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ against Benchmark 1.

Table 25 reports the resulting estimates. The mean submodel selects a βARMA(3, 2) structure with strong, alternating-sign AR/MA coefficients, consistent with the oscillatory short-run dynamics typical of this series; among the exogenous mean covariates, only IPI is significant (0.5467, s.e. = 0.0040), while CPI, IPG, and the COVID-19 dummy are retained by the best-subset search but are not individually significant once persistence and the IPI level are accounted for. The precision submodel shows a sharper contrast: CPI is precisely estimated and large (1.3861, s.e. = 0.0688), IR is not significant (0.0196, s.e. = 0.0690), and the COVID-19 dummy again enters with a large negative coefficient (−4.4942, s.e. = 0.6436), an order of magnitude larger than its mean submodel counterpart – confirming, as in the first application, that this disturbance was primarily a dispersion event that a constant precision model cannot represent.

Table 25. Parameter estimates and standard errors for the proposed model applied to the capacity utilization rate data.

Panel A: Mean submodel										
	Intercept	IPI	CPI	IPG	Dum	AR(1)	AR(2)	AR(3)	MA(1)	MA(2)
Estimate	2.5803	0.5467	0.0079	−0.0001	−0.0022	0.9963	0.9023	−0.8984	0.3072	−0.5942
Std. Error	3.7957	0.0040	0.0101	0.0002	0.0023	0.0384	0.0693	0.0361	0.0515	0.0419

Panel B: Precision submodel				
	Intercept	CPI	IR	Dum
Estimate	12.6790	1.3861	0.0196	−4.4942
Std. Error	0.0683	0.0688	0.0690	0.6436

Table 26. Joint Wald tests for coefficient blocks in the proposed model for the capacity utilization rate data. All tests use a χ^2 reference distribution.

Model	Block	Statistic	df	p-value
Proposed	Mean covariates (excl. intercept)	56283.88	4	0.000
	AR/MA dynamics	6550565.77	5	0.000
	Precision covariates (excl. intercept)	436.38	3	0.000
	All precision parameters	35118.13	4	0.000

The joint test of $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ for the precision covariate block is overwhelmingly rejected ($W = 436.38$ on 3 degrees of freedom, $p < 0.001$) as reported in Table 26 and consistent with the LR statistic of Table 24. That this joint rejection coexists with a non significant IR coefficient illustrates a general point: The relevant test is the joint test on the full precision covariate vector, since individual covariates may proxy for overlapping aspects of the same volatility regime. The very large mean covariate and AR/MA block statistics reflect both strong persistence and the large sample size, and should be read as confirmation that the selected dynamic structure is not spurious, rather than as a measure of the effect size.

The out-of-sample results for the capacity utilization rate (Tables 27–29), obtained under the same conditional forecasting convention as Section 8.3, sharpen the forecasting evidence considerably. Where the RMSE advantage in the first application ranged from roughly 12% to 53%, here, the proposed model's RMSE is smaller than the benchmarks' by a factor of roughly seven to eight at $h = 1$ and four to five at $h = 6$; MAE and MAPE (Table 27) show the same order-of-magnitude gap, with the proposed model's MAPE below 0.31% at every horizon versus over 1% for both benchmarks by $h = 6$. The probabilistic metrics (Table 28) mirror this pattern, and the much larger PredSD values for both benchmarks confirm that their inferior scores reflect genuinely wider, less informative predictive distributions. The DM tests (Table 29) confirm these differences are highly significant (CRPS statistics as extreme as -7.158 at $h = 1$, $p < 0.001$ at every horizon for both benchmarks). The larger relative gains here, compared with the first application, are consistent with the near-negligible in-sample improvement of Benchmark 2 documented above: When volatility is driven by identifiable, discrete shocks rather than by smoothly evolving own-series dynamics, a model that cannot access

exogenous conditions directly is at a substantially greater disadvantage.

Table 27. RMSE, MAE, and MAPE comparison across models and forecast horizons for a training window of size 370, evaluated under AIC- and BIC-selected models, for capacity utilization rate data.

