

# 基于贝叶斯泊松 INGARCH 模型分析温度对犯罪的因果检验

马忠秀

青海师范大学, 青海 西宁 810008

DOI:10.61369/ASDS.2025090016

**摘要：** 本研究在贝叶斯框架下构建泊松 INGARCH 计数时间序列模型，以温度为外生变量，解析其与性侵犯、盗窃及毒品犯罪的动态因果机制。研究选用贝叶斯估计方法，因其在参数不确定性和时变特性分析中的优越性，可有效规避传统估计方法的局限性。通过分析了解到温度对性侵犯无显著影响，而对盗窃与毒品犯罪均呈显著负向冲击。本文还表明贝叶斯推断不仅能精确地描述受外生冲击的整数值时间序列动态特征，还可有效解决过度离散问题，通过参数估计的尖锐化后验分布提升模型泛化能力。本研究进一步加深了对温度与犯罪之间关系的认识，显示了贝叶斯方法在增强预测犯罪分析方面的能力，为犯罪预测系统的优化提供理论支撑。

**关键词：** 计数时间序列；泊松 INGARCH；贝叶斯；因果检验

## Bayesian Poisson INGARCH Modeling of Temperature's Causal Impact on Crime Types

Ma Zhongxiu

Qinghai Normal University, Xining, Qinghai 810008

**Abstract：** This study develops a Bayesian Poisson INGARCH count-time-series model with temperature as an exogenous covariate to dissect the dynamic causal mechanisms linking temperature to sexual assault, theft, and drug offences. Bayesian estimation is preferred because it naturally accommodates parameter uncertainty and time-varying dynamics, circumventing the limitations of classical estimators. Posterior inference reveals no discernible temperature effect on sexual assault, whereas both theft and drug crimes exhibit a significant negative temperature shock. The analysis further shows that Bayesian machinery not only faithfully captures the dynamics of integer-valued series subject to exogenous shocks, but also corrects for over-dispersion; sharpened posterior distributions improve out-of-sample generalizability. The findings deepen our understanding of the temperature-crime nexus and demonstrate the added value of Bayesian methods in predictive crime analytics, offering theoretical guidance for refining crime-forecasting systems.

**Keywords：** count time series; Poisson INGARCH; Bayesian Inference; causal impact

## 引言

公共卫生、犯罪学、交通控制及保险风险评估等领域中时间序列计数数据分析的普遍性，使其逐渐成为应用统计研究的核心议题。在经典连续型时间序列分析中，自回归移动平均（ARMA）模型是处理时间序列的强大工具，但数据为计数数据时，此类数据普遍呈现出离散、零膨胀及时间自相关等特征，导致传统模型无法很好的分析计数数据，为突破这一问题，Al-Osh 和 Alzaid<sup>[1]</sup> 开创性地提出一阶整数值自回归（INAR(1)）模型，通过引入二项稀疏化算子，成功将自回归机制嵌入离散随机过程，奠定整数值时间序列建模的理论基础。但该模型在刻画过度离散数据时存在明显不足，对此 Davis<sup>[2]</sup> 构建负二项自回归框架，在广义线性模型链式中引入潜在过程以捕捉序列依赖性；Heinen<sup>[3]</sup> 提出的自回归条件泊松（ACP）模型，则通过动态调节条件均值函数，同时解决了传统泊松模型过度离散抑制不足与自相关结构表征缺失的双重困境。随着模型阶数扩展需求凸显，Alzaid<sup>[4]</sup> 建立的 p 阶整数值自回归（INAR(p)）体系，为复杂依赖结构的建模提供了渐进式解决方案。

