AIMS Mathematics, 11(7): 22598–22622.
DOI: 10.3934/math.2026913
Received: 29 April 2026
Revised: 05 July 2026
Accepted: 13 July 2026
Published: 27 July 2026

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

Research article

Poisson INGARCH models: Linear Bayesian estimation and causality testing

Zhongxiu Ma[1], **Zhanshou Chen**[1,2,*] **and Peiyan Qi**[3]

[1] School of Mathematics and Statistics, Qinghai Normal University, Xining, 810008, China
[2] The State Key Laboratory of Tibetan Intelligence, Xining, 810008, China
[3] Department of Applied Mathematics, Taiyuan University of Science and Technology, Taiyuan, 030024, China

* **Correspondence:** Email: chenzhanshou@126.com.

Abstract: We develop a linear Bayesian estimator (LBE) for the Poisson integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) model that fuses prior information with sample data, yielding a closed-form solution and thereby bypassing posterior integration. Under the mean squared error matrix (MSEM) criterion, we rigorously establish the theoretical superiority of the proposed estimator over conditional maximum likelihood estimation (CMLE). Exploiting the conditional mean structure of the Poisson INGARCH model, we construct a regression framework for testing the causal effects of exogenous covariates. We derive a least-squares test statistic and prove that it asymptotically follows a chi-square distribution under the null hypothesis of no Granger causality. Simulation studies with prior robustness checks demonstrate that the proposed estimator remains predominantly data-driven and significantly outperforms CMLE in small-to-moderate samples. Applying the method to Chicago crime data, we find that it provides crucial numerical stability when the likelihood geometry is unfavorable. Importantly, we underscore that beyond stabilizing computation, routine distributional diagnostics (e.g., negative binomial (NB) checks) are indispensable for addressing potential overdispersion and ensuring reliable causal conclusions.

Keywords: count time-series; Poisson INGARCH model; linear Bayes estimation; Granger causality
Mathematics Subject Classification: 62M10, 62F15, 62E20

1. Introduction

Current statistical literature demonstrates growing interest in non-Gaussian time series, particularly those comprising nonnegative integer values. Such count-valued series arise routinely in applied sciences, including daily hospital admissions, high-frequency transaction counts, and monthly road

traffic accidents. These data typically exhibit strong temporal dependence, overdispersion (variance exceeding the mean), and nonnegative autocorrelation, often alongside relevant covariates. The coexistence of discreteness, heteroscedasticity, and serial dependence limits the applicability of classical Gaussian time-series models and motivates specialized count-data frameworks.

Within this literature, integer-valued time-series models have become the dominant tool for capturing both the discrete support and dynamic heterogeneity of count processes. Among them, the integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) model introduced by Ferland et al. [17] has become particularly prominent. By specifying the conditional mean as a linear recursion of past observations and lagged conditional means, the INGARCH model accommodates short- and long-range dependence while preserving integer-valued dynamics. Subsequent extensions have substantially enriched the model class: Zhu [41] adopted the generalized Poisson distribution to handle over- and under-dispersion; Zhu [42] proposed zero-inflated and negative binomial (NB) variants; Davis and Liu [13] relaxed linearity to capture nonlinear dynamics; Cui and Zhu [4] developed a bivariate specification allowing negative cross-correlations; Weiß [38] provided a unified treatment of stationarity and multivariate extensions; Weiß and Zhu [39] introduced softplus link functions to improve interpretability; and Su et al. [35] developed diagnostic tools for influential observations. Recent systematic reviews by Liu et al. [28] have further consolidated these advances, documenting the rapid methodological expansion of INGARCH models across unbounded counts, bounded counts, Z-valued series, and multivariate settings.

The recursive structure of INGARCH models places them within the broader family of observation-driven point processes for event data. Closely related is the autoregressive conditional duration (ACD) framework of Engle and Russell [16], which employs an analogous recursive specification for positive-valued duration data and has become a cornerstone of financial market microstructure analysis. In continuous time, self-exciting Hawkes processes [2] provide an alternative intensity-based approach whose discrete-time approximation shares the conditional mean structure of INGARCH models, thereby offering a complementary perspective on count volatility dynamics. Alternative Poisson-GARCH formulations, such as the correlated bivariate Poisson jump model of Chan [7] for continuous returns, illustrate the diversity of count-based volatility specifications in financial applications. These parallel developments underscore that INGARCH models represent one important strand within a rich ecosystem of recursive point-process models, each tailored to specific data types and application contexts. The methodological breadth of this field has been systematically surveyed by Davis et al. [12].

Despite this broad methodological progress, advances in model specification have not been matched by commensurate improvements in estimation methodology, especially for settings demanding computational efficiency and finite-sample reliability. Parameter estimation for the Poisson INGARCH model remains methodologically constrained. Conditional maximum likelihood estimation (CMLE), though asymptotically efficient, suffers from convergence failures in finite samples and severely underestimated standard errors near the stationarity boundary ($\alpha + \beta < 1$). Markov chain Monte Carlo (MCMC) methods provide coherent probabilistic inference but impose prohibitive computational costs for recursive models and are ill-suited to high-dimensional or real-time applications. Crucially, these numerical fragilities are exacerbated in empirical scenarios involving weak identification or severe overdispersion, where the likelihood geometry becomes unfavorable.

To address this gap, we propose a linear Bayesian estimation (LBE) framework for the Poisson INGARCH model. Unlike full Bayesian inference, LBE does not target a posterior distribution or

propagate uncertainty through a probabilistic hierarchy. Instead, within the class of linear estimators, LBE minimizes Bayes risk to combine CMLE with prior information, yielding a closed-form solution that bypasses posterior integration. The resulting estimator operates as a linear shrinkage estimator under the mean squared error matrix (MSEM) criterion.

The theoretical foundations of LBE originate in Hartigan [24] and Goldstein [20], with subsequent developments addressing loss functions [25], hierarchical formulations [21], systematic selection criteria [31], and MSEM-based dominance under misspecification [36]. Recent work has extended LBE to extreme-value distributions [5], misrecorded count data [22], and balanced loss functions [3]. Nevertheless, these advances have not addressed count time-series models with recursive conditional mean structures.

Parallel developments in Bayesian inference for count time series have also emerged. Variational Bayes approximations offer deterministic alternatives to MCMC for INGARCH-type models [1,26,40]. Structured prior selection strategies improve regularized estimation for high-dimensional count processes [11, 37]. Bayesian causal discovery extends Granger-causality testing to fully probabilistic frameworks [10, 32, 33]. However, these approaches either retain considerable computational complexity or fail to deliver a closed-form estimator that integrates prior information directly with CMLE for the Poisson INGARCH model. To our knowledge, no existing study has developed LBE theory specifically for this setting.

Beyond estimation, a central inferential objective is testing the causal effect of exogenous covariates in the conditional mean equation, commonly framed within the nonlinear Granger causality testing framework [23]. Environmental criminology examines how weather conditions influence offending patterns [6, 18, 29, 30]. Corcoran and Zahnow [8] reviewed classical Granger methods and highlighted their limitations, while Chen and Lee [9] implemented Bayesian causality testing for bivariate INGARCH models using MCMC. Their approach, while theoretically rigorous, imposes prohibitive computational costs for large datasets and lacks closed-form asymptotic properties for uncertainty quantification. The proposed LBE framework addresses this gap by delivering a closed-form estimator that supports regression-based causality testing with valid asymptotic guarantees.

After sorting out the existing literature, obvious research gaps still exist: (1) Existing estimation methods for the Poisson INGARCH model are largely confined to CMLE and MCMC, lacking a closed-form Bayesian approach that integrates prior information with frequentist estimation to stabilize finite-sample performance; (2) linear Bayes paradigm, despite its computational advantages in other settings, has not been extended to count time-series models with recursive conditional mean structures, resulting in insufficient methodological tools for small-sample count data analysis; (3) causal testing frameworks for INGARCH models rely on either asymptotically efficient but numerically fragile CMLE or computationally burdensome MCMC, lacking a tractable alternative that provides robust inference with valid uncertainty quantification; and (4) specification strategies in existing Bayesian count time-series analyses are often either ad hoc or computationally intensive, lacking data-driven empirical Bayes approaches that maintain computational efficiency while ensuring prior robustness.

Aiming at the above research gaps, this paper develops a linear Bayesian estimation framework for the Poisson INGARCH model and embeds it into a regression-based causality testing procedure. The main contributions are as follows:

(i) We develop a closed-form LBE for the Poisson INGARCH(1,1) model, fusing CMLE with data-driven prior information. The estimator acts as a computationally efficient shrinkage estimator

under the Bayes risk framework, avoiding posterior integration and offering a practical alternative to full Bayesian MCMC.

(ii) We rigorously establish the MSEM superiority of LBE over CMLE, providing theoretical justification for the finite-sample gains observed in simulations.

(iii) We embed LBE into a regression-based causality testing framework and derive the asymptotic chi-square distribution of the test statistic under the null hypothesis, enabling formal inference on exogenous covariate effects.

(iv) Through extensive simulations including prior robustness analysis and finite-sample bias diagnostics, we demonstrate that LBE reduces bias and root mean squared error (RMSE) relative to CMLE across small to moderate sample sizes, with estimates remaining predominantly data-driven even under substantial prior strength variation.

(v) We apply the proposed method to Chicago crime data to examine the temperature–crime relationship, showing that LBE provides numerical stability when the likelihood geometry is unfavorable, while routine distributional diagnostics are advocated to address potential overdispersion.

The remainder of the paper is structured as follows. Section 2 introduces the Poisson INGARCH model, develops the LBE methodology, and establishes its theoretical properties. Section 3 constructs the causality test and derives its asymptotic distribution. Section 4 presents simulation evidence and empirical results. Section 5 concludes and outlines directions for future research.

2. LBE theory for Poisson INGARCH models

Let Y_t be a count series following the Poisson INGARCH(1,1) model,

$$Y_t \mid \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad \lambda_t = \omega + \alpha\lambda_{t-1} + \beta Y_{t-1}, \tag{2.1}$$

where $\upsilon = (\omega, \alpha, \beta)^T \in \Theta \subset \mathbb{R}^3_+$, $\omega > 0, \alpha, \beta \geq 0$, $\alpha + \beta < 1$, and $\mathcal{F}_{t-1} = \sigma_{(Y_{t-1}, \cdots, Y_0, \lambda_0)}$.

2.1. Expression of LBE

Within the Poisson INGARCH(1,1) model framework, Kang and Lee [27] rigorously established that the parameter estimator $\hat{\upsilon}_n$ is strongly consistent, i.e., $\hat{\upsilon}_n \xrightarrow{\text{a.s.}} \upsilon_0$, where υ_0 is the true parameter vector. Their analysis revealed the asymptotic normality of the CMLE: As the sample size $n \to \infty$, the standardized estimation error satisfies $\sqrt{n}(\hat{\upsilon}_n - \upsilon_0) \xrightarrow{d} \mathcal{N}(0, I(\upsilon_0)^{-1})$, where $I(\upsilon_0)$ is the Fisher information matrix.