		AIC			BIC		
Metric	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
RMSE	1	0.000503	0.003503	0.004032	0.000503	0.003426	0.004008
	2	0.000868	0.006688	0.006566	0.000868	0.006330	0.006566
	3	0.001326	0.008406	0.007860	0.001326	0.007257	0.007701
	4	0.001808	0.010160	0.008834	0.001808	0.007850	0.008611
	5	0.002298	0.012336	0.010262	0.002298	0.008800	0.009981
	6	0.002950	0.014459	0.011716	0.002950	0.009813	0.011480
MAE	1	0.000329	0.002550	0.002821	0.000329	0.002400	0.002841
	2	0.000623	0.005199	0.004778	0.000623	0.004480	0.004907
	3	0.000982	0.006845	0.005885	0.000982	0.005272	0.005844
	4	0.001385	0.008566	0.006714	0.001385	0.005792	0.006813
	5	0.001805	0.010734	0.007596	0.001805	0.006672	0.007841
	6	0.002349	0.012782	0.008576	0.002349	0.007623	0.009023
MAPE	1	0.043	0.333	0.368	0.043	0.314	0.371
	2	0.081	0.677	0.622	0.081	0.584	0.639
	3	0.128	0.889	0.765	0.128	0.686	0.761
	4	0.180	1.111	0.872	0.180	0.753	0.887
	5	0.234	1.392	0.985	0.234	0.867	1.019
	6	0.304	1.656	1.111	0.304	0.991	1.172

Table 28. CRPS, WIS, and PredSD comparison across models and forecast horizons for a training window of size 370, evaluated under AIC- and BIC-selected models, for capacity utilization rate data.

		AIC			BIC		
Metric	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
CRPS	1	0.000253	0.001884	0.002113	0.000253	0.001842	0.002113
	2	0.000487	0.003875	0.003530	0.000487	0.003378	0.003607
	3	0.000809	0.005311	0.004313	0.000809	0.003965	0.004331
	4	0.001187	0.006786	0.005066	0.001187	0.004369	0.005078
	5	0.001592	0.008829	0.005967	0.001592	0.005056	0.006067
	6	0.002122	0.010823	0.006889	0.002122	0.005884	0.007133
WIS	1	0.000153	0.001136	0.001267	0.000153	0.001119	0.001275
	2	0.000323	0.002310	0.002220	0.000323	0.002072	0.002189
	3	0.000604	0.003493	0.002678	0.000604	0.002509	0.002627
	4	0.000953	0.004774	0.003190	0.000953	0.002710	0.003131
	5	0.001341	0.006541	0.004078	0.001341	0.003216	0.003975
	6	0.001848	0.008430	0.004964	0.001848	0.003902	0.004960
PredSD	1	0.000412	0.004013	0.003367	0.000412	0.004097	0.003411
	2	0.000419	0.004000	0.004190	0.000419	0.004096	0.004209
	3	0.000426	0.003986	0.004202	0.000426	0.004095	0.004226
	4	0.000434	0.003973	0.004210	0.000434	0.004094	0.004259
	5	0.000442	0.003957	0.004207	0.000442	0.004092	0.004268
	6	0.000450	0.003943	0.004199	0.000450	0.004091	0.004267

Table 29. DM tests of the proposed model against Benchmark 1 and Benchmark 2, using capacity utilization rate data.

Training		Benchmark 1			Benchmark 2		
window size	Horizon	Mean Diff.	DM Stat.	p-value	Mean Diff.	DM Stat.	p-value
RMSE	1	−0.000011	−2.801	0.003	−0.000016	−3.306	0.000
	2	−0.000039	−2.973	0.001	−0.000042	−3.941	0.000
	3	−0.000051	−3.085	0.001	−0.000058	−3.482	0.000
	4	−0.000058	−2.596	0.005	−0.000071	−3.022	0.001
	5	−0.000072	−2.548	0.005	−0.000094	−2.702	0.003
	6	−0.000088	−2.517	0.006	−0.000123	−2.458	0.007
CRPS	1	−0.001589	−7.158	0.000	−0.001860	−6.468	0.000
	2	−0.002891	−5.561	0.000	−0.003121	−6.586	0.000
	3	−0.003156	−5.352	0.000	−0.003522	−6.222	0.000
	4	−0.003182	−4.386	0.000	−0.003891	−5.511	0.000
	5	−0.003465	−3.986	0.000	−0.004475	−4.339	0.000
	6	−0.003762	−3.336	0.000	−0.005011	−3.381	0.000