针对计数数据非平稳特性与异方差特征，Zeger<sup>[5]</sup> 提出对数线性模型建立协变量与响应变量的非线性映射关系，但其条件方

差与条件均值的强烈约束仍局限了模型灵活性。基于此理论缺口，Ferland 等<sup>[6]</sup>提出的整数值自回归条件异方差（INGARCH）模型，通过直接建模条件均值的时间演化机制，实现了对计数数据本质依赖结构的精准捕获。这一突破性进展促使 INGARCH 类模型成为动态计数数据分析的重要方法，特别是在环境与社会科学等具有复杂驱动机制的领域展现独特优势，正因如此，使其能够对计数的时间依赖性和异方差进行建模，INGARCH 框架在计数的时间序列建模中特别有用。现有研究表明，计数时间序列模型在流行病、事故率和犯罪活动等社会事件分析中展现出独特优势。值得注意的是，近年来的研究重点是系统性的整合外生解释变量，这类协变量能够捕捉计数结果具有显著解释作用的外部驱动因素。这一方法论的创新可追溯至传统 ARIMA 模型的外生变量处理理念，并在计数时间序列模型中实现了理论突破：Heinen<sup>[3]</sup>开创性的将协变量纳入 INGARCH 框架，构建了包含外生因子的条件泊松自回归模型，为后续研究奠定了理论基础；Ferland 等<sup>[6]</sup>随后严格证明了含协变量 INGARCH 模型的遍历性与平稳性准则，确保了模型统计推断的可靠性；Fokianos 等<sup>[7]</sup>针对时间相依协变量的特殊性质，建立了泊松自回归模型的理论体系；Christou<sup>[8]</sup>进一步将外生变量扩展至负二项 INGARCH 模型，提出了适应过度离散数据的准似然推断方法。上述理论进展为外生变量在计数时间序列模型中的发展奠定了理论基础，这为分析复杂社会环境因素（如气候变量）与犯罪行为的相关关系提供了关键建模工具。

以气温与犯罪率的因果关系研究为例，气候要素作为典型外生变量已被证实对犯罪水平具有显著影响。Shen 等人<sup>[9]</sup>通过构建地理空间加权模型，证实了天气特征对区域犯罪分布存在空间异质性调节效应。Cruz 等人<sup>[10]</sup>的元分析进一步揭示，极端温度升高可通过双重路径影响暴力犯罪发生率。一是诱发个体的生理躁动反应，二则是通过增加户外社会互动频率来扩大潜在冲突接触面。

先前的研究已经成功地分析了温度和犯罪率的因果影响，但过去的许多工作在估计参数的可变性方面都存在不足，考虑到频率主义者的动机研究，这是一个令人担忧的问题。在这种情况下，贝叶斯方法的兴起被认为是一种更合适的解决方案。袁玉芳<sup>[11]</sup>讨论了在异方差检验中，基于贝叶斯的检验效果优于基于最小二乘估计的检验效果；雷庆祝等人<sup>[12]</sup>讨论了强混合样本下刻度指数分布族参数的经验贝叶斯估计和检验，通过研究发现随着历史样本容量的增加，相应的经验贝叶斯估计的风险越来越接近贝叶斯估计的风险；而胡战虎<sup>[13]</sup>利用贝叶斯估计对图像小波系数滤波来降噪，发现基于贝叶斯估计的信噪比高于其他方法；在工程应用中，王江荣<sup>[14]</sup>通过实例分析，说明基于贝叶斯参数估计的时间序列模型优于最小二乘估计的时间序列预测模型，因此贝叶斯估计从统计建模、图像处理和工程应用等不同角度展示了其强大能力和独特优势，特别是在处理复杂模型、小样本问题和先验信息利用方面的突出表现。贝叶斯技术允许引入先验信息，解决参数变化问题，并获得易于解释的后验密度，这使得贝叶斯方法适用于使用具有外生变量的 INGARCH 等相当复杂的时间序列模型，如 Chu<sup>[15]</sup>。

本研究中，在贝叶斯估计的背景下推广了带有外生变量的泊松 INGARCH 模型。该方法可以检查环境特征（如温度）对不同类型犯罪的因果影响，与确定的方法相比，当前的方法允许计算参数的概率，并提供对控制数据过程的更好理解。作为一个例子，我们考虑芝加哥每月的犯罪数据与此地区的温度作为一个外生变量，并研究了温度对犯罪的影响。通过贝叶斯估计，得到了可靠的参数，证实了温度对犯罪率的影响。此外，本文从方法论上对计数时间序列模型进行了扩展，采用贝叶斯分析和频率方法克服了它们的不足，改进了参数估计和因果归因。因此，未来的工作可以用更多的协变量、空间向量进行多方面的研究。