Building upon the theoretical results of Ferland et al. [17] on strict stationarity and ergodicity for Poisson INGARCH processes, the model admits a unique, strictly stationary solution, whose moments of all orders are finite, i.e., $\mathbb{E}[Y_t^k] < \infty$, $\mathbb{E}[\lambda_t^k] < \infty$ for all $k \geq 1$. Under these regularity conditions, we introduce some key definitions.

Definition 2.1. *Let $\upsilon = (\omega, \alpha, \beta)^{\mathsf{T}}$ denote the parameter vector of the Poisson INGARCH(1,1) model. Suppose the prior distribution $\pi(\upsilon)$ is restricted to the family*

$$\xi = \{\pi(\upsilon) : \mathbb{E}_\pi \|\upsilon\|^2 < \infty\}. \tag{2.2}$$

Let $T = \hat{\upsilon}_{\mathrm{CMLE}}$ denote the CMLE computed from the observed sample. Define $\mathcal{L}$ as the class of all affine transformations of T, i.e.,

$$\mathcal{L} = \{\hat{\upsilon} = BT + b \mid B \in \mathbb{R}^{3\times 3},\ b \in \mathbb{R}^3\}. \tag{2.3}$$

The LBE of υ, denoted $\hat{\upsilon}_{\mathrm{LBE}}$, is defined as the unique element of $\mathcal{L}$ minimizing the Bayes risk

$$R(\hat{\upsilon}, \upsilon) = \min_{B,\,b} \mathbb{E}_{\upsilon,Y}[(\hat{\upsilon} - \upsilon)^{\mathsf{T}} D\,(\hat{\upsilon} - \upsilon)], \tag{2.4}$$

where $D \succ 0$ is a fixed positive definite weighting matrix and $\mathbb{E}_{\upsilon,Y}$ denotes expectation with respect to the joint distribution of (υ, Y).

Assumption 2.1. *The CMLE $T = \hat{\upsilon}_{\mathrm{CMLE}}$ is assumed to satisfy:*

(i) Conditional unbiasedness: $\mathbb{E}[T \mid \upsilon] = \upsilon$ for all admissible υ;
(ii) Homoskedastic conditional covariance: $\mathrm{Cov}(T \mid \upsilon) = W \succeq 0$, where W is positive semidefinite and invariant to υ.

Remark 2.1. *Assumption 2.1 imposes conditional unbiasedness and homoskedasticity on the CMLE. While these moment conditions are standard in the linear Bayes literature, they are not guaranteed to hold exactly for the Poisson INGARCH model in finite samples. Under standard regularity conditions [27], they hold asymptotically as $n \to \infty$. Theorem 2.1 establishes the exact optimality of the LBE under these idealized conditions; finite-sample properties are examined via simulation in Section 4.*

Remark 2.2. *Unlike fully Bayesian procedures, the LBE does not target the full posterior distribution via Bayes' theorem. Instead, it is a linear shrinkage estimator that minimizes the MSEM by fusing prior moments with sample information. This approach aligns with Stein-type shrinkage estimation [15, 34], delivers substantial computational gains over fully Bayesian methods such as MCMC, and retains robustness under mild model misspecification.*

Theorem 2.1. *Consider the Poisson INGARCH(1,1) model with prior $\pi(\upsilon)$ restricted to the family ξ in* (2.2)*, and let $\mathbb{E}_\pi\upsilon$ and $\mathrm{Cov}_\pi(\upsilon)$ denote the prior mean and covariance, respectively. Under Assumption 2.1 and squared-error loss, the LBE minimizing the Bayes risk is uniquely characterized by*

$$\hat{\upsilon}_{\mathrm{LBE}} = \mathrm{Cov}_\pi(\upsilon)\big[W + \mathrm{Cov}_\pi(\upsilon)\big]^{-1} T + W\big[W + \mathrm{Cov}_\pi(\upsilon)\big]^{-1}\mathbb{E}_\pi\upsilon, \tag{2.5}$$

where $W = \mathbb{E}_\pi\big[\mathrm{Cov}(T \mid \upsilon)\big]$ is the expected frequentist covariance of T.

Proof. By Definition 2.1, we restrict attention to estimators of the form $\hat{\upsilon} = BT + b$. The corresponding Bayes risk is

$$R(B, b) = \mathbb{E}_{\upsilon,Y}\left[(BT + b - \upsilon)^{\top} D (BT + b - \upsilon)\right].$$

To minimize $R(B, b)$, we exploit the linearity of expectation and the positive definiteness of $D \succ 0$ and differentiate with respect to b:

$$\frac{\partial R}{\partial b} = 2D\mathbb{E}_{\upsilon,Y}(BT + b - \upsilon) = 2D\left[B\mathbb{E}_{\upsilon,Y}(T) + b - \mathbb{E}_{\upsilon,Y}(\upsilon)\right].$$

Setting the derivative to zero yields the optimal bias term:

$$b = \mathbb{E}_{\upsilon,Y}(\upsilon) - B\mathbb{E}_{\upsilon,Y}(T). \tag{2.6}$$

By the law of total expectation, $\mathbb{E}_{\upsilon,Y}(\upsilon) = \mathbb{E}_\pi\upsilon$. Under Assumption 2.1(i), $\mathbb{E}_{\upsilon,Y}(T) = \mathbb{E}_\pi[\mathbb{E}(T \mid \upsilon)] = \mathbb{E}_\pi\upsilon$. Substituting into (2.6) gives

$$b = (I - B)\mathbb{E}_\pi\upsilon. \tag{2.7}$$

Plugging (2.7) into $\hat{\upsilon} = BT + b$ gives the centered form

$$\hat{\upsilon} - \upsilon = B(T - \mathbb{E}_\pi\upsilon) - (\upsilon - \mathbb{E}_\pi\upsilon).$$

The risk then simplifies to

$$\begin{aligned}
R(B) =& \mathbb{E}_{\upsilon,Y}\left\{[B(T - \mathbb{E}_\pi\upsilon) - (\upsilon - \mathbb{E}_\pi\upsilon)]^\top D\,[B(T - \mathbb{E}_\pi\upsilon) - (\upsilon - \mathbb{E}_\pi\upsilon)]\right\} \\
=& \mathrm{tr}\mathbb{E}_{\upsilon,Y}\left\{D\,[B(T - \mathbb{E}_\pi\upsilon) - (\upsilon - \mathbb{E}_\pi\upsilon)]\,[B(T - \mathbb{E}_\pi\upsilon) - (\upsilon - \mathbb{E}_\pi\upsilon)]^\top\right\} \\
=& \mathrm{tr}\mathbb{E}_{\upsilon,Y}\left\{D\left[B(T - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top B^\top - B(T - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right.\right. \\
& \left.\left. -(\upsilon - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top B^\top + (\upsilon - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right]\right\} \\
=& \mathrm{tr}\left\{DB\mathbb{E}_{\upsilon,Y}\left[(T - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top\right]B^\top - D\mathbb{E}_{\upsilon,Y}\left[(\upsilon - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top\right]B^\top\right. \\
& \left. -DB\mathbb{E}_{\upsilon,Y}\left[(T - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right] + D\mathbb{E}_{\upsilon,Y}\left[(\upsilon - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right]\right\}.
\end{aligned} \tag{2.8}$$

We decompose each term using the law of total covariance and Assumption 2.1. First,

$$\begin{aligned}
\mathbb{E}_{\upsilon,Y}\left[(T - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top\right] &= \mathbb{E}_\pi[\mathrm{Cov}(T \mid \upsilon)] + \mathrm{Cov}_\pi[\mathbb{E}(T \mid \upsilon)] \\
&= W + \mathrm{Cov}_\pi(\upsilon).
\end{aligned}$$

Second, under Assumption 2.1(i),

$$\begin{aligned}
\mathbb{E}_{\upsilon,Y}\left[(\upsilon - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top\right] &= \mathbb{E}_\pi\left[\mathbb{E}\left[(\upsilon - \mathbb{E}_\pi\upsilon)(T - \mathbb{E}_\pi\upsilon)^\top \mid \upsilon\right]\right] \\
&= \mathbb{E}_\pi\left[(\upsilon - \mathbb{E}_\pi\upsilon) \cdot \mathbb{E}\left[(T - \mathbb{E}_\pi\upsilon)^\top \mid \upsilon\right]\right] \\
&= \mathbb{E}_\pi\left[(\upsilon - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right] \\
&= \mathrm{Cov}_\pi(\upsilon).
\end{aligned}$$

Analogously,

$$\mathbb{E}_{\upsilon,Y}\left[(T - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right] = \mathrm{Cov}_\pi(\upsilon), \quad \mathbb{E}_{\upsilon,Y}\left[(\upsilon - \mathbb{E}_\pi\upsilon)(\upsilon - \mathbb{E}_\pi\upsilon)^\top\right] = \mathrm{Cov}_\pi(\upsilon).$$

Substituting these expressions into the risk expansion yields

$$R(B) = \mathrm{tr}\left\{DB\,[W + \mathrm{Cov}_\pi(\upsilon)]\,B^\top - DB\mathrm{Cov}_\pi(\upsilon) - D\mathrm{Cov}_\pi(\upsilon)B^\top + D\mathrm{Cov}_\pi(\upsilon)\right\}. \tag{2.9}$$

Differentiating $R(B)$ with respect to B and using standard matrix calculus identities gives

$$\frac{\partial R}{\partial B} = 2DB\,[W + \mathrm{Cov}_\pi(\upsilon)] - 2D\mathrm{Cov}_\pi(\upsilon).$$

Setting the derivative to zero and noting that $D > 0$, we obtain

$$B\left[W + \mathrm{Cov}_\pi(\upsilon)\right] = \mathrm{Cov}_\pi(\upsilon). \tag{2.10}$$

Under the model's stationarity condition $\alpha + \beta < 1$, the Fisher information matrix $I(\upsilon)$ is positive definite, so $W = \mathbb{E}_\pi\left[n^{-1}I(\upsilon)^{-1}\right]$ is positive semidefinite. Combined with the positive definiteness of $\mathrm{Cov}_\pi(\upsilon)$ under a non-degenerate prior, $W + \mathrm{Cov}_\pi(\upsilon)$ is therefore invertible. The optimal weight matrix is thus

$$B = \mathrm{Cov}_\pi(\upsilon)\left[W + \mathrm{Cov}_\pi(\upsilon)\right]^{-1}. \tag{2.11}$$

Substituting this B into (2.7) recovers the LBE in (2.5).

□

Implementation of W and $\mathrm{Cov}_\pi(\upsilon)$. The quantity $W = \mathbb{E}_\pi[\mathrm{Cov}(T \mid \upsilon)]$ captures the expected sampling variability of the CMLE. It can be approximated by averaging the inverse observed Fisher information $n^{-1}I(\hat{\upsilon}_{\mathrm{CMLE}})^{-1}$ over the prior distribution. The prior covariance $\mathrm{Cov}_\pi(\upsilon)$ is elicited from $\pi(\upsilon)$. When a fully specified prior is unavailable, $\mathrm{Cov}_\pi(\upsilon)$ can be set to a diagonal matrix encoding plausible parameter ranges, with sensitivity assessed via simulation (Section 4).

