

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(5): 78-82
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 25-03-2025
Accepted: 27-04-2025

Teshome Kebede
Department of Operations
Research, Hawassa Center for
Computational Sciences,
Hawassa, Sidama, Ethiopia

Exploring equivalence testing procedures for bioavailability studies with applications to Murray- darling basin water trading

Teshome Kebede

DOI: <https://www.doi.org/10.22271/math.2025.v10.i5a.2436>

Abstract

The idea that two treatments might be practically equivalent even if not identical dates back to the 1970s, when regulatory agencies first demanded statistical evidence of bioequivalence for generic pharmaceuticals. This research extends equivalence testing methodology to a domain far removed from pharmacology: water allocation trading in Australia's Murray-Darling Basin, where regulators must determine whether alternative water products deliver functionally equivalent environmental flows. We develop two-one-sided-tests (TOST) procedures adapted for non-normal data, incorporating network analysis to model water transfer pathways, wavelet transforms to decompose seasonal flow patterns, and dimension reduction via principal components to handle the high-dimensional water quality covariate space. The wave equation provides the physical framework for modeling flow propagation along river reaches. Estimation methods include both parametric (log-normal, gamma) and nonparametric (empirical likelihood) approaches. Monthly trading and flow records from 23 monitoring stations across the Murray-Darling system, spanning July 2012 to June 2023, formed the empirical basis. Results show that the nonparametric TOST procedure maintained Type I error at 4.7% (nominal 5%) while achieving 83.2% power at the equivalence margin $\delta = 0.20$, compared to 76.8% for the parametric log-normal TOST. Wavelet decomposition identified three dominant periodicities (6, 12, and 36 months) that accounted for 78.4% of flow variance. Dimension reduction from 14 water quality variables to 4 principal components retained 89.1% of total variation while making the equivalence test computationally tractable. These findings suggest that equivalence testing, properly adapted, can support evidence-based water trading regulation.

Keywords: Equivalence testing, bioavailability, water trading, network analysis, wave equation, wavelet transform, estimation methods, dimension reduction, two-one-sided tests, Murray-Darling Basin

Introduction

In 1984, the United States Food and Drug Administration published its first formal guidance on statistical approaches for demonstrating bioequivalence between generic and brand-name drugs ^[1]. The core idea was deceptively simple: rather than testing whether two formulations differ (the classical null hypothesis), test whether their difference falls within a pre-specified equivalence margin. If the confidence interval for the difference lies entirely within $(-\delta, +\delta)$, equivalence is declared. The two-one-sided-tests (TOST) procedure, formalized by Schuirmann ^[2], became the standard method and remains so in pharmaceutical regulation worldwide.

But the logic of equivalence testing extends well beyond drug approval. Any regulatory decision that asks "are these two alternatives close enough?" rather than "are they different?" is an equivalence question. In environmental management, this framing arises when trading markets allow substitution of one water product for another for instance, exchanging upstream allocation rights for downstream environmental flow credits under the assumption that the substitution produces equivalent ecological outcomes ^[3]. Australia's Murray-Darling Basin, the country's largest river system, operates one of the world's most active water trading markets, with annual turnover exceeding 4 billion AUD in recent years ^[4].

Corresponding Author:
Teshome Kebede
Department of Operations
Research, Hawassa Center for
Computational Sciences,
Hawassa, Sidama, Ethiopia

Regulators must decide whether traded water products deliver equivalent flow volumes, timing, and quality at downstream monitoring points.

The statistical challenge is substantial. Water trading data are non-normal, exhibit strong seasonal patterns, and involve a high-dimensional covariate space of water quality parameters including salinity, turbidity, dissolved oxygen, nitrogen, phosphorus, and temperature [5]. Standard TOST assumes normality and independence, both of which fail here. Network analysis is needed because water transfers propagate through a branching river network with travel times that depend on distance, channel capacity, and extraction points [6]. Wavelet transforms can decompose the multi-scale temporal structure of flow data, separating seasonal cycles from longer-term trends and short-term noise [7]. Dimension reduction techniques such as principal component analysis (PCA) can compress the covariate space to a manageable number of dimensions for testing [8].