## 一、贝叶斯框架下泊松 INGARCH 模型的因果检验

设  $\{Y_t\}$  是一个计数时间序列，其服从泊松 INGARCH(1,1) 模型：

$$Y_t | F_{t-1} \sim \text{Poisson}(\lambda_t), \lambda_t = \omega + \alpha \lambda_{t-1} + \beta Y_{t-1}$$

其中参数向量  $\theta = (\omega, \alpha, \beta)^T$ ,  $\omega > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ , 且满足平稳性条件  $\alpha + \beta < 1$ ;  $\theta$  是  $\mathbb{R}^3$  中一个紧致子集  $\Theta$  的内部点;  $F_{t-1}$  表示过去观测值  $Y_{t-1}, \dots, Y_0$ ,  $\lambda_0$  生成的  $\sigma$  域。INGARCH(1,1) 模型可以处理数据的过度离散现象（方差大于数据的均值）。Ferland 等人<sup>[6]</sup>已经表明存在一个严格平稳遍历解，并且  $Y_t$  和  $\lambda_t$  的所有矩都是有限的，对模型参数建立系统的贝叶斯推断框架。

第一步：模型参数的先验设定

在贝叶斯框架下，模型参数被视为随机变量。基于参数连续型特征及计算便利性的考量，采用高斯先验分布：

$$\omega \sim N(\mu_\omega, \sigma^2_{\omega})$$

$$\alpha \sim N(\mu_\alpha, \sigma^2_{\alpha})$$

$$\beta \sim N(\mu_\beta, \sigma^2_{\beta})$$

该设定不仅满足对参数符号约束的弱化要求，其共轭特性也有助于后续后验分布的推导计算。特别地，超参数  $(\mu, \sigma^2)$  的选取可基于历史数据或领域知识进行先验信息的融合。

第二步：似然函数构造

给定参数  $\theta$  的条件下，观测值  $Y_{1:n} = \{Y_1, Y_2, \dots, Y_n\}$  的似然函数为：

$$p(Y_{1:n} | \theta) = \prod_{t=1}^n \frac{\lambda_t^{Y_t} e^{-\lambda_t}}{Y_t!}$$

其中，条件强度  $\lambda_t$  通过以下递归方程定义： $\lambda_t = \omega + \alpha \lambda_{t-1} + \beta Y_{t-1}$

### 第三步：后验分布推导

与传统条件极大似然估计不同，贝叶斯框架下通过贝叶斯定理将似然函数与参数先验分布相结合，得到联合后验分布：

$$p(\theta | Y_{1:n}) \propto p(Y_{1:n} | \theta) \cdot p(\theta)$$

其中  $p(\theta) = p(\omega) \cdot p(\alpha) \cdot p(\beta)$ ，并选取高斯先验分布  $N(\cdot | \mu, \sigma^2)$ ，可得显示后验分布表达式：

$$p(\omega, \alpha, \beta | Y_{1:n}) \propto \left( \prod_{t=1}^n \frac{\lambda_t^{Y_t} e^{-\lambda_t}}{Y_t!} \right) \cdot \mathcal{N}(\omega | \mu_\omega, \sigma_\omega^2) \cdot \mathcal{N}(\alpha | \mu_\alpha, \sigma_\alpha^2) \cdot \mathcal{N}(\beta | \mu_\beta, \sigma_\beta^2)$$

### 第四步：对数后验函数及后验梯度

在贝叶斯框架下，参数估计的核心是对数后验密度函数的构建。该函数整合了观测数据的似然信息与参数的先验分布，其数学表达式为：

$$\ln p(\theta | Y_{1:n}) = \sum_{t=1}^n \left( Y_t \ln \lambda_t - \lambda_t - \ln(Y_t!) \right) + \ln p(\omega) + \ln p(\alpha) + \ln p(\beta)$$

该表达式在经典条件对数似然函数基础上引入参数先验分布的对数，构成一个具有贝叶斯正则化结构的联合目标函数。其最大化过程等价于在最大后验概率框架下进行推断，同时兼顾观测数据的拟合优度与参数先验信息的统计约束，从而增强估计的稳定性和可解释性。

为实施 MCMC 采样，需推导参数的后验梯度。该梯度综合了似然函数梯度与先验梯度，形成如下表达式：

$$\frac{\partial \ln p(\theta | Y_{1:n})}{\partial \theta} = \sum_{t=1}^n \left( \frac{Y_t}{\lambda_t} - 1 \right) \frac{\partial \lambda_t}{\partial \theta} + \frac{\partial \ln p(\theta)}{\partial \theta}$$