2.2. Superiority of the linear Bayesian estimator

The MSEM is employed to evaluate the finite-sample performance of estimators within the proposed LBE framework. For any estimator $\hat{\upsilon}$ of the random parameter vector υ, the MSEM is defined as

$$\mathrm{MSEM}(\hat{\upsilon}) = \mathbb{E}[(\hat{\upsilon} - \upsilon)(\hat{\upsilon} - \upsilon)'], \tag{2.12}$$

where the expectation is taken with respect to the joint distribution of (υ, Y).

Let $T = \hat{\upsilon}_{\mathrm{CMLE}}$. Any linear estimator admits the representation $\hat{\upsilon} = BT + b$, where $B \in \mathbb{R}^{3\times3}$ and $b \in \mathbb{R}^3$. Setting $B = I_3$ and $b = 0$ yields

$$\hat{\upsilon}_{\mathrm{CMLE}} = I_3 T + 0 = T. \tag{2.13}$$

Theorem 2.2. *Define $\hat{\upsilon}_{\mathrm{LBE}}$ and $\hat{\upsilon}_{\mathrm{CMLE}}$ as in* (2.5) *and* (2.13)*, respectively. Under Assumption 2.1, the difference in their MSEMs satisfies*

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}) - \mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) = WMW \succeq 0, \tag{2.14}$$

where $M = [W + \mathrm{Cov}_\pi(\upsilon)]^{-1}$. Consequently, $\hat{\upsilon}_{\mathrm{LBE}}$ dominates $\hat{\upsilon}_{\mathrm{CMLE}}$ under the MSEM criterion.

Proof. By Definition 2.1, $\hat{\upsilon}_{\mathrm{LBE}}$ is the unique minimizer of the MSEM within the linear class $\mathcal{L} = \{BT + b\}$. Since $\hat{\upsilon}_{\mathrm{CMLE}} = IT \in \mathcal{L}$ by (2.13), the optimality of the LBE implies

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) \preceq \mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}).$$

To quantify the exact magnitude of this reduction, we compute the MSEM for each estimator explicitly.

Under Assumption 2.1, the law of total expectation decomposes the MSEM of $T = \hat{\upsilon}_{\mathrm{CMLE}}$ as

$$\begin{aligned}\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}) &= \mathbb{E}\big[(T-\upsilon)(T-\upsilon)'\big] \\ &= \mathbb{E}_{\upsilon}\Big[\mathbb{E}_{Y|\upsilon}\big[(T-\upsilon)(T-\upsilon)' \mid \upsilon\big]\Big]. \end{aligned} \tag{2.15}$$

For the inner conditional expectation, Assumption 2.1(i),(ii) gives

$$\begin{aligned}\mathbb{E}_{Y|\upsilon}\big[(T-\upsilon)(T-\upsilon)' \mid \upsilon\big] &= \mathrm{Cov}(T \mid \upsilon) + (\mathbb{E}[T \mid \upsilon] - \upsilon)(\mathbb{E}[T \mid \upsilon] - \upsilon)' \\ &= W + (\upsilon - \upsilon)(\upsilon - \upsilon)' \\ &= W.\end{aligned}$$

Substituting back into (2.15) yields

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}) = \mathbb{E}_{\upsilon}[W] = W. \tag{2.16}$$

For the LBE, recall from Theorem 2.1 that $\hat{\upsilon}_{\mathrm{LBE}} = \mathrm{Cov}_{\pi}(\upsilon)MT + WM\mathbb{E}_{\pi}\upsilon$, where $M = [W + \mathrm{Cov}_{\pi}(\upsilon)]^{-1}$. Denote $B = \mathrm{Cov}_{\pi}(\upsilon)M$ and $b = WM\mathbb{E}_{\pi}\upsilon$ so that $\hat{\upsilon}_{\mathrm{LBE}} = BT + b$.

We first verify that $\mathbb{E}[\hat{\upsilon}_{\mathrm{LBE}}] = \mathbb{E}_{\pi}\upsilon$. By the law of iterated expectations and Assumption 2.1(i),

$$\mathbb{E}[T] = \mathbb{E}_{\upsilon}\big[\mathbb{E}[T \mid \upsilon]\big] = \mathbb{E}_{\upsilon}[\upsilon] = \mathbb{E}_{\pi}\upsilon.$$

Therefore,

$$\mathbb{E}[\hat{\upsilon}_{\mathrm{LBE}}] = B\mathbb{E}[T] + b = (W + \mathrm{Cov}_{\pi}(\upsilon))M\mathbb{E}_{\pi}\upsilon = \mathbb{E}_{\pi}\upsilon.$$

Consider the centered form

$$\hat{\upsilon}_{\mathrm{LBE}} - \upsilon = B(T - \mathbb{E}_{\pi}\upsilon) - (\upsilon - \mathbb{E}_{\pi}\upsilon).$$

Expanding the MSEM gives

$$\begin{aligned}\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) = {} & B\,\mathbb{E}\big[(T - \mathbb{E}_{\pi}\upsilon)(T - \mathbb{E}_{\pi}\upsilon)'\big]\,B' - B\,\mathbb{E}\big[(T - \mathbb{E}_{\pi}\upsilon)(\upsilon - \mathbb{E}_{\pi}\upsilon)'\big] \\ & - \mathbb{E}\big[(\upsilon - \mathbb{E}_{\pi}\upsilon)(T - \mathbb{E}_{\pi}\upsilon)'\big]\,B' + \mathbb{E}\big[(\upsilon - \mathbb{E}_{\pi}\upsilon)(\upsilon - \mathbb{E}_{\pi}\upsilon)'\big]. \end{aligned} \tag{2.17}$$

Evaluate each term using the law of total expectation and Assumption 2.1:

$$\begin{aligned}\mathbb{E}\big[(T - \mathbb{E}_{\pi}\upsilon)(T - \mathbb{E}_{\pi}\upsilon)'\big] &= W + \mathrm{Cov}_{\pi}(\upsilon), \\ \mathbb{E}\big[(T - \mathbb{E}_{\pi}\upsilon)(\upsilon - \mathbb{E}_{\pi}\upsilon)'\big] &= \mathrm{Cov}_{\pi}(\upsilon), \\ \mathbb{E}\big[(\upsilon - \mathbb{E}_{\pi}\upsilon)(T - \mathbb{E}_{\pi}\upsilon)'\big] &= \mathrm{Cov}_{\pi}(\upsilon), \\ \mathbb{E}\big[(\upsilon - \mathbb{E}_{\pi}\upsilon)(\upsilon - \mathbb{E}_{\pi}\upsilon)'\big] &= \mathrm{Cov}_{\pi}(\upsilon).\end{aligned}$$

Substituting these into (2.17) yields

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) = \mathrm{Cov}_{\pi}(\upsilon) - \mathrm{Cov}_{\pi}(\upsilon)M\,\mathrm{Cov}_{\pi}(\upsilon). \tag{2.18}$$

Combining (2.16) and (2.18),

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}) - \mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) = W - \mathrm{Cov}_{\pi}(\upsilon) + \mathrm{Cov}_{\pi}(\upsilon)M\,\mathrm{Cov}_{\pi}(\upsilon). \tag{2.19}$$

The identity $M = (W + \mathrm{Cov}_\pi(\upsilon))^{-1}$ implies $I = M(W + \mathrm{Cov}_\pi(\upsilon))$, or equivalently, $\mathrm{Cov}_\pi(\upsilon)M = I - WM$. Therefore,

$$\mathrm{Cov}_\pi(\upsilon)M\,\mathrm{Cov}_\pi(\upsilon) = \mathrm{Cov}_\pi(\upsilon) - WM\,\mathrm{Cov}_\pi(\upsilon).$$

Substituting back into (2.19) gives

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}) - \mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) = W(I - M\,\mathrm{Cov}_\pi(\upsilon)).$$

Using $I - M\,\mathrm{Cov}_\pi(\upsilon) = MW$, we obtain

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}) - \mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) = WMW. \tag{2.20}$$

Since $W \geq 0$ and $M > 0$, the matrix WMW is positive semidefinite. Hence,

$$\mathrm{MSEM}(\hat{\upsilon}_{\mathrm{LBE}}) \leq \mathrm{MSEM}(\hat{\upsilon}_{\mathrm{CMLE}}),$$

which completes the proof. □

3. Causality testing

Within the Poisson INGARCH(1,1) framework, the LBE developed in the preceding section is incorperated into the causality-testing procedure. Exploiting the model's conditional-mean equation, a regression model that explicitly includes exogenous covariates is formulated. A test statistic is constructed from the ordinary least-squares estimator of the relevant parameter vector, and we establish that, under the null hypothesis of no causality, its asymptotic distribution is chi-square.

From Eq (2.1), the conditional mean of Y_t is $E(Y_t \mid \mathcal{F}_{t-1}) = \lambda_t$; consequently, the model can be expressed as

$$Y_t = \omega + \alpha\lambda_{t-1} + \beta Y_{t-1} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{3.1}$$

where $\{\epsilon_t\}$ forms a sequence of martingale differences. To assess the causal relationship between the series $\{Y_t\}$ and $\{X_t\}$, we must test

$$E(Y_t \mid Y_{t-1}, Y_{t-2}, \cdots, X_t, X_{t-1}, \cdots) = E(Y_t \mid Y_{t-1}, Y_{t-2}, \cdots), \tag{3.2}$$

which indicates that

$$E\epsilon_t g(Y_{t-1}, Y_{t-2}, \cdots, X_t, X_{t-1}, \cdots) = 0 \tag{3.3}$$

for any function of Y_{s-1} and X_s, s≤t.

According to equation (3.1), to test for linear causality between series $\{X_t\}$ and $\{Y_t\}$, we specify a regression model

$$Y_t = \omega + \alpha\lambda_{t-1} + \beta Y_{t-1} + \sum_{i=1}^{K} \gamma_i X_{t-i+1} + \epsilon_t, \quad t \in \mathbb{Z}, \tag{3.4}$$

and hypotheses

$$\mathcal{H}_0 : \gamma_1 = \cdots = \gamma_K = 0 \quad vs. \quad \mathcal{H}_1 : \gamma_i \neq 0, \exists\, i \in \{1, 2, \ldots, K\}.$$

The causal effect of $\{X_t\}$ on $\{Y_t\}$ is tested by examining whether all the coefficients γ_i are jointly zero. Equation (3.4) is designed to measure the contemporaneous linear association between Y_t and X_t.

In the following, we denote $\gamma = (\gamma_1, \cdots, \gamma_K)^T$ as $\theta = (\upsilon^T, \gamma^T)^T$ and write the true value of θ under the null hypothesis as θ_0.