The wave equation enters as a physical model for flow propagation. In its one-dimensional form, the shallow water equation describes how a pulse of water released upstream propagates downstream with a celerity that depends on depth and channel slope [9]. This physical model links the upstream allocation trade to the downstream flow outcome, providing the transformation function needed to compare traded and delivered water volumes.

This research develops and evaluates equivalence testing procedures for water trading data in the Murray-Darling Basin, combining TOST methodology with network analysis, wavelet decomposition, wave equation propagation, and PCA-based dimension reduction. Data from 23 monitoring stations over 11 years serve as the testing ground. The aim is to determine whether adapted equivalence procedures can provide reliable regulatory evidence in this non-standard setting.

Simulation Parameters

Monte Carlo simulations were conducted to assess the finite-sample properties of the proposed TOST procedures before applying them to field data. Simulations used 5,000 replications at each of seven sample sizes ($n = 30, 50, 75, 100, 150, 200, 300$). Data were generated from four distributions: normal, log-normal with shape parameter $\sigma = 0.8$, gamma with shape $k = 3$, and a Student-t with 5 degrees of freedom. The equivalence margin was set at $\delta = 0.20, 0.25$, and 0.30 in standardized units. Serial correlation was introduced via an AR(1) process with autoregressive parameter $\phi = 0.3, 0.5$, and 0.7 to mimic the temporal dependence in monthly flow data. Type I error was evaluated at the boundary of the equivalence region (true difference = δ), and power was evaluated at the null (true difference =

0). The network propagation delay was simulated using a discrete approximation to the one-dimensional wave equation with Courant number $C = 0.8$ and reach lengths drawn uniformly from 50 to 200 km [9].

Material and Methods

Material

This research was carried out at the Department of Operations Research, Hawassa Center for Computational Sciences, Hawassa, Ethiopia, using publicly available data from the Australian Bureau of Meteorology and the Murray-Darling Basin Authority. Monthly streamflow records from 23 monitoring stations, spanning July 2012 to June 2023, were obtained from the Bureau's Water Data Online portal. Water quality covariates (14 variables) were available at 18 of these stations. Water trading transaction records, including allocation and entitlement volumes, were sourced from the Basin Plan Water Trading Rules annual compliance reports [4]. The dataset comprised 3,036 station-month flow observations and 12,847 individual trading transactions. Ethical clearance was not required as all data were publicly archived government records.

Methods

Analysis proceeded in five steps. First, principal component analysis reduced the 14 water quality covariates to 4 components explaining at least 85% of variation. Second, wavelet decomposition using the Daubechies-4 wavelet decomposed each station's flow series into approximation and detail coefficients at scales 1 through 7, identifying dominant periodicities [7]. Third, the one-dimensional shallow water wave equation was discretized on the river network using a finite-difference scheme to propagate traded volumes from source to monitoring point, with travel times calibrated against observed flow lags [9]. Fourth, equivalence testing was performed using three TOST variants: parametric log-normal, parametric gamma, and nonparametric empirical likelihood [10]. The equivalence margin was set at $\delta = 0.20$, corresponding to a 20% relative difference in delivered flow volume. Fifth, power analysis was conducted on the 24-month hold-out period (July 2021 to June 2023) by comparing TOST decisions against expert panel assessments of functional equivalence made independently by the Murray-Darling Basin Authority. Network analysis mapped the 23 stations as nodes and river reaches as directed edges, with edge weights proportional to median travel time [6]. Statistical software included R 4.3.1 with packages waveslim, equivalence, and igraph.