The ex ante comparison in Table 35, for the AIC-selected specifications, qualifies these results substantially. Once the covariates must be forecast, the errors of all three models increase considerably and the differences between them narrow. The proposed model is the most accurate from $h = 2$ to $h = 6$, whereas Benchmark 1 is clearly less accurate at longer horizons (by about 23% at $h = 6$). For this series, the large ex post gains therefore reflect access to contemporaneous covariate information to a large extent, and the ex ante gain is modest.

The one-step calibration diagnostics in Table 36 qualify the ex post results further. Both PIT tests reject calibration for the proposed model ($p < 0.001$); its mean PIT of 0.377 indicates that the outcomes tend to fall in the lower part of the predictive distribution, and its central intervals under-cover (78%, 81%, and 88% at nominal levels of 80%, 90%, and 95%). Benchmark 2 is better calibrated by the uniformity criterion ($p = 0.262$ and 0.537), although the Berkowitz test rejects it for both benchmarks. For this series, the proposed model delivers much sharper predictive distributions, but they are not well calibrated, and its lower proper scores under realized covariates should not be read as evidence of better calibration.

Taken together, the two applications support a clear in-sample conclusion and a more qualified out-of-sample one. In both series, with the mean submodel held fixed, exogenous precision covariates improve the likelihood substantially, whereas a lagged variance proxy adds little (Tables 13 and 24); the joint Wald and LR tests reject constant precision. In both cases, the COVID-19 period appears mainly as a reduction in precision rather than as a shift in level. Out of sample, the proposed specification forecasts the unemployment rate more accurately than the benchmarks whether or not the covariates are known in advance, whereas for capacity utilization, most of its advantage depends on access to the realized covariates, and the ex ante gain from $h = 2$ onward is small. In both series, its predictive distributions are sharper than the benchmarks' but tend to under-cover. The reallocation of covariates between the mean and precision submodels in the unemployment application is consistent with the transmission mechanism documented in Experiments S4 and S8, although in real data, it cannot be

separated from differences in the selected mean dynamics.

Table 30. MAPE (%) comparison across models and forecast horizons for three training window sizes, for monthly Malaysian unemployment rate data.

Train		AIC			BIC		
Size	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
105	1	1.245	1.755	1.621	0.886	1.858	1.539
	2	1.906	2.663	2.328	1.498	2.564	2.175
	3	2.501	3.657	3.038	1.926	3.227	2.951
	4	3.137	4.526	3.760	2.334	3.902	3.668
	5	3.607	5.181	4.296	2.627	4.372	4.374
	6	4.352	5.815	4.864	3.364	4.863	4.905
111	1	0.983	1.754	1.603	0.920	1.764	1.603
	2	1.709	3.062	2.533	1.580	3.142	2.533
	3	2.303	4.308	3.566	2.201	4.468	3.566
	4	3.002	5.370	4.477	2.770	5.561	4.477
	5	3.785	6.484	5.561	3.159	6.737	5.561
	6	4.498	7.451	6.420	3.956	7.736	6.420
118	1	1.006	2.211	2.172	1.006	2.179	2.051
	2	1.780	4.024	3.903	1.780	4.110	3.637
	3	2.400	5.633	5.351	2.400	5.801	4.677
	4	3.132	7.227	6.728	3.132	7.430	6.037
	5	4.034	8.592	7.902	4.034	8.843	6.864
	6	5.040	10.008	9.188	5.040	10.289	7.600

Table 31. MAE comparison across models and forecast horizons, for monthly Malaysian unemployment rate data.