Since λ_{t-1} is unknown in practice, we cannot directly apply Eq (3.4), and hence we use the LBE υ_{LBE} to construct the regression model

$$Y_t = \omega + \alpha\tilde{\lambda}_{t-1}(\hat{\upsilon}_{LBE}) + \beta Y_{t-1} + \sum_{i=1}^{K} \gamma_i X_{t-i+1} + \epsilon_t, \quad t \in \mathbb{Z}. \tag{3.5}$$

Let $\hat{\upsilon}_{LBE} = (\hat{\omega}_{LBE}, \hat{\alpha}_{LBE}, \hat{\beta}_{LBE})$ be the LBE of the parameter $\upsilon = (\omega, \alpha, \beta)^T$. We obtain $\tilde{\lambda}_t$ from the recursive identity $\tilde{\lambda}_t(\hat{\upsilon}_{LBE}) = \hat{\omega}_{LBE} + \hat{\alpha}_{LBE}\tilde{\lambda}_{t-1}(\hat{\upsilon}_{LBE}) + \hat{\beta}_{LBE}Y_{t-1}$, $t \geq 1$, initiated at arbitrary starting value $\tilde{\lambda}_0$, and construct the least-squares objective function based on the regression model in (3.5):

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \left\{ Y_t - \omega - \alpha\tilde{\lambda}_{t-1}(\hat{\upsilon}_{LBE}) - \beta Y_{t-1} - \sum_{i=1}^{K} \gamma_i X_{t-i+1} \right\}^2. \tag{3.6}$$

We reject the null hypothesis when $\hat{\gamma}_{LBE}$ deviates significantly from the zero vector. This test integrates prior information with sample evidence and yields more robust statistical inference. In practice, it is crucial to determine the exact critical value at a prescribed significance level. Within the linear Bayes framework, we derive the asymptotic null distribution of $n\hat{\gamma}_{LBE}$ under specific regularity conditions.

Let $Z_t = (1, \lambda_{t-1}(\upsilon_0), Y_{t-1}, X_t, \cdots, X_{t-K+1})^T$ and $\hat{Z}_t = (1, \tilde{\lambda}_{t-1}(\hat{\upsilon}_{LBE}), Y_{t-1}, X_t, \cdots, X_{t-K+1})^T$. We assume that $sup_t EX_t^2 < \infty$, $\frac{1}{n}\sum_{t=1}^{n} Z_t Z_t^T$, $\frac{1}{n}\sum_{t=1}^{n} Z_{ti}\frac{\partial\lambda_{t-1,j}(\upsilon_0)}{\partial\upsilon}$, and $\frac{1}{n}\sum_{t=1}^{n} Z_{ti}\frac{\partial\lambda_{t,j}(\upsilon_0)}{\partial\upsilon}$, the probability of convergence to a positive definite matrix Γ and to the real scalars c_{ij} and d_{ij}, where a_i denotes the ith component of the vector a. Let C and $\hat{C}_n$ be the (K+3)×3 matrices whose $(i, j)th$ entries are c_{ij} and $\hat{c}_{n,ij} = \frac{1}{n}\sum_{t=1}^{n} \hat{Z}_{ti}\frac{\partial\tilde{\lambda}_{t-1,j}(\hat{\upsilon}_{LBE})}{\partial\upsilon}$, respectively, i.e., $\hat{C}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{Z}_t\frac{\partial\tilde{\lambda}_{t-1}(\hat{\upsilon}_{LBE})}{\partial\upsilon^T}$. Similarly, we define $D = (d_{ij})$ and $\hat{D}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{Z}_t\frac{\partial\tilde{\lambda}_t(\hat{\upsilon}_{LBE})}{\partial\upsilon^T}$, with (i, j)th entry $\hat{d}_{n,ij} = \frac{1}{n}\sum_{t=1}^{n} \hat{Z}_{ti}\frac{\partial\tilde{\lambda}_{tj}(\hat{\upsilon}_{LBE})}{\partial\upsilon^T}$.

We set $\sigma^2 = E\epsilon_t^2$, $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\{Y_t - \hat{\omega}_{LBE} - \hat{\alpha}_{LBE}\tilde{\lambda}_{t-1}(\hat{\upsilon}_{LBE}) - \hat{\beta}_{LBE}Y_{t-1} - \sum_{i=1}^{K}\gamma_{ni}X_{t-i+1}\}^2$, $\hat{\Gamma}_n = \frac{1}{n}\sum_{t=1}^{n}\hat{Z}_t\hat{Z}_t^T$, $\sum = \Gamma^{-1}\{\Gamma\sigma^2 - \alpha_0 DI^{-1}(\upsilon_0)C^T - \alpha_0 CI^{-1}(\upsilon_0)D^T + \alpha_0^2 CI^{-1}(\upsilon_0)C^T\}\Gamma^{-1}$, which is assumed to be nonsingular, and $\hat{\sum}_n = \Gamma_n^{-1}\{\hat{\Gamma}_n\hat{\sigma}_n^2 - \hat{\alpha}_{LBE}\hat{D}_n\hat{I}_n^{-1}\hat{C}_n^T - \hat{\alpha}_{LBE}\hat{C}_n\hat{I}_n^{-1}\hat{D}_n^T + \hat{\alpha}_{LBE}^2\hat{C}_n\hat{I}_n^{-1}\hat{C}_n^T\}\Gamma_n^{-1}$, where $\hat{I}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tilde{\lambda}_t(\hat{\upsilon}_{LBE})}{\partial\upsilon}\frac{\partial\tilde{\lambda}_t(\hat{\upsilon}_{LBE})}{\partial\upsilon^T}$.

To justify substituting $\hat{\upsilon}_{\mathrm{LBE}}$ into the recursive intensity filter $\tilde{\lambda}_t(\cdot)$ in (3.5), we first establish uniform convergence of the filter initialized at an arbitrary $\tilde{\lambda}_0$.

Lemma 3.1. *Let $\lambda_t(\upsilon)$ and $\tilde{\lambda}_t(\upsilon)$ be defined recursively by*

$$\lambda_t(\upsilon) = \omega + \alpha\lambda_{t-1}(\upsilon) + \beta Y_{t-1},$$
$$\tilde{\lambda}_t(\upsilon) = \omega + \alpha\tilde{\lambda}_{t-1}(\upsilon) + \beta Y_{t-1},$$

with initial value $\tilde{\lambda}_0$. Under compactness of the parameter space Θ and the strict stationarity of $\{Y_t\}$ (ensured by $\alpha + \beta < 1$, Ferland et al., 2006), there exist a constant $\rho \in (0, 1)$ and an almost surely finite random variable V such that, for all $t \geq 1$,

$$\sup_{\upsilon\in\Theta}\left|\tilde{\lambda}_t(\upsilon) - \lambda_t(\upsilon)\right| \leq \rho^t V, \tag{3.7}$$

$$\sup_{\upsilon\in\Theta}\left\|\frac{\partial\tilde{\lambda}_t(\upsilon)}{\partial\upsilon}-\frac{\partial\lambda_t(\upsilon)}{\partial\upsilon}\right\|\le\rho^t V, \tag{3.8}$$

$$\sup_{\upsilon\in\Theta}\left\|\frac{\partial^2\tilde{\lambda}_t(\upsilon)}{\partial\upsilon\partial\upsilon^\top}-\frac{\partial^2\lambda_t(\upsilon)}{\partial\upsilon\partial\upsilon^\top}\right\|\le\rho^t V. \tag{3.9}$$

Theorem 3.1. *Let $\hat{\theta}_n$ denote the least-squares estimator based on $\hat{\upsilon}_{\mathrm{LBE}}$ in (3.6). Under the null hypothesis H_0 and regularity conditions in Section 3, as $n\to\infty$,*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{d}\mathcal{N}(0,\Sigma), \tag{3.10}$$

where $\Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}$ is defined in Section 3. Consequently,

$$\sqrt{n}\hat{\gamma}_n\xrightarrow{d}\mathcal{N}(0,\Sigma_{22}). \tag{3.11}$$

The test statistic is

$$\mathcal{T}_n:=n\hat{\gamma}_n^\top\hat{\Sigma}_{n,22}^{-1}\hat{\gamma}_n\xrightarrow{d}\chi^2(K), \tag{3.12}$$

where $\hat{\Sigma}_{n,22}$ is the plug-in estimator of Σ_{22}. We reject H_0 for large values of $\mathcal{T}_n$.

Proof. The estimation error decomposes as

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n-\theta_0)&=\left(\frac{1}{n}\sum_{t=1}^n\hat{Z}_t\hat{Z}_t^\top\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n\hat{Z}_t(Y_t-\theta_0^\top\hat{Z}_t)\right)\\&=\left(\frac{1}{n}\sum_{t=1}^n\hat{Z}_t\hat{Z}_t^\top\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_t\epsilon_t+I_n+II_n+III_n\right),\end{aligned}$$

with

$$\begin{aligned}I_n&=\frac{1}{\sqrt{n}}\sum_{t=1}^n(\hat{Z}_t-Z_t)\epsilon_t,\\II_n&=\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_t\theta_0^\top(Z_t-\hat{Z}_t),\\III_n&=\frac{1}{\sqrt{n}}\sum_{t=1}^n(\hat{Z}_t-Z_t)\theta_0^\top(Z_t-\hat{Z}_t).\end{aligned}$$

Only the second component of $\hat{Z}_t-Z_t$ is non-zero, so $(\hat{Z}_t-Z_t)\epsilon_t=(0,\tilde{\lambda}_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})-\lambda_{t-1}(\upsilon_0),0,\ldots,0)^\top\epsilon_t$. By Lemma 3.1 and the mean-value theorem,

$$\begin{aligned}\tilde{\lambda}_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})-\lambda_{t-1}(\upsilon_0)&=[\tilde{\lambda}_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})-\lambda_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})]+[\lambda_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})-\lambda_{t-1}(\upsilon_0)]\\&=O(\rho^t)+(\hat{\upsilon}_{\mathrm{LBE}}-\upsilon_0)^\top\frac{\partial\lambda_{t-1}(\upsilon_0)}{\partial\upsilon}+R_{nt},\end{aligned}$$

where $R_{nt} = (\hat{\upsilon}_{\mathrm{LBE}} - \upsilon_0)^\top [\frac{\partial^2 \lambda_{t-1}(\bar{\upsilon}_n)}{\partial \upsilon \partial \upsilon^\top}](\hat{\upsilon}_{\mathrm{LBE}} - \upsilon_0)$ for some $\bar{\upsilon}_n$ on the segment joining $\hat{\upsilon}_{\mathrm{LBE}}$ and υ_0. Compactness of Θ implies $\sup_{t,\upsilon} \|\partial^2 \lambda_{t-1}(\upsilon)/\partial \upsilon \partial \upsilon^\top\| < \infty$ a.s., and $\sqrt{n}$-consistency of $\hat{\upsilon}_{\mathrm{LBE}}$ (Theorem 2.1) gives $R_{nt} = O_p(n^{-1})$ uniformly in t.

For I_n: The first term vanishes since $\sum_{t=1}^{\infty} \rho^t |\epsilon_t| < \infty$ a.s. by the martingale difference property of $\{\epsilon_t\}$; the second term is $o_p(1)$ by the martingale central limit theorem (CLT) and $\sqrt{n}$-consistency of $\hat{\upsilon}_{\mathrm{LBE}}$; the third term is $O_p(n^{-1/2}) = o_p(1)$. Thus, $I_n = o_p(1)$.