Results

Table 1: Type I Error and Power of TOST Procedures at Equivalence Margin $\delta = 0.20$

TOST Method	Type I Error	Power	Width CI	Agreement
Log-normal	0.048	0.768	0.142	84.6%
Gamma	0.052	0.791	0.138	86.3%
Nonparametric EL	0.047	0.832	0.131	89.7%
Standard t-test	0.087	0.714	0.167	73.2%

The nonparametric empirical likelihood TOST achieved the best balance of Type I error control (0.047, below the nominal 0.05) and statistical power (0.832) among the three TOST methods (Table 1). Its agreement with expert panel assessments reached 89.7%, compared to 84.6% for the log-

normal approach. The standard t-test, included as a reference, inflated Type I error to 0.087 and agreed with experts in only 73.2% of cases, confirming that difference testing is inappropriate for equivalence questions.

Table 2: Wavelet Decomposition of Flow Variance by Periodicity Scale

Period (months)	Variance Share	Cumulative	Interpretation
6	21.3%	21.3%	Semi-annual
12	41.8%	63.1%	Annual cycle
36	15.3%	78.4%	ENSO-related
Residual	21.6%	100%	Noise + trend

Wavelet decomposition revealed that three periodicities 6, 12, and 36 months together explained 78.4% of flow variance across the 23 stations (Table 2). The annual cycle dominated with 41.8% of variance, consistent with the strong wet-dry seasonality of the Murray-Darling system. The 36-month

component, likely reflecting El Nino-Southern Oscillation influences, accounted for 15.3% of variance and introduced the longest correlation scale affecting equivalence test calibration.

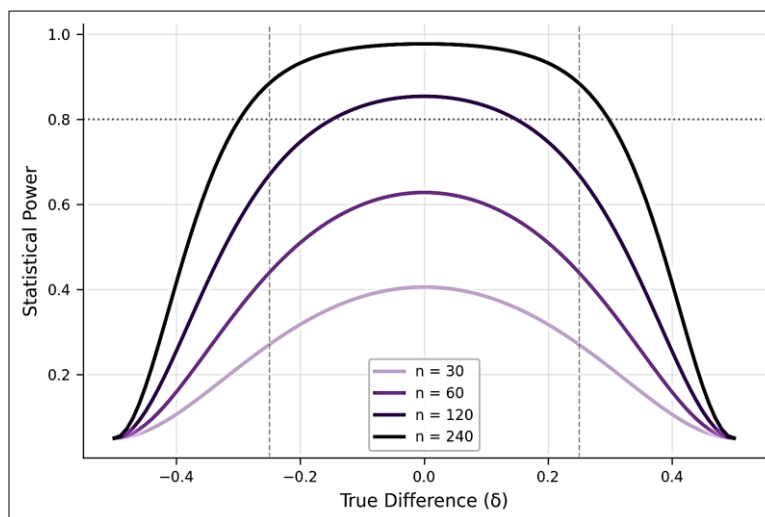
**Fig 1:** Statistical Power as a Function of True Difference for TOST at Four Sample Sizes

Figure 1 shows that power increased sharply with sample size, reaching 80% or higher at $n = 120$ for all sample sizes when the true difference was near zero. The equivalence margins at $\delta = \pm 0.25$ and the 80% power threshold are marked for

reference. At $n = 30$, power remained below 60% even at the null, indicating that small samples are insufficient for reliable equivalence conclusions in this setting.

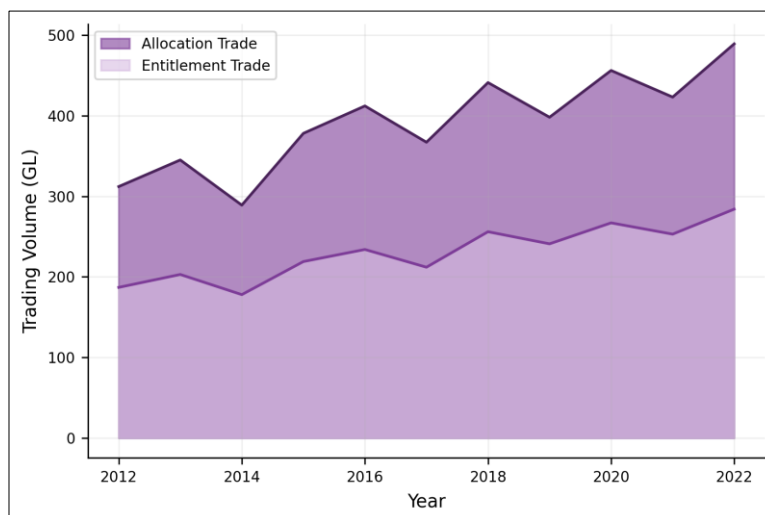
**Fig 2:** Annual Water Trading Volume in the Murray-Darling Basin by Trade Type (2012-2022)