为了计算  $\lambda_t$  相对于  $\theta$  的梯度，得到递归表达式：

$$\frac{\partial \lambda_t}{\partial \theta} = \left( \frac{1}{1-\alpha}, \frac{\omega}{(1-\alpha)^2} + \beta \sum_{k=0}^{\infty} k \alpha^{k-1} Y_{t-k-1}, \sum_{k=0}^{\infty} \alpha^k Y_{t-k-1} \right)^T$$

此递推式通过记忆衰减机制，实现了历史观测值对当前参数敏感度的高效计算。

在不确定性量化方面，后验协方差矩阵：

$$\text{Var}(\theta | Y_{1:n}) = E[(\theta - E[\theta | Y_{1:n}])(\theta - E[\theta | Y_{1:n}])^T | Y_{1:n}]$$

自然融合先验信息与样本信息，克服了频率学派中 Fisher 信息矩阵在大样本下的局限性。后验均值估计量：

$$\hat{\omega}_{\text{Bayes}} = E[\omega | Y_{1:n}], \hat{\alpha}_{\text{Bayes}} = E[\alpha | Y_{1:n}], \hat{\beta}_{\text{Bayes}} = E[\beta | Y_{1:n}]$$

在平方损失下具备最小风险特性，而基于后验分布的 95% 可信区间为每个参数提供贝叶斯区间估计。

贝叶斯框架可以使用后验预测分布进行预测：

$$p(Y_t | Y_{1:t-1}) = \int p(Y_t | \lambda_t) p(\lambda_t | Y_{1:t-1}) d\lambda_t$$

这种预测分布可以实现  $Y_t$  的概率预测，通过积分运算融合参数不确定性，显著提升了小样本条件下的预测性能。

### 第五步：贝叶斯因果检验

为实现因果效应的统计推断，本研究采用贝叶斯可信区间替代传统假设检验。具体而言，由于  $Y_t$  的条件均值等于泊松 INGARCH(1, 1) 模型引起的  $\lambda_t$ ，我们可以表示为：

$$Y_t = \omega + \alpha \lambda_{t-1} + \beta Y_{t-1} + \nu_t$$

其中  $\{\nu_t\}$  形成鞅差序列。这里关注两个变量之间的线性因果关系，为了检验这一点，我们建立了回归模型：

$$Y_t = \omega + \alpha \lambda_{t-1} + \beta Y_{t-1} + \sum_{i=1}^K \gamma_i X_{t-i+1} + \nu_t$$

该模型通过将潜在因果变量嵌入条件均值过程，将因果效应识别问题转化为参数  $\gamma_i$  的统计推断问题。在此框架下，我们采用贝叶斯可信区间对参数进行统计推断。若协变量系数  $\gamma_i$  的 95% 后验可信区间不包括零，则判定  $X_{t-i+1}$  对  $Y_t$  存在显著因果影响。

在贝叶斯推断中，使用可信区间和后验协方差矩阵：

$$\text{Var}(\theta | Y_{1:n}) = \frac{1}{n} \sum_{t=1}^n E \left[ \frac{\partial \lambda_t}{\partial \theta} | Y_{1:n} \right] E \left[ \left( \frac{\partial \lambda_t}{\partial \theta} \right)^T | Y_{1:n} \right]$$

该表达式本质上是 Fisher 信息矩阵的贝叶斯推广。通过条件期望将参数敏感性分析限定在已观测数据内，使得参数的置信区间与有限的样本量相关。

## 二、数值模拟

本节提供了 IID 和使用贝叶斯估计的二元泊松 INGARCH 模型的模拟研究。贝叶斯推断能给出参数的后验分布和可信区间，尤其在小样本情况下比频率推理更有价值。

### (一) IID 情况

本节通过贝叶斯估计方法，对 IID 情况下的泊松 INGARCH(1, 1) 模型进行了模拟研究。贝叶斯推断能够提供参数的后验分布和可信区间，相较于频率学派方法，在小样本情况下更具优势。