Train		AIC			BIC		
Size	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
105	1	0.000392	0.000547	0.000507	0.000276	0.000581	0.000480
	2	0.000597	0.000820	0.000723	0.000468	0.000798	0.000671
	3	0.000781	0.001124	0.000939	0.000600	0.000999	0.000908
	4	0.000976	0.001388	0.001160	0.000724	0.001204	0.001126
	5	0.001114	0.001581	0.001317	0.000815	0.001339	0.001338
	6	0.001339	0.001770	0.001486	0.001038	0.001484	0.001495
111	1	0.000303	0.000536	0.000493	0.000283	0.000540	0.000493
	2	0.000528	0.000933	0.000774	0.000487	0.000958	0.000774
	3	0.000712	0.001311	0.001090	0.000678	0.001360	0.001090
	4	0.000926	0.001629	0.001363	0.000852	0.001688	0.001363
	5	0.001163	0.001963	0.001690	0.000971	0.002041	0.001690
	6	0.001375	0.002251	0.001946	0.001207	0.002339	0.001946
118	1	0.000303	0.000666	0.000654	0.000303	0.000656	0.000618
	2	0.000537	0.001211	0.001175	0.000537	0.001237	0.001096
	3	0.000724	0.001694	0.001609	0.000724	0.001745	0.001410
	4	0.000943	0.002170	0.002018	0.000943	0.002231	0.001813
	5	0.001210	0.002570	0.002363	0.001210	0.002646	0.002052
	6	0.001506	0.002986	0.002739	0.001506	0.003069	0.002272

Table 32. WIS comparison across models and forecast horizons, for monthly Malaysian unemployment rate data.

Train		AIC			BIC		
Size	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
105	1	0.000226	0.000246	0.000228	0.000183	0.000258	0.000223
	2	0.000377	0.000321	0.000288	0.000298	0.000296	0.000291
	3	0.000470	0.000432	0.000383	0.000349	0.000377	0.000390
	4	0.000599	0.000581	0.000515	0.000417	0.000460	0.000514
	5	0.000778	0.000752	0.000661	0.000505	0.000564	0.000677
	6	0.000980	0.000910	0.000815	0.000689	0.000671	0.000824
111	1	0.000189	0.000248	0.000226	0.000181	0.000246	0.000226
	2	0.000331	0.000353	0.000329	0.000305	0.000358	0.000329
	3	0.000428	0.000498	0.000462	0.000400	0.000512	0.000462
	4	0.000592	0.000696	0.000629	0.000517	0.000718	0.000629
	5	0.000803	0.000945	0.000864	0.000634	0.000983	0.000864
	6	0.001014	0.001180	0.001084	0.000842	0.001231	0.001084
118	1	0.000184	0.000278	0.000272	0.000184	0.000271	0.000257
	2	0.000337	0.000433	0.000450	0.000337	0.000437	0.000457
	3	0.000438	0.000654	0.000656	0.000438	0.000669	0.000584
	4	0.000601	0.000985	0.000924	0.000601	0.001013	0.000861
	5	0.000849	0.001353	0.001262	0.000849	0.001402	0.001090
	6	0.001118	0.001752	0.001628	0.001118	0.001813	0.001298

Table 33. PredSD comparison across models and forecast horizons, for monthly Malaysian unemployment rate data. Smaller values indicate sharper predictive distributions.

