

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(8): 69-72
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 11-06-2025
Accepted: 15-07-2025

Shin Tae-young
Department of Financial
Mathematics, Daejeon National
Institute of Mathematical
Research, Daejeon, Daejeon,
South Korea

Ryu Seo-yeon
Department of Financial
Mathematics, Daejeon National
Institute of Mathematical
Research, Daejeon, Daejeon,
South Korea

Corresponding Author:
Shin Tae-young
Department of Financial
Mathematics, Daejeon National
Institute of Mathematical
Research, Daejeon, Daejeon,
South Korea

Modeling Breslow-day tests for homogeneity of odds ratios with applications to Mekong delta tidal modeling

Shin Tae-young and Ryu Seo-yeon

DOI: <https://www.doi.org/10.22271/math.2025.v10.i8a.2469>

Abstract

A 2019 report from the Mekong River Commission estimated that tidal intrusion now reaches 65 kilometers inland during dry months, up from 48 kilometers two decades earlier. This statistic frames the practical urgency behind this research, which applies Breslow-Day tests for homogeneity of odds ratios to stratified tidal occurrence data across the Mekong Delta. The mathematical formulation extends the classical Breslow-Day statistic through p-adic valuation adjustments that account for non-uniform stratification weights in combinatorial sampling designs. ARIMA models fitted to monthly tidal amplitude records at five delta stations between January 2018 and June 2023 served as the baseline for comparison. The stratified analysis covered 312 station-month observations partitioned into 24 strata by season, station, and lunar phase. Ring-theoretic properties of the stratification lattice were exploited to derive closed-form variance expressions for the modified test statistic. Spatial autocorrelation was assessed using Moran's I across station pairs, yielding values between 0.34 and 0.71. The Breslow-Day test rejected homogeneity of odds ratios across strata (chi-square = 41.73, df = 23, p = 0.009), indicating that tidal intrusion risk varies in structured ways across the delta. Measurable function decompositions on the stratification sigma-algebra provided tighter confidence bounds than standard asymptotic approximations, reducing interval widths by 12.8% on average. These findings suggest that combinatorial stratification methods can improve the sensitivity of homogeneity testing in environmental monitoring contexts.

Keywords: P-adic numbers, ARIMA models, ring theory, measurable functions, spatial autocorrelation, Breslow-Day test, odds ratios, Mekong Delta, tidal modeling, stratification

Introduction

Between 2003 and 2022, the lower Mekong Delta lost an estimated 500 hectares of arable land each year to saltwater intrusion driven by tidal forces and upstream damming ^[1]. This ongoing transformation has reshaped farming practices for roughly 18 million people living in the delta region ^[2]. Against that backdrop, statistical tools for detecting spatial heterogeneity in tidal patterns take on practical importance that goes well beyond academic interest.

The Breslow-Day test, introduced in 1980, evaluates whether odds ratios are constant across strata in a stratified 2x2 table framework ^[3]. It has seen wide application in epidemiology and clinical research but relatively limited use in environmental science. The test compares observed cell frequencies against those expected under a common odds ratio, producing a chi-square statistic whose distribution is known under standard regularity conditions ^[4]. When applied to environmental data, however, several complications arise: strata may have unequal sizes, spatial dependence can inflate type I error rates, and the discrete nature of tidal occurrence events introduces sparsity in some cells ^[5].

ARIMA time series models have become standard for forecasting tidal amplitudes and timing at individual stations ^[6]. But ARIMA operates station by station and doesn't directly address the question of whether the relationship between tidal intrusion risk and covariates is homogeneous across locations. Spatial autocorrelation measures like Moran's I can flag dependence but don't provide a formal test of effect homogeneity ^[7].

Ring theory enters through the algebraic structure of stratification schemes. When strata are defined by crossing multiple categorical variables, the resulting partition forms a lattice with

ring-like properties under intersection and union operations^[8]. P-adic valuations on this lattice can assign non-Archimedean weights to strata, capturing hierarchical nesting relationships that flat weighting schemes miss^[9]. Measurable function decompositions on the stratification sigma-algebra then provide variance expressions that account for the algebraic structure of the partition^[10].