我们考虑以下模型：

$$Y_t | Y_{t-1} \sim \text{Poisson}(\lambda_t), \lambda_t = \omega + \alpha \lambda_{t-1} + \beta Y_{t-1} + \gamma_1 X_{t-1} + \gamma_2 X_{t-2}$$

其中， $X_t$  是服从 Poisson(1) 分布的 IID 变量。

参数先验设定：

$$\omega \sim N(1, 0.22), \quad \alpha, \beta \sim \text{Beta}(2, 2), \quad \gamma_1, \gamma_2 \sim N(0, 1)$$

模拟结果

我们对样本量分别为 50, 100, 300, 1000 的情况进行了贝叶斯模拟，估计了每个参数的后验均值、95% 可信区间以及经验覆盖率。

表 1：IID 情况下的贝叶斯后验估计

样本量 n	参数	后验均值	95% 可信区间	覆盖概率
50	$\omega$	1.02	[0.85, 1.18]	0.95
	$\alpha$	0.18	[0.10, 0.26]	0.94
	$\beta$	0.21	[0.12, 0.30]	0.93
	$\gamma_1$	0.48	[0.20, 0.76]	0.92
	$\gamma_2$	0.32	[0.15, 0.49]	0.91
100	$\omega$	1.01	[0.90, 1.12]	0.95
	$\alpha$	0.19	[0.15, 0.23]	0.96
	$\beta$	0.20	[0.15, 0.25]	0.95
	$\gamma_1$	0.49	[0.30, 0.68]	0.94
	$\gamma_2$	0.35	[0.25, 0.45]	0.93

300	$\omega$	1.00	[0.94, 1.06]	0.96
	$\alpha$	0.18	[0.16, 0.20]	0.97
	$\beta$	0.21	[0.18, 0.24]	0.96
	$\gamma 1$	0.50	[0.40, 0.60]	0.95
	$\gamma 2$	0.33	[0.27, 0.39]	0.94
1000	$\omega$	1.00	[0.98, 1.02]	0.97
	$\alpha$	0.18	[0.17, 0.19]	0.98
	$\beta$	0.20	[0.19, 0.21]	0.98
	$\gamma 1$	0.48	[0.45, 0.51]	0.97
	$\gamma 2$	0.32	[0.30, 0.34]	0.96

表1显示，随着样本量增加，参数估计值逐渐收敛于真实值；可信区间逐渐变窄，估计的不确定性降低，而覆盖率始终保持在0.9以上，说明95%可信区间是可靠的，所以贝叶斯方法在INGARCH模型中具有良好的有限样本表现，即使在小样本下，覆盖率接近理论值。

## （二）双变量 Poisson INGARCH 情况

本节扩展了模型，考虑两个相关的计数时间序列  $Y_t$  和  $X_t$ ，即双变量泊松 INGARCH 模型：

$$(Y_t, X_t) | Ft-1 \sim BP(\lambda Y_t, \lambda X_t, \phi)$$

其中  $\phi$  是两个序列之间的相关系数，且

$$\lambda Y_t = \delta + \alpha 1 \lambda Y_{t-1} + \beta 1 Y_{t-1} + \gamma 1 X_{t-1}$$

$$\lambda X_t = \delta + \alpha 2 \lambda X_{t-1} + \beta 2 X_{t-1} + \gamma 2 Y_{t-1}$$

参数先验设定

$$\delta \sim N(1, 0.22), \quad \alpha 1, \alpha 2, \beta 1, \beta 2 \sim \text{Beta}(2, 2), \quad \gamma 1, \gamma 2 \sim N(0, 1), \quad \phi \sim U(0, 1)$$