Train		AIC			BIC		
Size	Horizon	Proposed	Benchmark 1	Benchmark 2	Proposed	Benchmark 1	Benchmark 2
105	1	0.000437	0.000999	0.000920	0.000368	0.001042	0.000936
	2	0.000445	0.001004	0.000718	0.000369	0.001046	0.000730
	3	0.000426	0.001008	0.000729	0.000360	0.001049	0.000740
	4	0.000437	0.001011	0.000738	0.000361	0.001052	0.000753
	5	0.000320	0.001014	0.000751	0.000312	0.001054	0.000764
	6	0.000317	0.001016	0.000759	0.000307	0.001056	0.000772
111	1	0.000319	0.000985	0.000918	0.000316	0.001003	0.000918
	2	0.000316	0.000990	0.000700	0.000313	0.001009	0.000700
	3	0.000313	0.000996	0.000713	0.000309	0.001014	0.000713
	4	0.000311	0.001001	0.000730	0.000307	0.001020	0.000730
	5	0.000308	0.001005	0.000745	0.000304	0.001024	0.000745
	6	0.000307	0.001009	0.000757	0.000303	0.001028	0.000757
118	1	0.000313	0.000968	0.000949	0.000313	0.000985	0.000931
	2	0.000315	0.000976	0.000909	0.000315	0.000994	0.000806
	3	0.000317	0.000983	0.000910	0.000317	0.001001	0.000799
	4	0.000320	0.000989	0.000908	0.000320	0.001008	0.000829
	5	0.000314	0.000994	0.000901	0.000314	0.001013	0.000811
	6	0.000317	0.000999	0.000900	0.000317	0.001018	0.000791

Table 34. DM tests (CRPS loss) comparing the proposed model against Benchmark 1 and Benchmark 2, for monthly Malaysian unemployment rate data.

Train		Benchmark 1			Benchmark 2		
Size	Horizon	Mean Diff.	DM Stat.	p-value	Mean Diff.	DM Stat.	p-value
105	1	-1.85×10^{-4}	-3.76	0.000	-1.20×10^{-4}	−3.07	0.001
	2	-1.60×10^{-4}	-1.98	0.024	-1.12×10^{-4}	−1.52	0.064
	3	-1.96×10^{-4}	-1.84	0.033	-1.64×10^{-4}	−1.52	0.065
	4	-2.56×10^{-4}	-1.69	0.045	-2.56×10^{-4}	−1.63	0.052
	5	-2.99×10^{-4}	-1.33	0.091	-3.52×10^{-4}	−1.39	0.082
	6	-2.34×10^{-4}	-0.86	0.194	-3.10×10^{-4}	−0.98	0.165
111	1	-1.68×10^{-4}	-3.49	0.000	-1.34×10^{-4}	−3.07	0.001
	2	-2.75×10^{-4}	-2.54	0.006	-1.81×10^{-4}	−2.34	0.010
	3	-3.96×10^{-4}	-2.03	0.021	-2.25×10^{-4}	−1.70	0.045
	4	-5.33×10^{-4}	-1.59	0.056	-3.18×10^{-4}	−1.48	0.070
	5	-7.27×10^{-4}	-1.37	0.086	-4.86×10^{-4}	−1.48	0.069
	6	-7.85×10^{-4}	-1.28	0.100	-5.06×10^{-4}	−1.43	0.076
118	1	-2.13×10^{-4}	-4.25	0.000	-1.98×10^{-4}	−4.88	0.000
	2	-3.95×10^{-4}	-3.79	0.000	-2.76×10^{-4}	−3.37	0.000
	3	-6.55×10^{-4}	-3.96	0.000	-3.53×10^{-4}	−3.32	0.000
	4	-8.97×10^{-4}	-3.21	0.001	-4.91×10^{-4}	−2.78	0.003
	5	-1.04×10^{-3}	-2.19	0.014	-3.97×10^{-4}	−3.77	0.000
	6	-1.17×10^{-3}	-1.82	0.034	-4.35×10^{-4}	−2.44	0.007

Table 35. Ex ante out-of-sample point-forecast RMSE ($\times 10^3$) by horizon for the U.S. capacity utilization rate, with the covariates replaced by their forecasts.

h	Proposed	Benchmark 1	Benchmark 2
1	6.873	6.475	6.245
2	7.233	7.734	7.430
3	7.881	9.139	8.234
4	8.819	10.205	8.888
5	9.609	12.104	10.210
6	10.446	13.495	10.911

Note: AIC-selected specifications; expanding window from $n_{\text{train}} = 370$. Non-deterministic covariates are replaced by automatic ARIMA forecasts [20].