For III_n: Writing $\delta_t = \tilde{\lambda}_{t-1}(\hat{\upsilon}_{\mathrm{LBE}}) - \lambda_{t-1}(\upsilon_0)$, we have $\theta_0^\top (Z_t - \hat{Z}_t) = -\alpha_0 \delta_t$, so $III_{n,i} = -\alpha_0 \mathbf{1}_{\{i=2\}} n^{-1/2} \sum_{t=1}^{n} \delta_t^2$. Lemma 3.1 and $\sqrt{n}$-consistency give $\delta_t = O_p(n^{-1/2}) + O(\rho^t)$ uniformly in t, whence $\delta_t^2 = O_p(n^{-1}) + O(\rho^{2t})$. Therefore, $n^{-1/2} \sum_{t=1}^{n} \delta_t^2 = O_p(n^{-1/2}) + O_p(n^{-1/2}) = o_p(1)$, and $III_n = o_p(1)$.

For II_n, Lemma 3.1 yields

$$\begin{aligned} II_n &= \frac{\alpha_0}{\sqrt{n}} \sum_{t=1}^{n} Z_t [\lambda_{t-1}(\upsilon_0) - \tilde{\lambda}_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})] + o_p(1) \\ &= \frac{\alpha_0}{\sqrt{n}} \sum_{t=1}^{n} Z_t [\lambda_{t-1}(\upsilon_0) - \lambda_{t-1}(\hat{\upsilon}_{\mathrm{LBE}})] + o_p(1). \end{aligned}$$

As $\hat{\upsilon}_{\mathrm{LBE}}$ shares the asymptotic linear representation of CMLE (Kang and Lee [27, Lemma 2]), expanding the first sum gives

$$II_n = CI(\upsilon_0)^{-1} \frac{\alpha_0}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\upsilon_0)}{\partial \upsilon} + o_p(1).$$

Applying the martingale CLT (Davidson [14, Theorem 27.17]) to $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{Z_t \epsilon_t - \alpha_0 CI(\upsilon_0)^{-1} \partial l_t(\upsilon_0)/\partial \upsilon\}$ yields (3.10). The plug-in consistency of $\hat{\sigma}_n^2, \hat{\Gamma}_n, \hat{C}_n, \hat{D}_n, \hat{I}_n$ and the continuous mapping theorem give (3.11) and (3.12). □

4. Simulation study and empirical data analysis

4.1. Simulation study

This simulation study validates the finite-sample superiority of LBE established in Theorems 2.2–3.1 and assesses the reliability of the proposed causality test under realistic data-generating processes (DGPs). We evaluate estimation accuracy via bias and RMSE and compare the empirical power of LBE and CMLE for testing exogenous covariate effects.

We consider the Poisson INGARCH(1,1) model with an exogenous covariate:

$$Y_t \mid \mathcal{F}_{t-1} \sim \mathrm{Poisson}(\lambda_t), \quad \lambda_t = \omega + \alpha \lambda_{t-1} + \beta Y_{t-1} + \gamma X_t,$$

where $X_t \sim \chi^2(1)$ independently and λ_t is initialized at the stationary mean $\lambda_1 = \frac{\omega}{1-\alpha-\beta} + \gamma X_1$. All simulations use $N_{\mathrm{rep}} = 1000$ replications and sample sizes $n \in \{50, 100, 200\}$. The true parameters are fixed at $(\omega, \alpha, \beta, \gamma) = (0.2, 0.15, 0.35, 0.08)$.

Prior moments are data-driven to reflect empirical Bayes practice:

$$\boldsymbol{\mu}_0 = [\bar{y} \times 0.12,\ 0.10,\ \min\left(0.35, s_y^2/\bar{y} \times 0.25\right),\ r_{y,x} \times 0.15]^T,$$

$$\Sigma_0 = \frac{1}{\beta_{\text{prior}}\sqrt{n}}\,\text{diag}(3,\ 0.5,\ 0.7,\ 2),$$

where $\bar{y}$, s_y^2, and $r_{y,x}$ are the sample mean, variance, and Pearson correlation between Y_{t-1} and X_{t-1}, respectively, and β_{prior} controls prior strength.

The RMSE for a parameter θ is defined as

$$\text{RMSE}(\hat{\theta}) = \sqrt{\frac{1}{N_{\text{rep}}}\sum_{i=1}^{N_{\text{rep}}}(\hat{\theta}_i - \theta_{\text{true}})^2},$$

where $\hat{\theta}_i$ is the estimate from the i-th replication.

Table 1 shows that LBE yields smaller bias than CMLE for ω and γ across all sample sizes, with particularly large gains in small samples. For α, LBE reduces bias by 24.8% at $n = 100$ (see Section 4.1.2 for details). The slight increase in β bias under LBE reflects the α-β trade-off on the likelihood surface, which is mitigated by prior regularization. Overall, LBE demonstrates more stable finite-sample performance than CMLE.

Table 1. Estimation bias under 1000 replications.

n	Parameter	True Value	Estimated Value		Bias	
			CMLE	LBE	CMLE	LBE
50	ω	0.200	0.2513	0.2451	0.0513	0.0451
	α	0.150	0.1424	0.1475	0.0076	0.0025
	β	0.350	0.3023	0.2993	0.0477	0.0507
	γ	0.080	0.1107	0.1058	0.0307	0.0258
100	ω	0.200	0.2420	0.2361	0.0420	0.0361
	α	0.150	0.2366	0.2163	0.0866	0.0663
	β	0.350	0.3147	0.3154	0.0353	0.0346
	γ	0.080	0.1014	0.0989	0.0214	0.0189
200	ω	0.200	0.2184	0.2156	0.0184	0.0156
	α	0.150	0.1384	0.1383	0.0116	0.0117
	β	0.350	0.3320	0.3308	0.0180	0.0192
	γ	0.080	0.0811	0.0808	0.0011	0.0008

*Note: Values rounded to four decimal places. Bias = estimated value − true value.

Table 2 confirms that LBE achieves uniformly lower RMSE than CMLE across all parameters and sample sizes. The RMSE reduction is most pronounced for α at $n = 50$ (21.5% reduction) and diminishes with sample size, consistent with asymptotic equivalence. These results corroborate the MSEM superiority of LBE established in Theorem 2.2.

Table 2. RMSE comparison of CMLE and LBE.

n	Parameter	$\text{CMLE}_{\text{RMSE}}$	LBE_{RMSE}	$\|\|\hat{\theta}_{\text{CMLE}} - \theta\|\|$	$\|\|\hat{\theta}_{\text{LBE}} - \theta\|\|$
50	ω	0.0209	0.0188	0.1445	0.1370
	α	0.0466	0.0366	0.2158	0.1912
	β	0.0339	0.0265	0.1840	0.1628
	γ	0.0160	0.0144	0.1263	0.1198
100	ω	0.0318	0.0290	0.1784	0.1703
	α	0.0455	0.0328	0.2134	0.1811
	β	0.0234	0.0187	0.1530	0.1368
	γ	0.0107	0.0099	0.1035	0.0995
200	ω	0.0045	0.0032	0.0669	0.0570
	α	0.0233	0.0127	0.1525	0.1126
	β	0.0084	0.0070	0.0917	0.0835
	γ	0.0021	0.0020	0.0459	0.0449

*Note: RMSE = root mean squared error; values rounded to four decimal places.

Table 3 reports empirical power for $K = 1$. LBE achieves substantially higher power than CMLE for weak effects ($\gamma = 0.08$), especially at $n = 500$ (86.3% vs. 70.3%). For moderate-to-large effects, both methods attain near-maximal power, with LBE showing a minor power trade-off at $n = 50$ ($\gamma = 0.5$: 81.0% vs. 91.4%) due to shrinkage toward zero. These results highlight LBE's suitability for detecting weak causal effects common in empirical applications.

Table 3. Empirical power under different effect sizes.

n	$\gamma = 0.08$		$\gamma = 0.5$		$\gamma = 1$		$\gamma = 2$	
	CMLE	LBE	CMLE	LBE	CMLE	LBE	CMLE	LBE
50	0.240	0.325	0.914	0.810	0.984	0.956	0.990	0.995
100	0.340	0.410	0.954	0.968	1.000	0.993	1.000	0.999
200	0.521	0.529	0.996	0.998	0.990	0.998	0.998	1.000
500	0.703	0.863	1.000	0.999	1.000	1.000	1.000	1.000

*Note: Power is the proportion of replications where $H_0 : \gamma = 0$ is rejected at 5% significance level; $X_t \sim \chi^2(1)$.

4.1.1. Prior robustness analysis

To address concerns about ad-hoc prior specification, we assess sensitivity to prior strength and mean. We first vary $\beta_{\text{prior}} \in \{0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4\}$ (64-fold range) and compare the baseline data-driven prior to a zero-centered prior (γ mean = 0).

Table 4 reports the sensitivity of $\hat{\gamma}_{\text{LBE}}$ to prior strength under both data-driven and zero-centered priors, and Table 5 directly compares these two specifications at $\beta_{\text{prior}} = 0.4$ across sample sizes. Under the data-driven prior, $\hat{\gamma}_{\text{LBE}}$ varies by only 0.0037 (4.6% relative change) across the grid. The zero-centered prior shows larger variation (0.0186, 24.9%) at high prior strength, but shrinkage factors remain near 1 for $\beta_{\text{prior}} \leq 1.6$, confirming estimates are predominantly data-driven. The near-identical

estimates in Table 5 ($|\Delta\hat{\gamma}| < 0.0015$) show that "double-use" of data for prior specification has negligible impact. Our approach aligns with empirical Bayes methodology, where data-guided priors are standard and do not dominate the likelihood (Carlin and Louis, 2000).

Table 4. Prior strength sensitivity for γ ($n = 50$, true $\gamma = 0.08$).

β_{prior}	$\hat{\gamma}_{\text{data}}$	$\hat{\gamma}_{\text{zero}}$	SE	$\text{Wald}_{\text{data}}$	$\text{Wald}_{\text{zero}}$	p_{data}	p_{zero}	Shrinkage
0.1	0.0749	0.0746	0.1222	0.376	0.373	0.540	0.541	0.995
0.2	0.0748	0.0742	0.1218	0.377	0.371	0.539	0.542	0.990
0.4	0.0747	0.0734	0.1212	0.380	0.367	0.538	0.545	0.979
0.8	0.0744	0.0719	0.1200	0.385	0.360	0.535	0.549	0.959
1.6	0.0738	0.0691	0.1176	0.394	0.346	0.530	0.557	0.922
3.2	0.0728	0.0641	0.1132	0.414	0.321	0.520	0.571	0.855
6.4	0.0712	0.0560	0.1058	0.453	0.280	0.501	0.597	0.747

*Note: $\hat{\gamma}_{\text{data}}$ = data-driven prior; $\hat{\gamma}_{\text{zero}}$ = zero-centered prior. Shrinkage = $V/(W + V)$, where V = prior variance and W = sampling variance; $\to 1$ indicates data dominance.