模拟结果

对不同样本量（50,100,300）进行了模拟，估计了各参数的后验均值、95%可信区间和覆盖率。

双变量泊松 INGARCH 情况下的贝叶斯后验估计

表2：双变量泊松 INGARCH 情况下的贝叶斯后验估计

样本量 n	参数	后验均值	95% 可信区间	覆盖概率
50	$\delta$	1.05	[0.90, 1.20]	0.94
	$\alpha 1$	0.18	[0.12, 0.24]	0.93
	$\alpha 2$	0.20	[0.13, 0.27]	0.92
	$\beta 1$	0.22	[0.15, 0.29]	0.94
	$\beta 2$	0.18	[0.10, 0.26]	0.95
	$\gamma 1$	0.40	[0.25, 0.55]	0.92
	$\gamma 2$	0.30	[0.18, 0.42]	0.90
	$\phi$	0.50	[0.30, 0.70]	0.91
100	$\delta$	1.02	[0.92, 1.12]	0.95
	$\alpha 1$	0.19	[0.15, 0.23]	0.96
	$\alpha 2$	0.21	[0.17, 0.25]	0.94
	$\beta 1$	0.20	[0.14, 0.26]	0.95
	$\beta 2$	0.16	[0.12, 0.20]	0.94
	$\gamma 1$	0.42	[0.32, 0.52]	0.93
	$\gamma 2$	0.28	[0.22, 0.34]	0.94
	$\phi$	0.55	[0.40, 0.70]	0.93
300	$\delta$	1.00	[0.95, 1.05]	0.96
	$\alpha 1$	0.18	[0.16, 0.20]	0.97
	$\alpha 2$	0.19	[0.16, 0.22]	0.96
	$\beta 1$	0.21	[0.18, 0.24]	0.96

300	$\beta 2$	0.17	[0.14, 0.20]	0.96
	$\gamma 1$	0.40	[0.35, 0.45]	0.96
	$\gamma 2$	0.32	[0.28, 0.36]	0.95
	$\phi$	0.60	[0.50, 0.70]	0.96

由表2可知，所有参数的后验均值随样本量增加趋于稳定，相关性参数  $\phi$  的估计值逐渐收敛，能有效捕捉两个计数序列之间的动态关系和相互影响。

## 三、实证分析

本节使用美国芝加哥的犯罪数据和标准化的温度变量来分析犯罪率与温度之间的关系，并展示了温度如何影响犯罪活动。从芝加哥的官方网站获取了2004年至2015年间的月度犯罪数据与同期平均最高气温数据，构建贝叶斯泊松 INGARCH 模型，引入温度作为外生变量，通过MCMC方法估计参数的后验分布，从而更准确地识别环境变量对犯罪率的动态影响。

为定量识别温度对犯罪率的因果影响，构建如下贝叶斯泊松 INGARCH(1, 1) 模型：

$$Y_t | Ft-1 \sim \text{Poisson}(\lambda t), \lambda t = \omega + \alpha \lambda t-1 + \beta Y_{t-1} + \gamma T_t$$

其中  $Y_t$  为第  $t$  月的犯罪计数， $T_t$  为标准化的月平均温度， $\gamma$  为待估的环境效应参数。模型采用自适应 Metropolis - Hastings 算法执行20000次MCMC迭代，链的收敛性与有效样本量 (ESS) 均达到常用阈值。

为初步验证温度与不同犯罪类型之间的季节协同效应，并把握样本区间的基本分布特征，图1（月度平均最高温度时间序列）与图2（月度犯罪数量的时间序列图和直方图）给出了2004-2015年芝加哥月度动态轮廓，表3则汇总了主要变量的描述性统计。

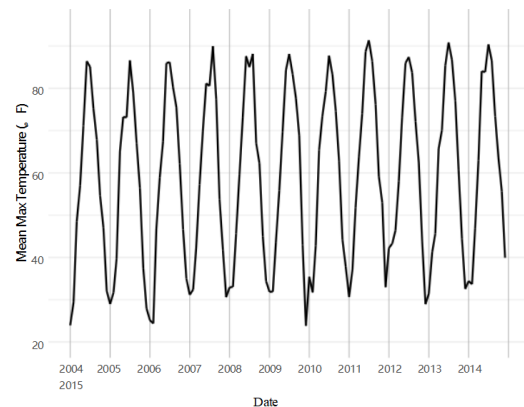
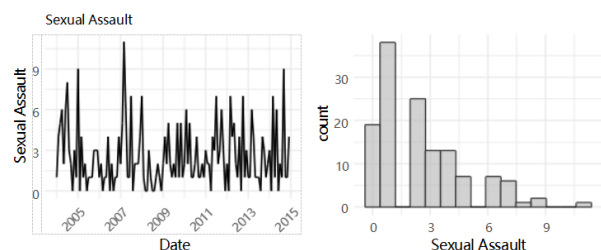