Table 36. Calibration of one-step-ahead predictive distributions for the U.S. capacity utilization rate.

		Probability integral transform				Coverage		
Criteria	Model	Mean	$\sup_u \lvert\bar{F}(u) - u\rvert$	p (sup)	p (Berkowitz)	80%	90%	95%
AIC	Benchmark 1	0.433	0.169	0.037	<0.001	93%	94%	96%
	Benchmark 2	0.511	0.121	0.262	0.013	85%	93%	97%
	Proposed	0.377	0.245	<0.001	<0.001	78%	81%	88%
BIC	Benchmark 1	0.502	0.149	0.097	0.005	90%	94%	97%
	Benchmark 2	0.503	0.097	0.537	0.006	87%	93%	94%
	Proposed	0.377	0.245	<0.001	<0.001	78%	81%	88%

Note: One-step-ahead ex post forecasts from $n_{\text{train}} = 370$. $\bar{F}$: non-randomized PIT [26]; p-values from a parametric bootstrap with 2000 draws.

9. Conclusions

This paper proposed a dual exogenous βARMA framework in which exogenous covariates enter both the conditional mean and the precision submodel through independent, observation-driven recursions. Unlike existing dynamic precision extensions, which let ϕ_t evolve only through its own lagged values, a single lagged variance proxy, or deterministic seasonal terms [1, 2], the proposed specification allows the precision to be driven by the same class of external information used to model the mean. The theoretical contributions include deriving the conditional likelihood, recursive score equations, and analytic Fisher information matrix; establishing the asymptotic properties; and formulating hypothesis tests for precision covariates. Most consequentially, we showed that the mean and precision blocks need not be information-orthogonal and that precision submodel misspecification can render the mean submodel estimator inconsistent when it induces a nonzero mean score bias.

The Monte Carlo study supports these results. Bias and RMSE decline with the sample size and CI coverage approaches the nominal level; the tests for the precision covariates are mildly oversized in small samples but approach the nominal size and have high power in moderate samples; and information criterion selection recovers the true precision covariate in more than 90% of replications for $n \geq 250$. The BIC selects the true logarithmic precision link in about two-thirds of replications, so the link is only moderately well identified, although the misspecified links approximate the dispersion structure closely. Omitting a relevant precision covariate reduces the efficiency of the mean submodel slope estimator and, when the omitted covariate is correlated with the mean covariate, biases it, with the bias growing in both the strength of the omitted effect and the correlation.

In two applications with contrasting persistence and shock structure, the Malaysian unemployment rate and the US capacity utilization rate, exogenous precision covariates improve the likelihood substantially relative to constant precision when the mean submodel is held fixed, whereas a lagged variance proxy adds little. In both series, the COVID-19 period appears mainly as a fall in conditional precision. The forecasting evidence is more qualified. For the unemployment rate, the proposed specification is more accurate than the benchmarks whether or not the future covariates are known; for capacity utilization, its large advantage under realized covariates shrinks to a small gain at horizons of two months or more once the covariates must be forecast, and it is the least accurate specification

one month ahead. In both series its predictive distributions are sharper than those of the benchmarks but tend to under-cover. Exogenous information in the precision submodel is therefore valuable for describing the dispersion of bounded time series and can improve forecasts, but the size of the forecasting gain depends on the predictability of the covariates, and sharper predictive distributions do not by themselves guarantee better calibration.

Three limitations suggest directions for future work. The proposed framework assumes that the response is strictly within (0,1); zero one inflated outcomes, which are common in many practical settings, are not directly accommodated. Moreover, although the asymptotic theory permits correlated regressors, high collinearity can inflate variance estimates and complicate parameter identification; regularization or penalized estimation offers a promising avenue for such cases. Finally, multi-step predictive distributions are approximated by plug-in beta laws rather than by the exact mixtures, which may contribute to under-coverage at longer horizons.