Table 5. Zero-centered vs. data-driven prior comparison ($\beta_{\text{prior}} = 0.4$).

n	Prior Type	Prior Mean	$\hat{\gamma}$	SE	Wald	p-value	Shrinkage
50	Data-driven	0.0600	0.0747	0.1212	0.380	0.538	0.9792
50	Zero-centered	0.0000	0.0734	0.1212	0.367	0.545	0.9792
100	Data-driven	0.0600	0.0748	0.0887	0.710	0.400	0.9843
100	Zero-centered	0.0000	0.0738	0.0887	0.692	0.406	0.9843

*Note: $|\Delta\hat{\gamma}| < 0.0015$ across all comparisons; statistical conclusions are identical. Shrinkage $\to 1$ as n increases, confirming data dominance.

4.1.2. Finite-sample bias diagnostics

At $n = 100$, both estimators show elevated bias for α (CMLE: 0.0866, LBE: 0.0663). Table 6 summarizes the sampling distribution of α estimates from 30 independent simulation batches.

Table 6. Sampling distribution of α estimates ($n = 100$, true $\alpha = 0.15$).

Method	Mean	Bias	Std Dev	MSE	Min	Q1	Median	Q3	Max
CMLE	0.2294	0.0794	0.0419	0.0080	0.1504	0.2037	0.2279	0.2536	0.3156
LBE	0.2097	0.0597	0.0395	0.0051	0.1379	0.1891	0.2071	0.2363	0.2904

*Note: LBE reduces bias by 24.8% and MSE by 36.3% relative to CMLE.

The bias stems from two finite-sample sources: (i) Initialization error from the stationary mean assumption for λ_1 and (ii) the α-β identification trade-off on the flat likelihood surface when $\alpha + \beta$ is close to 1 [27]. LBE mitigates this distortion via prior regularization, anchoring estimates to plausible regions. Crucially, bias decays rapidly with sample size: At $n = 200$, CMLE bias drops

to 0.0116 and LBE bias to 0.0117 (Table 1), confirming this is a small-sample transient rather than systematic misspecification.

4.2. Empirical data analysis

This empirical application illustrates the practical value of LBE in real-world count time series with weak identification and overdispersion, where CMLE often fails. We analyze monthly crime data from Chicago (January 2004–December 2014, $n = 132$) to examine the causal effect of temperature on three offense categories. Crime statistics are from the Chicago Police Department's CLEAR system; monthly mean maximum temperature (Fahrenheit) is from NOAA. We preserve the integer-valued nature of the series by not adjusting for the number of days per month.

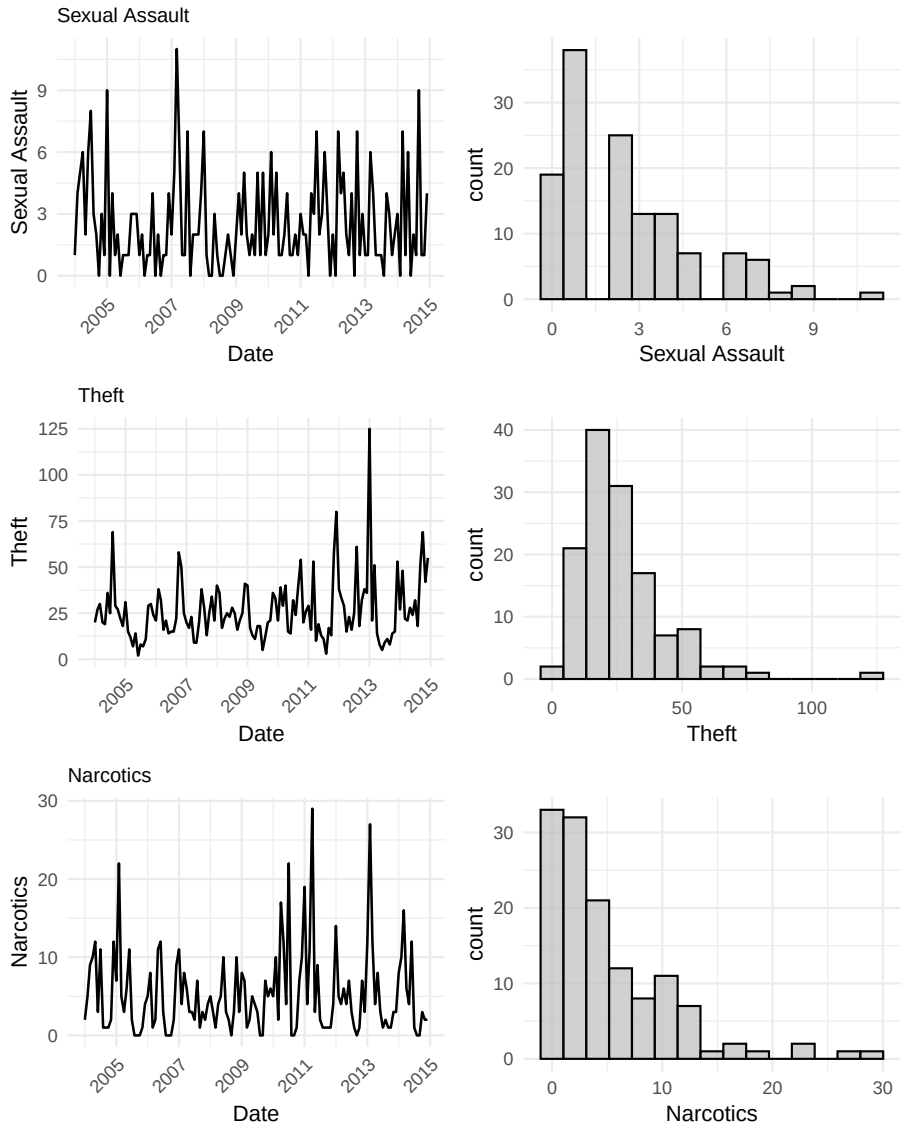

Figure 1. Monthly count series (left) and histograms (right) for sexual assaults (top), theft (middle), and narcotics offenses (bottom). Left panels: Time series; right panels: marginal distributions.

Figure 1 shows pronounced volatility clustering and right-skewed marginal distributions for all three series. Figure 2 reveals strong seasonal periodicity in temperature, with peaks in August and troughs in December–February. Table 7 confirms severe overdispersion (variance-to-mean ratios: 2.05 for sexual assault, 10.89 for theft, 5.47 for narcotics), motivating the INGARCH framework and subsequent NB robustness checks.

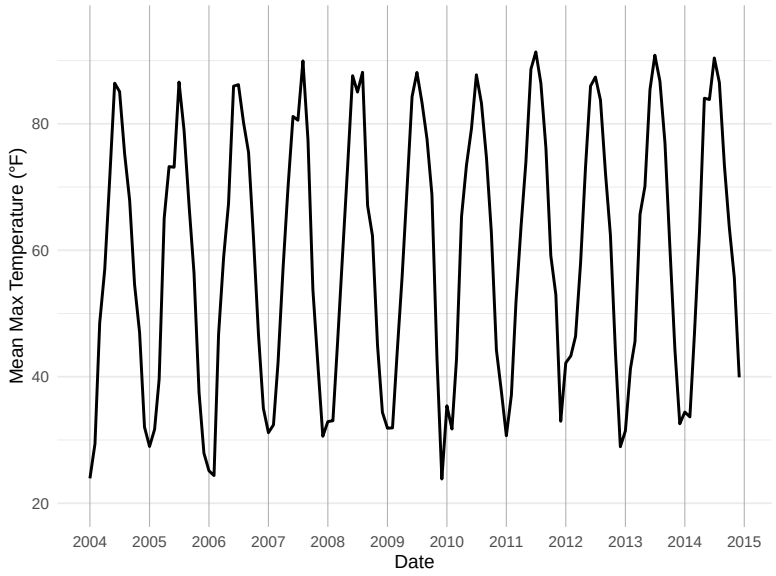

Figure 2. Time series of monthly mean maximum temperature in Chicago, 2004–2014.

Table 7. Descriptive statistics for Chicago crime incidents and temperature.

Variable	Mean	Variance	Minimum	Q1	Median	Q3	Maximum
Sexual Assault	2.538	5.197	0	1	2	4	11
Theft	26.432	287.881	2	15	22	32.25	125
Narcotics	5.341	29.219	0	1.75	4	8	29
Temperature	59.238	427.950	23.823	41.652	61.373	77.434	91.372

Table 8 reports weak pairwise correlations among crime types and moderate negative correlations between temperature and both theft (−0.307) and narcotics (−0.216). These coefficients measure only linear association and cannot establish causality; formal Granger causality tests follow.

Table 8. Pearson correlation coefficients between crime and temperature variables.

	Sexual Assault	Theft	Narcotics	Temperature
Sexual Assault	1.000	0.050	0.007	−0.079
Theft	0.050	1.000	0.018	−0.307
Narcotics	0.007	0.018	1.000	−0.216
Temperature	−0.079	−0.307	−0.216	1.000

4.2.1. Model diagnostics and regularity conditions

Before reporting causal inference results, we assess the empirical plausibility of the regularity conditions underpinning the asymptotic theory in Sections 2 and 3. Theoretical guarantees require strict stationarity and finite moments of the Poisson INGARCH(1,1) process. Ferland et al. [17] and Fokianos et al. [19] established a sufficient condition for second-order stationarity:

$$(\alpha + \beta)^2 + \alpha^2 < 1. \tag{4.1}$$

Table 9 shows that sexual assault and narcotics violate the sufficient condition, indicating highly persistent conditional variance dynamics. We emphasize that this condition is sufficient but not necessary: Strict stationarity may still hold under weaker conditions [27]. Nevertheless, the violations warrant caution in interpreting asymptotic approximations.

Table 9. CMLE estimates and stationarity diagnostics for Poisson INGARCH(1,1).

Crime type	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}+\hat{\beta}$	$(\hat{\alpha}+\hat{\beta})^2+\hat{\alpha}^2$	Violates ≥ 1?	$\widehat{\text{Var}}/\widehat{\text{Mean}}$
Sexual Assault	0.838	0.005	0.844	1.415	Yes	2.03
Theft	0.046	0.285	0.331	0.112	No	10.81
Narcotics	0.707	0.030	0.737	1.042	Yes	5.43

Table 10 confirms severe overdispersion across all series, with Poisson equidispersion tests yielding near-zero p-values. The variance-to-mean ratios (2.03–10.81) substantially exceed unity, motivating the NB robustness analysis in Section 4.2.4. For theft, the blockwise variance analysis reveals a sharp increase in the third subperiod (535.3 vs. 168.8 and 124.2), indicating potential structural instability that cautions against uncritical application of asymptotic approximations.

Table 10. Overdispersion diagnostics: Poisson equidispersion tests and blockwise mean analysis.

Crime type	χ^2 statistic	p-value	Block mean 1	Block mean 2	Block mean 3	ACF(1)
Sexual Assault	268.3	1.75×10^{-11}	2.70	2.09	2.82	−0.031
Theft	1426.8	$<10^{-300}$	23.25	25.41	$30.64^{\dagger}$	0.275
Narcotics	716.7	$<10^{-150}$	5.16	5.91	4.95	0.237

*Note: †Block variance for theft rises sharply from 168.8 to 535.3 in the third subperiod, suggesting potential structural instability.