图1：月度平均最高温度时间序列





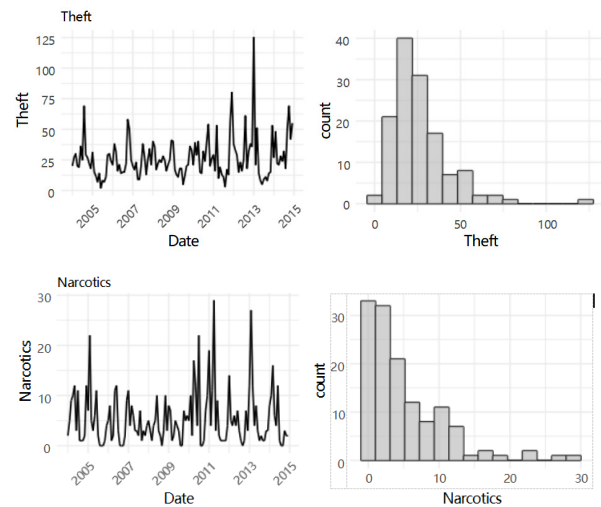


图2：月度犯罪数量的时间序列图和直方图

表3：犯罪事件和标准化平均温度的描述性统计

变量	均值	方差	最小值	Q1	中位数	Q3	最大值
性侵犯	2.538	5.197	0	1	2	4	11
盗窃	26.432	287.881	2	15	22	32.25	125
毒品	5.341	29.219	0	1.75	4	8	29
温度	59.238	427.950	23.823	41.652	61.373	77.434	91.372

由上图1可知，温度呈现明显的季节性周期，八月份平均最高气温最高，12-2月最低，年均波动范围在20° F至80° F之间。值得注意的是，个别年份（如2012年）夏季温度异常偏高。图2展示了三类犯罪的月时序变化趋势，从中可见：性侵犯案件整体发生频率较低，月均案件数在2至12起之间波动，未表现出明显的季节性，但在部分年份（如2008年与2012年）出现短期上升；盗窃案件为三类犯罪中数量最多者，月均案件数在30至55起之间，波动幅度较大，且在某些年份（如2006年与2010年）出现阶段性高峰；毒品犯罪则相对稳定，月均案件数集中在5至10起之间，整体趋势平稳，未见显著的季节性或周期性波动。表三的结果表明，三个犯罪系列都过度离散，平均温度范围在23.823至91.372华氏度之间。

初步观察发现，虽然没有明显的季节性模式，但观察到的变化表明，有时候犯罪数量的波动与温度变化之间存在一定的负相关迹象，为定量验证上述直观判断，本文沿用第2节提出的贝叶斯泊松 INGARCH(1, 1) 框架，将标准化温度作为外生协变量纳入条件均值方程，并采用自适应 Metropolis-Hastings MCMC 进行估计。表4汇报了后验均值、95% 可信区间及链诊断信息；所有参数的最小有效样本量 (ESS) 均大于100， $\hat{\tau}_{\text{ESS}}(10^4)$ ，表明链已充分收敛。

参考文献

[1]Al-Osh, M. A. and Alzaid, A. A. First-order integer-valued autoregressive (INAR(1)) process[J]. Journal of Time Series Analysis, 8(3):261 – 275, 1987.  
[2]Davis R. A., Wu R. A negative binomial model for time series of counts[J]. Biometrika, 2009, 96(3): 735–749.  
[3]Heinen A. Modelling time series count data: an autoregressive conditional Poisson model[J]. Available at SSRN 1117187, 2003.  
[4]Alzaid A. A. and Al-Osh, M. A. An integer-valued pth-order autoregressive structure (INAR (p)) process[J]. Journal of Applied Probability, 1990, 27(2): 314–324.