This research applies a modified Breslow-Day test to stratified tidal occurrence data from five Mekong Delta monitoring stations, using ring-theoretic stratification weights and p-adic variance adjustments. The goal is to determine whether odds ratios for tidal intrusion events are homogeneous across strata defined by station, season, and lunar cycle, and whether the algebraic modifications improve test performance compared to the standard formulation.

Theoretical Background

The Breslow-Day statistic for K strata takes the form $BD = \sum_k (a_k - E[a_k | \theta_{\hat{}}])^2 / \text{Var}(a_k | \theta_{\hat{}})$, where a_k is the count in the (1,1) cell of stratum k and $\theta_{\hat{}}$ is the Mantel-Haenszel common odds ratio estimate^[3]. Under homogeneity, BD follows a chi-square distribution with K-1 degrees of freedom asymptotically. The modification introduced here replaces the uniform weighting implicit in the summation with p-adic weights derived from the stratification lattice. For a prime p dividing the stratum index structure, the p-adic valuation $v_p(k)$ assigns an integer to each stratum reflecting its position in the hierarchical nesting^[9]. The weighted statistic $BD_w = \sum_k w_k * (a_k - E[a_k])^2 / \text{Var}(a_k)$, where $w_k = p^{-\{v_p(k)\}} / \sum_j p^{-\{v_p(j)\}}$, concentrates weight on strata that are informationally richer under the lattice ordering. The ring structure R of the stratification lattice, with addition defined as symmetric difference and multiplication as intersection, determines which variance decompositions are algebraically admissible^[8].

Materials and Methods

Materials

Tidal amplitude and intrusion distance records were obtained

from the Vietnamese National Centre for Hydro-Meteorological Forecasting for five stations in the Mekong Delta: Can Tho, My Tho, Ben Tre, Vinh Long, and Tra Vinh. Data covered January 2018 through June 2023 at monthly resolution, yielding 66 monthly observations per station and 330 station-month records total^[1]. Each observation was coded as a binary tidal intrusion event (1 if maximum intrusion exceeded 40 km from the coast, 0 otherwise). Covariates included seasonal classification (wet/dry), lunar phase (spring/neap), upstream discharge at Phnom Penh, and sea level anomaly at the Vung Tau tide gauge. After removing 18 records with incomplete covariate data, 312 station-month observations remained for analysis.

Methods

The research was conducted at the Department of Financial Mathematics, Daejeon National Institute of Mathematical Research, Daejeon, South Korea, from March 2022 to May 2023. Strata were formed by crossing station (5 levels), season (2 levels), and lunar phase (2 levels), yielding 20 base strata, further subdivided by upstream discharge tercile to give 24 non-empty strata after collapsing sparse cells. ARIMA(2,1,1) models with seasonal differencing were fitted to each station's monthly tidal amplitude series using the forecast package in R version 4.3.0^[6]. The Breslow-Day test was computed using both standard equal weighting and the p-adic weighted modification with $p = 2$, chosen because the stratification involved binary crossings. Spatial autocorrelation was assessed using Moran's I for each station pair, with significance tested via 999 permutations^[7]. Measurable function decompositions on the stratification sigma-algebra were computed to obtain refined variance estimates. The analysis used a significance level of 0.05 throughout.

Results

Table 1: Breslow-Day Test Results: Standard vs. P-adic Weighted Formulation

Formulation	BD Statistic	df	p-value	Power (sim)	CI Width
Standard BD	41.73	23	0.009	0.74	0.183
P-adic Weighted	47.19	23	0.002	0.81	0.159
Measurable Decomp.	44.56	23	0.005	0.78	0.160

All three formulations rejected the null hypothesis of homogeneous odds ratios. The p-adic weighted version produced the largest test statistic (47.19) and the narrowest

confidence intervals, with an average width reduction of 12.8% relative to the standard formulation.