Author contributions

Elsidieg I. Belhaj: Conceptualization, data curation, formal analysis, investigation, software, validation, visualization, writing – original draft; Majid Khan Majahar Ali: Supervision, methodology, validation, writing – review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors used AI tools exclusively for English language correction. AI tools were not used to generate scientific content, data analysis, mathematical derivations, numerical simulations, figures, or conclusions. The authors assume full responsibility for all content in the manuscript.

Conflict of interest

The authors declare no conflicts of interest in this paper.

References

1. F. M. Bayer, R. J. Cintra, F. Cribari-Neto, Beta seasonal autoregressive moving average models, *Commun. Stat.-Simul. Comput.*, **47** (2018), 1314–1337. https://doi.org/10.1080/00949655.2018.1491974
2. V. T. Scher, F. Cribari-Neto, F. M. Bayer, Generalized barma model for double bounded time series forecasting, *Int. J. Forecasting*, **40** (2024), 721–734. https://doi.org/10.1016/j.ijforecast.2023.05.005
3. G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time series analysis: Forecasting and control*, John Wiley & Sons, 2015.
4. A. V. Rocha, F. Cribari-Neto, Beta autoregressive moving average models, *Test*, **18** (2009), 529–545. https://doi.org/10.1007/s11749-008-0112-z

5. S. L. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, *J. Appl. Stat.*, **31** (2004), 799–815. https://doi.org/10.1080/0266476042000214501
6. V. T. Scher, F. Cribari-Neto, G. Pumi, F. M. Bayer, Goodness-of-fit tests for βARMA hydrological time series modeling, *Environmetrics*, **31** (2020), e2607. https://doi.org/10.1002/env.2607
7. A. B. Simas, W. Barreto-Souza, A. V. Rocha, Improved estimators for a general class of beta regression models, *Comput. Stat. Data Anal.*, **54** (2010), 348–366. https://doi.org/10.1016/j.csda.2009.08.017
8. R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.*, **16** (1995), 1190–1208. https://doi.org/10.1137/0916069
9. A. V. Rocha, F. Cribari-Neto, Erratum to: Beta autoregressive moving average models, *Test*, **26** (2017), 451–459. https://doi.org/10.1007/s11749-017-0528-4
10. P. McCullagh, *Generalized linear models*, Chapman and Hall, London, 1989.
11. F. Zhang, *The Schur complement and its applications*, Springer Science & Business Media, 2006.
12. P. J. Bickel, C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, et al., *Efficient and adaptive estimation for semiparametric models*, vol. 4, Johns Hopkins University Press Baltimore, 1993.
13. E. B. Andersen, Asymptotic properties of conditional maximum-likelihood estimators, *J. Roy. Stat. Soc. B*, **32** (1970), 283–301. https://doi.org/10.1111/j.2517-6161.1970.tb00842.x
14. Y. Pawitan, *In all likelihood: Statistical modelling and inference using likelihood*, Oxford University Press, 2001. https://doi.org/10.1093/oso/9780198507659.001.0001
15. P. J. Bickel, K. A. Doksum, *Mathematical statistics: Basic ideas and selected topics, volumes I-II package*, Chapman and Hall/CRC, 2015. https://doi.org/10.1201/9781315369266
16. G. Casella, R. Berger, *Statistical inference*, Chapman and Hall/CRC, 2024. https://doi.org/10.1201/9781003456285
17. H. Akaike, A new look at the statistical model identification, *IEEE T. Automat. Control*, **19** (1974), 716–723. https://doi.org/10.1109/TAC.1974.1100705
18. G. Schwarz, Estimating the dimension of a model, *Ann. Stat.*, **6** (1978), 461–464. https://doi.org/10.1214/aos/1176344136
19. R. J. Hyndman, G. Athanasopoulos, *Forecasting: Principles and practice*, OTexts, 2018.
20. R. J. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for R, *J. Stat. Softw.*, **27** (2008), 1–22. https://doi.org/10.18637/jss.v027.i03
21. D. Koutsandreas, E. Spiliotis, F. Petropoulos, V. Assimakopoulos, On the selection of forecasting accuracy measures, *J. Oper. Res. Soc.*, **73** (2022), 937–954. https://doi.org/10.1080/01605682.2021.1892464
22. T. Gneiting, F. Balabdaoui, A. E. Raftery, Probabilistic forecasts, calibration and sharpness, *J. Roy. Stat. Soc. B*, **69** (2007), 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x
23. J. Bracher, E. L. Ray, T. Gneiting, N. G. Reich, Evaluating epidemic forecasts in an interval format, *PLoS Comput. Biol.*, **17** (2021), e1008618. https://doi.org/10.1371/journal.pcbi.1008618

24. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian data analysis*, Chapman and Hall/CRC, 1995. https://doi.org/10.1201/9780429258411
25. T. Gneiting, A. E. Raftery, A. H. Westveld III, T. Goldman, Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation, *Mon. Weather Rev.*, **133** (2005), 1098–1118. https://doi.org/10.1175/MWR2904.1
26. C. Czado, T. Gneiting, L. Held, Predictive model assessment for count data, *Biometrics*, **65** (2009), 1254–1261. https://doi.org/10.1111/j.1541-0420.2009.01191.x
27. J. Berkowitz, Testing density forecasts, with applications to risk management, *J. Bus. Econ. Stat.*, **19** (2001), 465–474. https://doi.org/10.1198/07350010152596718
28. P. Good, *Permutation, parametric and bootstrap tests of hypotheses*, Springer, 2005.
29. F. X. Diebold, R. S. Mariano, Comparing predictive accuracy, *J. Bus. Econ. Stat.*, **20** (2002), 134–144. https://doi.org/10.1198/073500102753410444
30. D. Harvey, S. Leybourne, P. Newbold, Testing the equality of prediction mean squared errors, *Int. J. Forecasting*, **13** (1997), 281–291. https://doi.org/10.1016/S0169-2070(96)00719-4
31. D. I. Harvey, S. J. Leybourne, Y. Zu, Tests for equal forecast accuracy under heteroskedasticity, *J. Appl. Econ.*, **39** (2024), 850–869. https://doi.org/10.1002/jae.3050
32. K. H. D. Tang, Movement control as an effective measure against Covid-19 spread in Malaysia: An overview, *J. Public Health*, **30** (2022), 583–586. https://doi.org/10.1007/s10389-020-01316-w
33. J. Bai, P. Perron, Estimating and testing linear models with multiple structural changes, *Econometrica*, 47–78. https://doi.org/10.2307/2998540
34. J. Bai, P. Perron, Computation and analysis of multiple structural change models, *J. Appl. Econ.*, **18** (2003), 1–22. https://doi.org/10.1002/jae.659
35. A. Zeileis, F. Leisch, K. Hornik, C. Kleiber, B. Hansen, E. C. Merkle, et al., Package 'strucchange', *R package version*, 1–5.
36. S. M. Abd Gani, M. Alied, N. T. Huy, Pandemic to endemic: A successful transition in Malaysia, *J. Global Health*, **13** (2023), 03063. https://doi.org/10.7189/jogh.13.03063
37. G. E. Box, G. C. Tiao, Intervention analysis with applications to economic and environmental problems, *J. Am. Stat. Assoc.*, **70** (1975), 70–79. https://doi.org/10.1080/01621459.1975.10480264
38. C. Chen, L. M. Liu, Joint estimation of model parameters and outlier effects in time series, *J. Am. Stat. Assoc.*, **88** (1993), 284–297. https://doi.org/10.1080/01621459.1993.10594321
39. P. L. Espinheira, S. L. Ferrari, F. Cribari-Neto, Influence diagnostics in beta regression, *Comput. Stat. Data Anal.*, **52** (2008), 4417–4431. https://doi.org/10.1016/j.csda.2008.02.028
40. G. M. Ljung, G. E. Box, On a measure of lack of fit in time series models, *Biometrika*, **65** (1978), 297–303. https://doi.org/10.1093/biomet/65.2.297



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)