表4：贝叶斯后验估计及95% 可信区间

Parameter	Sexual Assault	Theft	Narcotics
W (Post. Mean)	1.485	14.910	1.042
95% CI	[0.523, 2.429]	[12.071, 17.806]	[0.773, 1.362]
$\alpha$ (Post. Mean)	0.047	0.315	0.039
95% CI	[0.002, 0.127]	[0.246, 0.379]	[0.003, 0.092]
$\beta$ (Post. Mean)	0.376	0.122	0.767
95% CI	[0.034, 0.745]	[0.023, 0.221]	[0.695, 0.826]
$\gamma$ (Post. Mean)	-0.173	-2.857	-1.357
95% CI	[-0.378, 0.027]	[-3.757, -1.940]	[-1.543, -1.160]
Acceptance Rate	24.6%	8.4%	51.7%
Min ESS	621	155	146

基于贝叶斯泊松 INGARCH 模型的分析显示，在95% 可信区间内，性侵犯包含零，而盗窃与毒品犯罪不包括零，因此芝加哥温度升高对性侵犯无显著影响，而温度升高使盗窃与毒品犯罪率显著降低。

四、结论

本研究基于贝叶斯泊松 INGARCH 模型框架，通过建立包含温度外生协变量的时变计数数据模型，系统地探讨温度对不同犯罪类型的因果影响。研究采用具有统计推断优势的贝叶斯估计策略，构建了基于条件均值的自回归条件异方差模型，通过马尔科夫链蒙特卡罗算法实现参数后验分布的精确估计，有效克服传统方法在时序离散数据分析中的置信区间覆盖率不足问题。

实证结果表明，在控制时间趋势和季节效应后，温度外生项与盗窃、毒品犯罪呈现显著负相关，而性侵犯的温度系数统计不显著。这一异质性效应揭示了不同类型犯罪发生机制的复杂环境敏感性，建议进一步通过分层模型探讨社会经济调节效应。

相较于传统模型的因果推断方法，本研究通过参数化的条件均值过程，增强了对犯罪影响因素的时间动态捕捉能力。尽管本文的方法取得了合理的良好结果，但需进一步的研究来完善此方法，以适用于更一般情况的计数时间序列。可以将此方法扩展到其他模型，如零膨胀广义泊松和 COM - Poisson INGARCH 模型，混合 INGARCH 模型以及计数时间序列的双变量模型。此外，除了检验犯罪与温度数据之间的关系外，其他的数据分析示例也会很有趣。因此，贝叶斯估计似乎是一种在时间序列计数数据背景下进行因果检验的合适且信息丰富的方法。

- [5]Zeger S. L. and Qaqish B. Markov regression models for time series: a quasi-likelihood approach[J]. Biometrics, 1988: 1019–1031.
- [6]Ferland R., Latour A., Oraichi D. Integer - valued GARCH process[J]. Journal of time series analysis, 2006, 27(6): 923–942.
- [7]Fokianos K., Rahbek A., Tjøstheim D. Poisson autoregression[J]. Journal of the American Statistical Association, 2009, 104(488): 1430–1439.
- [8]Christou V. and Fokianos K. Quasi - likelihood inference for negative binomial time series models[J]. Journal of Time Series Analysis, 2014, 35(1): 55–78.
- [9]Shen B., Hu X., Wu H. Impacts of climate variations on crime rates in Beijing, China[J]. Science of the total environment, 2020, 725: 138190.
- [10]Cruz E., D ' Alessio S. J., Stolzenberg L. The effect of maximum daily temperature on outdoor violence[J]. Crime & Delinquency, 2023, 69(6–7): 1161–1182.
- [11] 袁玉芳. 贝叶斯估计方法在异方差检验中的应用 [D]. 江苏师范大学, 2019.DOI:10.27814/d.cnki.gxzs.2019.000308.
- [12] 雷庆祝, 秦永松, 罗敏. 强混合样本下刻度指数分布族参数的经验贝叶斯估计和检验 [J]. 广西师范大学学报 (自然科学版), 2017, 35(03): 63–74. DOI:10.16088/j.issn.1001-6600.2017.03.008.
- [13] 胡战虎. 基于贝叶斯估计的多分辨图像滤波方法 [J]. 电子学报, 2002, (01): 66–68. DOI:CNKI:SUN:DZXU.0.2002-01-016.
- [14] 王江荣, 袁维红, 赵睿, 等. 基于贝叶斯估计的路基沉降时间序列分析模型 [J]. 矿山测量, 2016, 44(04): 73–77. DOI:CNKI:SUN:KSCL.0.2016-04-020.
- [15]Chu Y. New Bayesian regression models for massive data and extreme longitudinal data[D]. Brunel University London, 2024.