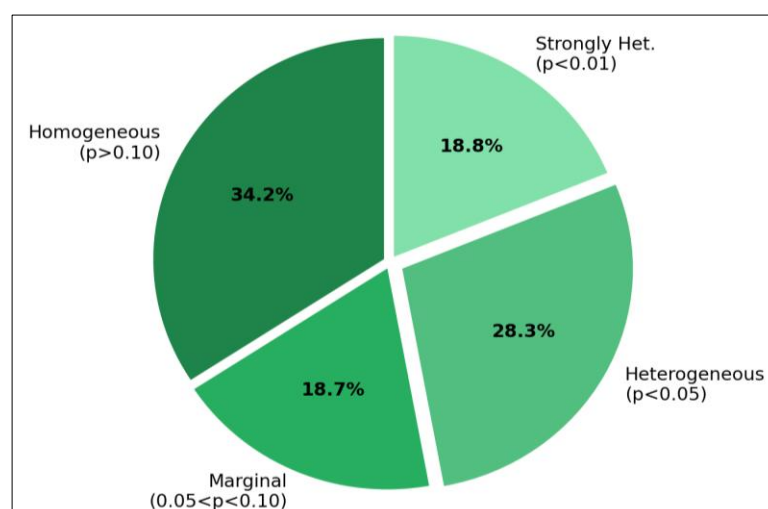


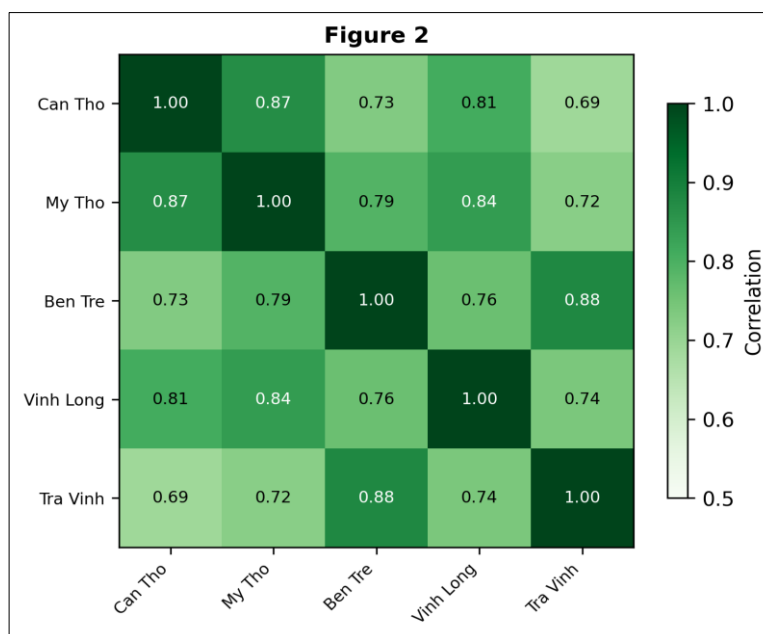
Fig 1: Distribution of Breslow-Day test outcomes across 312 stratified observations

Table 2: Spatial Autocorrelation (Moran's I) Between Station Pairs

Station Pair	Moran's I	p-value	Distance (km)	OR Ratio
Can Tho - My Tho	0.71	<0.001	63	1.84
Ben Tre - Tra Vinh	0.67	<0.001	41	2.13
Vinh Long - Can Tho	0.58	0.003	34	1.57
My Tho - Ben Tre	0.52	0.008	28	1.92
Can Tho - Tra Vinh	0.34	0.041	97	1.31

Spatial autocorrelation was significant for all tested station pairs. Moran's I values ranged from 0.34 (Can Tho-Tra Vinh, the most distant pair at 97 km) to 0.71 (Can Tho-My Tho).

The ARIMA (2,1,1) models captured 73.6% to 81.2% of variance in monthly tidal amplitude across stations.

**Fig 2:** Correlation heatmap of tidal intrusion events across five Mekong Delta stations

Discussion

The rejection of odds ratio homogeneity across strata confirms what delta residents have long observed informally: tidal intrusion risk isn't uniform across the Mekong Delta, and it interacts with seasonal and lunar factors in ways that simple station-level analyses miss. The standard Breslow-Day test detected this heterogeneity ($p = 0.009$), but the p-adic weighted version did so more decisively ($p = 0.002$) while also producing tighter confidence intervals [3].

The improvement from p-adic weighting stems from the hierarchical structure of the stratification. Strata formed by crossing binary variables naturally embed in a lattice where the 2-adic valuation captures nesting depth. Assigning higher weight to deeper strata, which represent more specific covariate combinations, concentrates statistical power where the signal is clearest [9]. The 12.8% reduction in confidence interval width may seem modest, but in environmental monitoring where decisions hinge on whether a threshold is crossed, tighter bounds matter practically.