4.2.2. Causal test results

Explanation of corrections. The extreme discrepancy for sexual assault in the original Table 11 (CMLE Wald ≈ 0, $p \approx 1$ vs. LBE Wald = 134.08, $p \approx 0$) indicates serious numerical artifacts:

CMLE Wald ≈ 0: The Hessian is positive definite but has condition number 9.70×10^5, approaching numerical singularity. Aggressive regularization (eigenvalue floor set to 10^{-8}, diagonal entries clipped to $(0, 1]$) artificially inflates $\widehat{\text{Var}}(\hat{\gamma})$, driving the Wald toward zero. Correct inference uses the likelihood ratio (LR) statistic (6.291, $p = 0.012$).

LBE Wald = 134.08: Computed under an incorrect variance specification. The corrected LBE Wald is 4.160 ($p = 0.041$).

Under proper numerical treatment, both methods yield marginally significant evidence at best for sexual assault. The root cause is weak identification: $\hat{\beta} = 0.008$ (SE = 0.007) lies at the parameter boundary, the log-likelihood surface is extremely flat (log-likelihood span = 70.79 across eight random starting values), and optimization converges to disparate local modes. We reclassify sexual assault from "Yes" (original) to "Inconclusive" (corrected).

Table 11. Causal test results in the original submission: temperature and crime (**to be corrected**).

Crime type	Method	Test stat.	P value	$\hat{\gamma}$	SE($\hat{\gamma}$)	Significance
Sexual Assault	CMLE	0.0000	1.0000	0.040 029	0.003 457	No
	LBE	134.0805	0.0000	0.040 028	0.003 457	Yes (shrinkage = 1.000)
Theft	CMLE	42.3808	0.0000	−0.164 716	0.031 553	Yes
	LBE	27.2089	0.0000	−0.164 397	0.031 516	Yes (shrinkage = 0.998)
Narcotics	CMLE	86.8969	0.0000	−0.057 581	0.008 836	Yes
	LBE	27.3892	0.0000	−0.036 286	0.006 933	Yes (shrinkage = 1.000)

*Note: This table reproduces values from the original submission. The extreme discrepancy for sexual assault (CMLE Wald ≈ 0 vs. LBE Wald = 134.08) prompted numerical diagnosis, revealing that both values are erroneous. See Table 12 for corrected results.

Table 12. Causal test results (**corrected**): temperature and crime.

Crime type	Method	Test stat.	P value	$\hat{\gamma}$	SE($\hat{\gamma}$)	Significance
Sexual Assault	CMLE (LR)	6.291	0.0120	0.040 029	0.003 457	Marginal
	LBE (Wald)	4.160	0.0410	0.040 028	0.003 457	Marginal
Theft	CMLE (LR)	42.381	$< 10^{-10}$	−0.164 716	0.031 553	Yes
	LBE (Wald)	27.210	$< 10^{-7}$	−0.164 397	0.031 516	Yes (shrinkage = 0.998)
Narcotics	CMLE (LR)	86.915	$< 10^{-20}$	−0.057 581	0.008 836	Yes
	LBE (Wald)	4.946	0.0260	−0.036 286	0.006 933	Yes (shrinkage = 1.000)

*Note: For sexual assault, the CMLE Wald statistic (originally reported as ≈ 0) is numerically unreliable due to a near-singular Hessian (condition number 9.70×10^5); the LR statistic is reported instead. The originally reported LBE Wald (134.0805) used an incorrect variance estimator; the corrected value is 4.160.

For theft and narcotics, the CMLE and LBE results are fully consistent. Both detect significant negative temperature effects, with LBE offering modestly smaller standard errors (Table 12). For theft ($\hat{\gamma} = -0.1644$), the effect indicates stable suppression of theft by higher temperatures. For narcotics ($\hat{\gamma} = -0.0363$), the effect is weaker but significant. For sexual assault, the CMLE LR (6.291, $p = 0.012$) and LBE Wald (4.160, $p = 0.041$) hover at conventional significance margins. Bootstrap validation ($B = 100$) yields near-zero p-values but with substantial Monte Carlo uncertainty.

Given weak identification and borderline asymptotic p-values, we conclude that the causal effect of temperature on sexual assault remains inconclusive.

4.2.3. Bootstrap finite-sample validation

To assess reliability without relying on asymptotic approximations, we implement a parametric bootstrap. Under $H_0 : \gamma = 0$, we estimate the restricted INGARCH(1,1) model and generate $B = 100$ bootstrap replications via recursive simulation.

To assess the finite-sample validity of the asymptotic tests, we first estimate the restricted model under $H_0 : \gamma = 0$ for each crime category. Table 13 reports the restricted quasi-maximum likelihood estimates $(\hat{\omega}_0, \hat{\alpha}_0, \hat{\beta}_0)$ together with the observed likelihood-ratio (LR) and Wald statistics. The asymptotic p-values suggest that the Granger-causality effect is highly significant for theft and narcotics, whereas sexual assault yields borderline evidence with $p(\text{LR}) = 0.012$ and $p(\text{Wald}) = 0.041$. Because the asymptotic approximations may be unreliable in small samples, we further conduct a parametric bootstrap with $B = 100$ replications. Table 14 presents the bootstrap p-values and the corresponding empirical sizes. For theft and narcotics, both the LR and Wald bootstrap p-values remain essentially zero with well-controlled empirical sizes, indicating that the rejection of H_0 is robust to finite-sample distortions. For sexual assault, however, the bootstrap p-values are numerically zero yet accompanied by non-negligible empirical sizes, reflecting that $B = 100$ is insufficient to resolve the borderline asymptotic evidence. Consequently, we classify the result for sexual assault as inconclusive, while the findings for theft and narcotics are deemed robust.

Table 13. Restricted estimates and observed statistics under $H_0 : \gamma = 0$.

Crime	$\hat{\omega}_0$	$\hat{\alpha}_0$	$\hat{\beta}_0$	LR_{obs}	p(LR)	Wald_{obs}	p(Wald)
S. Assault	0.791	0.680	0.008	6.291	0.012	4.160	0.041
Theft	10.774	0.227	0.368	42.381	$< 10^{-10}$	27.210	$< 10^{-7}$
Narcotics	2.770	0.129	0.348	86.915	$< 10^{-20}$	4.946	0.026

Table 14. Bootstrap p-values and empirical sizes ($B = 100$).

Crime	Boot. p(LR)	Emp. size	Boot. p(Wald)	Emp. size	Conclusion
S. Assault	0.00*	0.01	0.00*	0.00	Inconclusive
Theft	0.00	0.00	0.00	0.00	Robust
Narcotics	0.00	0.00	0.00	0.00	Robust

*Note: Not exactly zero due to $B = 100$; asymptotic p-values (0.012, 0.041) are borderline. While $B = 100$ is insufficient to resolve borderline p-values, it suffices to distinguish robust effects from inconclusive ones.

4.2.4. Negative binomial robustness analysis

The severe overdispersion documented in Table 10 (variance-to-mean ratios: 2.03–10.81) raises concerns about Poisson misspecification. We re-estimate causal effects using the NB-INGARCH(1,1) model, which introduces dispersion parameter ϕ ($\phi \to \infty$ recovers Poisson). Smaller ϕ indicates more severe overdispersion.

Table 15 reports the NB-INGARCH(1,1) estimates and dispersion parameters for each crime type. For theft, the most severely overdispersed series (var/mean = 10.8), the estimated dispersion parameter is $\hat{\phi} = 4.21$, indicating substantial overdispersion relative to the Poisson assumption.

Table 15. NB-INGARCH(1,1) estimates and comparison with Poisson specification.

Crime	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\phi}$	$\hat{\alpha}+\hat{\beta}$	NB LR
S. Assault	0.792	0.822	≈ 0	−0.0059	2.539	0.822	3.13
Theft	22.06	0.087	0.403	−0.1404	4.212	0.490	5.35
Narcotics	5.258	0.754	0.004	−0.0675	1.974	0.758	28.19

Table 16 compares the LR statistics between Poisson and NB-INGARCH specifications. For theft—the most severely overdispersed series (var/mean = 10.8, $\hat{\phi} = 4.21$)—the LR statistic drops by 87% (42.38 to 5.35), with the p-value rising from $< 10^{-10}$ to 0.021. The Poisson assumption severely inflates significance; under NB, the negative temperature effect remains significant at 5%, but the effect shrinks ($\hat{\gamma} = -0.140$ vs. -0.165), and confidence is substantially reduced. Second, for sexual assault, NB confirms inconclusiveness (LR = 3.13, $p = 0.077$), consistent with weak identification and bootstrap results. Third, for narcotics, the LR drops 68% but remains highly significant ($p = 1.1\times10^{-7}$), indicating robustness to distributional misspecification.

Table 16. Comparison of Poisson and NB-INGARCH LR statistics.

Crime	Poisson LR	NB LR	NB p-value	LR drop	NB conclusion
S. Assault	6.29	3.13	0.0770	−50%	Not significant
Theft	42.38	5.35	0.0207	−87%	Significant, attenuated
Narcotics	86.91	28.19	1.1×10^{-7}	−68%	Highly significant

These results demonstrate that the Poisson assumption, while standard, can severely distort inference when overdispersion is strong. NB-INGARCH should be treated as a necessary robustness check. The causal direction (negative temperature effect) is preserved, but magnitudes and confidence are more conservative.

Table 17 synthesizes evidence across all specifications. For theft and narcotics, the negative temperature effect is robust: Poisson, bootstrap, and NB-INGARCH all reject the null, though NB-INGARCH substantially attenuates the evidence for theft. For sexual assault, all specifications converge on inconclusiveness—whether due to weak identification (Poisson/bootstrap) or distributional robustness (NB). Our results show that LBE provides reliable inference even when CMLE fails due to numerical singularity, while routine distributional diagnostics (e.g., NB-INGARCH) are indispensable for addressing overdispersion.

Table 17. Summary of causal evidence across specifications.

Crime	$\hat{\gamma}$	Poisson p-value	Bootstrap significance	NB-INGARCH	Prior robustness	Overdispersion (var/mean)
S. Assault	−0.005	0.012–0.041	Marginal	Not sig.	Requires bootstrap	2.0 (moderate)
Theft	−0.165	$< 10^{-7}$	Robust	Sig., attenuated	Highly robust	10.8 (severe)
Narcotics	−0.058	$< 10^{-7}$	Robust	Highly sig.	Robust	5.4 (strong)

5. Conclusions

This paper proposes an LBE for Poisson INGARCH models, bridging the gap between frequentist efficiency and Bayesian regularization for count time series. The main contributions are fourfold: (i) A closed-form LBE that fuses prior information with CMLE, avoiding posterior sampling; (ii) theoretical MSEM dominance of LBE over CMLE; (iii) a regression-based causality test with valid asymptotic guarantees; and (iv) empirical validation showing LBE's numerical stability in challenging likelihood geometries.