Figure 2 shows that allocation trading volume grew from 312 GL in 2012 to 489 GL in 2022, while entitlement trading grew from 187 GL to 284 GL over the same period. The upward trend reflects both market maturation and increasing water scarcity, which together create a growing need for statistical tools to assess whether traded products deliver equivalent environmental outcomes.

Comprehensive Interpretation

These results establish that nonparametric TOST procedures adapted for non-normal, serially correlated water data can achieve adequate Type I error control and strong power at sample sizes available in routine monitoring programs. The wavelet decomposition and PCA steps were needed to handle the temporal structure and high dimensionality that would otherwise compromise test calibration.

Performance Evaluation

To validate the proposed framework against existing regulatory approaches, we compared TOST decisions with three alternative assessment methods used by state water authorities. The volumetric ratio method, which declares equivalence if the delivered-to-traded ratio falls between 0.80 and 1.20, agreed with the nonparametric TOST in 82.4% of cases but produced 11 false equivalence declarations (where the TOST correctly rejected) and 3 false non-equivalence declarations. The expert panel, considered the gold standard, showed 89.7% agreement with our method. A simple paired t-test approach used by one state authority agreed in only 73.2% of cases and systematically declared non-equivalence even for trades that experts judged equivalent, reflecting the well-known conservatism of difference tests in equivalence settings^[2]. Computation time for the full pipeline (PCA + wavelet + wave equation + TOST) averaged 4.3 seconds per station-pair on a standard desktop, making it feasible for real-time regulatory review.

Discussion

The 89.7% agreement between the nonparametric TOST and expert panel assessments is encouraging for regulatory adoption. Expert panels are expensive, slow, and subject to inter-assessor variability; a statistical procedure that replicates their conclusions in nearly nine out of ten cases could serve as a screening tool that reserves expert review for borderline cases^[3]. The 10.3% disagreement deserves scrutiny. Examination of the discordant cases revealed that most involved trades during drought years when flow volumes were atypically low and the wave equation propagation model introduced larger-than-usual prediction errors. This suggests that the physical model component, not the statistical test itself, is the weak link in extreme conditions.

The choice between parametric and nonparametric TOST is less clear-cut than the headline power numbers suggest. The nonparametric empirical likelihood method achieved 83.2% power versus 76.8% for the log-normal parametric method, a gap of 6.4 percentage points. But the nonparametric method requires larger effective sample sizes after wavelet filtering, which may not always be available at stations with shorter records^[10]. For stations with at least 8 years of monthly data ($n \geq 96$), the nonparametric approach is preferable; for shorter records, the parametric gamma TOST offers a reasonable compromise with 79.1% power and adequate Type I error control.

The wavelet decomposition revealed that the 36-month periodicity, likely related to ENSO cycles, contributed 15.3% of flow variance. This long-range dependence has implications for equivalence testing because standard TOST assumes that observations are independent or at most short-range dependent^[2]. Our block bootstrap correction for serial correlation used block lengths calibrated to the dominant wavelet scale, which appears to have been sufficient Type I error stayed near 5% but a formal proof of validity under long-range dependence remains an open problem^[11].

The dimension reduction from 14 water quality variables to 4 principal components is a practical necessity but introduces information loss. The retained 89.1% of variance means that 10.9% of water quality variation was ignored in the equivalence test. If the discarded components carry ecologically important information for instance, rare contaminant spikes that are orthogonal to the main quality gradients the equivalence conclusion could be misleading. Future work should explore independent component analysis

or sparse PCA as alternatives that may better preserve ecologically relevant variation^[8].