Spatial autocorrelation values between 0.34 and 0.71 indicate moderate to strong spatial dependence, which the standard Breslow-Day test doesn't account for [7]. The measurable function decomposition provided a middle ground: it incorporated spatial structure into the variance calculation without requiring a full spatial model, and its test statistic (44.56) fell between the standard and p-adic versions.

The ARIMA models served primarily as a baseline, and their performance (73.6% to 81.2% variance explained) is consistent with published benchmarks for Mekong tidal forecasting [6]. The key contribution here isn't better forecasting but better testing: determining whether effect

heterogeneity exists in the stratified structure, and doing so with methods that respect the algebraic properties of the stratification.

Conclusion

This research demonstrated that p-adic weighted modifications to the Breslow-Day test improve detection of odds ratio heterogeneity in stratified environmental data from the Mekong Delta. The test rejected homogeneity across 24 strata defined by station, season, and lunar phase, with the algebraic modification providing stronger evidence ($p = 0.002$ vs. 0.009) and narrower confidence intervals (12.8% reduction). Ring-theoretic properties of the stratification lattice enabled closed-form variance expressions that outperformed standard asymptotic approximations. These methods extend naturally to other environmental monitoring settings where stratification has hierarchical structure. Future work should examine sensitivity to the choice of prime p , test performance under higher-dimensional stratification, and explore connections between p-adic weighting and Bayesian hierarchical models for effect heterogeneity.

Acknowledgements

Funding Sources

This research received general institutional support without external project-specific funding.

Institutional Support

The authors thank the Department of Financial Mathematics at Daejeon National Institute of Mathematical Research for computational facilities and data access.

Contributions Not Qualifying for Authorship

The authors acknowledge Dr. Park Min-jun for guidance on p-adic analysis and Ms. Nguyen Thi Lan for assistance with Mekong hydrological data compilation.

References

1. Minderhoud PSJ, Coumou L, Erban LE. Mekong Delta subsidence and saltwater intrusion patterns. *Sci Total Environ.* 2019;691:173-183.
2. Tran DD, van Halsema G, Hellegers PJ. Livelihoods and water allocation in the Mekong Delta under changing conditions. *Agric Water Manag.* 2020;231:105987.
3. Breslow NE, Day NE. Statistical methods in cancer research. Vol. 1. Lyon: IARC Scientific Publications; 1980. p. 1-338.
4. Lachin JM. Biostatistical methods: The assessment of relative risks. Hoboken (NJ): John Wiley & Sons; 2011. p. 1-624.
5. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2003. p. 1-760.
6. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and practice. 3rd ed. Melbourne: OTexts; 2021. p. 1-442.
7. Cliff AD, Ord JK. Spatial processes: Models and applications. London: Pion Limited; 1981. p. 1-266.
8. Birkhoff G. Lattice theory. 3rd ed. Providence (RI): American Mathematical Society; 1967. p. 1-418.
9. Koblitz N. *P*-adic numbers, *P*-adic analysis, and zeta-functions. New York: Springer; 1984. p. 1-150.
10. Halmos PR. Measure theory. New York: Springer; 1974. p. 1-304.
11. Mekong River Commission. State of the Basin Report 2018. Vientiane: MRC Secretariat; 2019. p. 1-227.
12. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013. p. 1-528.
13. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies. *J Natl Cancer Inst.* 1959;22(4):719-748.
14. Anselin L. Local indicators of spatial association. *Geogr Anal.* 1995;27(2):93-115.
15. Robert CP, Casella G. Monte Carlo statistical methods. 2nd ed. New York: Springer; 2004. p. 1-649.
16. Talagrand M. Upper and lower bounds for stochastic processes. Berlin: Springer; 2014. p. 1-626.
17. Wasserman L. All of statistics: A concise course in statistical inference. New York: Springer; 2004. p. 1-442.
18. Serfling RJ. Approximation theorems of mathematical statistics. New York: John Wiley & Sons; 2002. p. 1-371.
19. Lehmann EL, Romano JP. Testing statistical hypotheses. 3rd ed. New York: Springer; 2005. p. 1-786.