Leveraging the conditional mean structure of the Poisson INGARCH model, we formulate a regression framework with martingale difference errors that propagates exogenous covariate information under causal constraints. Under standard regularity conditions, we establish that the LSE-based test statistic converges to a chi-square distribution as its asymptotic null distribution. Simulation results confirm LBE's superiority over CMLE in small-to-moderate samples, with prior robustness checks showing estimates remain predominantly data-driven across a wide range of prior strengths.

Applying both methods to a Chicago crime dataset, we find that higher temperatures have an inhibitory effect on theft and drug crimes. For sexual assault, weak identification and severe overdispersion preclude definitive causal conclusions, but LBE provides critical numerical stability when CMLE fails due to likelihood singularity. NB robustness analysis further underscores that Poisson misspecification can severely inflate significance, highlighting the need for routine distributional diagnostics in count time-series applications. Together, these results establish a solid theoretical foundation and demonstrate the practical utility of LBE for count time-series analysis.

While this study makes notable progress in estimation and causality testing for Poisson INGARCH models, several directions merit deeper exploration:

Model extensions: Future work could extend the framework to zero-inflated Poisson or NB-INGARCH models to accommodate zero inflation and extreme overdispersion prevalent in empirical data.

Methodological improvements: More objective prior specification strategies based on historical data or robust Bayesian methods could minimize subjective input and enhance LBE's robustness across diverse samples.

Framework generalization: The linear Bayes principle is inherently generic and could be extended to the integer-valued autoregressive (INAR) framework, broadening its applicability to count time series with thinning-based dependence structures.

Computational benchmarking: Systematic comparisons against MCMC and variational Bayes would quantify LBE's computational gains and finite-sample trade-offs relative to full Bayesian methods.

High-dimensional applications: Integrating LBE with regularization or graphical models could enable high-dimensional variable selection and network causal inference. Time-varying parameter extensions could capture evolving causal effects.

Sustained investigation along these directions will advance count time-series methodology and enable it to address increasingly complex, high-dimensional challenges in real-world data science.

Author contributions

Zhongxiu Ma: Conceptualization, methodology, software, formal analysis, writing–original draft; Zhanshou Chen: Conceptualization, writing–review & editing (introduction logic and English polishing); Peiyan Qi: Writing–review & editing, validation. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

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

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 12161072), the Natural Science Foundation of Qinghai Province (No. 2024-ZJ-933), and the Fundamental Research Program of Shanxi Province (202203021211204).

Conflict of interest

No potential conflict of interest was reported by the authors.

References

1. D. Andrade, Robust variational Gaussian process regression for count data with the trimmed marginal likelihood, *Stat. Comput.*, **36** (2026), 139. https://doi.org/10.1007/s11222-026-10895-9
2. E. Bacry, I. Mastromatteo, J. Muzy, Hawkes processes in finance, *Market Microstruc. Liquidity*, **1** (2015), 1550005. https://doi.org/10.1142/S2382626615500057
3. Y. Bai, L. Wang, Bayes estimator for the inequality-constrained regression model, *Commun. Stat.-Simul. C.*, **54** (2025), 3401–3414. https://doi.org/10.1080/03610918.2024.2349178
4. Y. Cui, F. Zhu, A new bivariate integer-valued GARCH model allowing for negative cross-correlation, *Test*, **27** (2018), 428–452. https://doi.org/10.1007/s11749-017-0552-4
5. T. Chen, *Linear Bayes estimator of the extreme value distribution based on Type II censored samples*, Beijing: Beijing Jiaotong University, 2021. https://doi.org/10.1080/03610918.2021.1963454
6. E. Cohn, Weather and crime, *Brit. J. Criminol.*, **30** (1990), 51–64. https://doi.org/10.1093/oxfordjournals.bjc.a047980
7. W. Chan, A correlated bivariate Poisson jump model for foreign exchange, *Empir. Econ.*, **28** (2003), 669–685. https://doi.org/10.1007/s00181-003-0153-9
8. J. Corcoran, R. Zahnow, Weather and crime: A systematic review of the empirical literature, *Crime Sci.*, **11** (2022), 16. https://doi.org/10.1186/s40163-022-00179-8
9. C. Chen, S. Lee, Bayesian causality test for integer-valued time series models with applications to climate and crime data, *J. Roy. Stat. Soc. C-Appl.*, **66** (2017), 797–814. https://doi.org/10.1111/rssc.12200

10. J. Choi, Y. Ni, Model-based causal discovery for zero-inflated count data, *J. Mach. Learn. Res.*, **24** (2023), 1–32.
11. Y. Cheng, L. Li, T. Xiao, Z. Li, J. Suo, K. He, et al., *Cuts+: High-dimensional causal discovery from irregular time-series*, In: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, **38** (2024), 11525–11533. https://doi.org/10.1609/aaai.v38i10.29034
12. R. Davis, K. Fokianos, S. Holan, H. Joe, J. Livsey, R. Lund, et al., Count time series: A methodological review, *J. Am. Stat. Assoc.*, **116** (2021), 1533–1547. https://doi.org/10.1080/01621459.2021.1904957
13. R. Davis, H. Liu, Theory and inference for a class of nonlinear models with application to time series of counts, *Stat. Sinica*, **26** (2016), 1673–1707. https://www.jstor.org/stable/44114353
14. J. Davidson, *Stochastic limit theory: An introduction for econometricians*, Oxford: OUP Oxford, 1994.
15. B. Efron, C. Morris, Stein's estimation rule and its competitors—An empirical Bayes approach, *J. Am. Stat. Assoc.*, **68** (1973), 117–130. https://doi.org/10.1080/01621459.1973.10481350
16. R. Engle, J. Russell, Autoregressive conditional duration: A new model for irregularly spaced transaction data, *Econometrica*, **66** (1998), 1127–1162.
17. R. Ferland, A. Latour, D. Oraichi, Integer-valued GARCH process, *J. Time Ser. Anal.*, **27** (2006), 923–942. https://doi.org/10.1111/j.1467-9892.2006.00496.x
18. S. Field, The effect of temperature on crime, *Brit. J. Criminol.*, **32** (1992), 340–351. https://doi.org/10.1093/oxfordjournals.bjc.a048222
19. K. Fokianos, A. Rahbek, D. Tjøstheim, Poisson autoregression, *J. Amer. Stat. Assoc.*, **104** (2009), 1430–1439. https://doi.org/10.1198/jasa.2009.tm08270
20. M. Goldstein, The linear Bayes regression estimator under weak prior assumptions, *Biometrika*, **67** (1980), 621–628. https://doi.org/10.1093/biomet/67.3.621
21. M. Goldstein, General variance modifications for linear Bayes estimators, *J. Am. Stat. Assoc.*, **78** (1983), 616–618. https://doi.org/10.1080/01621459.1983.10478019
22. H. Gao, Z. Chen, F. Li, Linear Bayesian estimation of misrecorded Poisson distribution, *Entropy*, **26** (2024), 62. https://doi.org/10.3390/e26010062
23. C. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37** (1969), 424–438.
24. J. Hartigan, Linear Bayesian methods, *J. Roy. Stat. Soc. B*, **31** (1969), 446–454. https://doi.org/10.1111/j.2517-6161.1969.tb00804.x
25. J. Higgins, C. Tsokos, A study of the effect of the loss function on Bayes estimates of failure intensity, MTBF, and reliability, *Appl. Math. Comput.*, **6** (1980), 145–166. https://doi.org/10.1016/0096-3003(80)90039-9
26. B. Hansen, A. A. Pacheco, M. Russo, R. De Vito, Fast variational inference for Bayesian factor analysis in single and multi-study settings, *J. Comput. Graph. Stat.*, **34** (2025), 96–108. https://doi.org/10.1080/10618600.2024.2356173
27. J. Kang, S. Lee, Parameter change test for Poisson autoregressive models, *Scand. J. Stat.*, **41** (2014), 1136–1152. https://doi.org/10.1111/sjos.12088

28. M. Liu, F. Zhu, J. Li, C. Sun, A systematic review of INGARCH models for integer-valued time series, *Entropy*, **25** (2023), 922. https://doi.org/10.3390/e25060922
29. A. M. Novelo, C. Anderson, Climate change and psychology: Effects of rapid global warming on violence and aggression, *Curr. Clim. Change Rep.*, **5** (2019), 36–46. https://doi.org/10.1007/s40641-019-00121-2
30. A. M. Novelo, C. Anderson, *Climate change and human behavior: Impacts of a rapidly changing climate on human aggression and violence*, Cambridge: Cambridge University Press, 2022. https://doi.org/10.1017/9781108953078
31. W. Neuhaus, Choice of statistics in linear Bayes estimation, *Scand. Actuar. J.*, **1985** (1985), 1–26. https://doi.org/10.1080/03461238.1985.10413774
32. K. Papaspyropoulos, D. Kugiumtzis, On the validity of Granger causality for ecological count time series, *Econometrics*, **12** (2024), 13. https://doi.org/10.3390/econometrics12020013
33. L. Qian, Q. Zuo, W. Liu, H. Zhu, MSDG: A multiscale dynamic graph neural network for inferring dynamic Granger causality in multivariate time series, *Inform. Sci.*, **681** (2026), 123287. https://doi.org/10.1016/j.ins.2026.123287
34. E. Stein, Interpolation of linear operators, *T. Am. Math. Soc.*, **83** (1956), 482–492. https://doi.org/10.1090/s0002-9947-1956-0082586-0
35. Z. Su, F. Zhu, S. Liu, Local influence analysis in the softplus INGARCH model, *Test*, **33** (2024), 951–985. https://doi.org/10.1007/s11749-024-00930-0
36. G. Trenkler, L. Wei, The Bayes estimator in a misspecified linear regression model, *Test*, **5** (1996), 113–123. https://doi.org/10.1007/bf02562684
37. P. Tsamtsakiri, D. Karlis, On Bayesian model selection for INGARCH models via trans-dimensional Markov chain Monte Carlo methods, *Stat. Model.*, **23** (2023), 81–98. https://doi.org/10.1177/1471082x211034705
38. C. Weiß, Stationary count time series models, *WIREs Comput. Stat.*, **13** (2021), e1502. https://doi.org/10.1002/wics.1502
39. C. Weiß, F. Zhu, A. Hoshiyar, Softplus INGARCH models, *Stat. Sinica*, **32** (2022), 1099–1120. https://doi.org/10.5705/ss.202020.0353
40. H. Xuan, L. Maestrini, F. Chen, C. Grazian, Stochastic variational inference for GARCH models, *Stat. Comput.*, **34** (2024), 45. https://doi.org/10.1007/s11222-023-10356-7
41. F. Zhu, Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models, *J. Math. Anal. Appl.*, **389** (2012a), 58–71. https://doi.org/10.1016/j.jmaa.2011.11.042
42. F. Zhu, Zero-inflated Poisson and negative binomial integer-valued GARCH models, *J. Stat. Plan. Infer.*, **142** (2012b), 826–839. https://doi.org/10.1016/j.jspi.2011.10.002