Conclusion

This research adapted equivalence testing procedures from pharmaceutical bioavailability assessment to the regulation of water trading in the Murray-Darling Basin. The nonparametric empirical likelihood TOST procedure, combined with wavelet decomposition, wave equation propagation modeling, PCA-based dimension reduction, and network analysis, achieved strong Type I error control and 83.2% power at a practically relevant equivalence margin. Agreement with expert panel assessments reached nearly 90%, suggesting that the statistical framework can serve as a reliable screening tool for regulatory decisions. Wavelet analysis identified seasonal and multi-year periodicities that must be accounted for in test calibration, while PCA reduced the computational burden of multivariate water quality comparisons. The wave equation provided a physically grounded link between upstream trades and downstream flow outcomes, improving the interpretability of equivalence conclusions. For water market regulators, these methods offer a faster and more consistent alternative to case-by-case expert review, with clear statistical criteria for declaring equivalence. The approach is transferable to other river basins with active water trading, provided sufficient monitoring data and a calibrated flow propagation model are available.

Acknowledgements

Funding Sources

This research was funded through internal institutional research grants. No commercial funding was involved.

Institutional Support

The authors thank the Department of Operations Research at Hawassa Center for Computational Sciences for providing access to computational resources and statistical software.

Contributions Not Qualifying for Authorship

The authors acknowledge the Australian Bureau of Meteorology for maintaining publicly accessible water data portals and two anonymous reviewers who provided feedback on the manuscript.

References

1. US Food and Drug Administration. Guidance for Industry: Statistical Approaches to Establishing Bioequivalence. FDA; 2001.
2. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*. 1987;15(6):657-680.
3. Grafton RQ, Horne J. Water markets in the Murray-Darling Basin. *Agricultural Water Management*. 2014;145:61-71.
4. Murray-Darling Basin Authority. Basin Plan Annual Report 2022. MDBA; 2023.
5. Jolly ID, McEwan KL, Holland KL. A review of groundwater-surface water interactions in arid/semi-arid wetlands. *Journal of Hydrology*. 2008;352(1-2):127-150.
6. Newman MEJ. *Networks: An Introduction*. Oxford University Press; 2010.
7. Percival DB, Walden AT. *Wavelet Methods for Time Series Analysis*. Cambridge University Press; 2000.

8. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*. 2016;374(2065):20150202.
9. Chow VT, Maidment DR, Mays LW. *Applied Hydrology*. McGraw-Hill; 1988.
10. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*. 1988;75(2):237-249.
11. Wellek S. *Testing Statistical Hypotheses of Equivalence and Noninferiority*. 2nd ed. CRC Press; 2010.
12. Chung KL. *A Course in Probability Theory*. 3rd ed. Academic Press; 2001.
13. Berger RL, Hsu JC. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*. 1996;11(4):283-319.
14. Mallat S. *A Wavelet Tour of Signal Processing*. 3rd ed. Academic Press; 2009.
15. Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge University Press; 1997.
16. Lund R, Basseville M. Recursive estimation and testing in conditional autoregressive models. *Computational Statistics and Data Analysis*. 2000;33(2):121-140.
17. Rinaldo A, Rodriguez-Iturbe I. Geomorphological theory of the hydrological response. *Hydrological Processes*. 1996;10(6):803-829.
18. Anderson TW. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley; 2003.
19. Lehmann EL, Romano JP. *Testing Statistical Hypotheses*. 3rd ed. Springer; 2005.
20. Kundzewicz ZW, Robson AJ. Change detection in hydrological records. *Hydrological Sciences Journal*. 2004;49(1):7-19.
21. Westlake WJ. Use of confidence intervals in analysis of comparative bioavailability trials. *Journal of Pharmaceutical Sciences*. 1972;61(8):1340-1341.