Statistica Sinica Preprint No: SS-2026-0333	
Title	Diagnostic Checking for Wasserstein Autoregression
Manuscript ID	SS-2026-0333
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202026.0333
Complete List of Authors	Chenxiao Dai, Feiyu Jiang and Dong Li
Corresponding Authors	Feiyu Jiang
E-mails	jiangfy@fudan.edu.cn
Notice: Accepted author version.	

Statistica Sinica: Newly accepted Paper

Diagnostic Checking for Wasserstein Autoregression

Chenxiao Dai[1], Feiyu Jiang[2], and Dong Li[1]

[1] *Tsinghua University,* [2] *Fudan University*

Abstract: Wasserstein autoregression provides a robust framework for modeling serial dependence among probability distributions, with wide-ranging applications in economics, finance, and climate science. In this paper, we develop portmanteau-type diagnostic tests for assessing the adequacy of the first-order autoregressive transport map (ATM(1)) model. By defining autocorrelation functions for model errors and residuals in the Wasserstein space, we construct two related tests: one analogous to the classical McLeod type test, and the other based on the sample-splitting approach of Davis and Fernandes (2025). We establish that, under mild regularity conditions and an additional cancellation condition for the splitting test statistic, the corresponding test statistics converge in distribution to chi-square limits. Simulation studies and empirical applications demonstrate that the proposed tests effectively detect model mis-specification, offering a principled and reliable diagnostic tool for distributional time series analysis.

Key words and phrases: Distributional Data, Model Checking, Optimal Transport, Random Object, Wasserstein Distance

1. Introduction

Statistical modeling and analysis of object-valued time series has attracted a lot of attention lately due to the increasing encounter of such data in many modern applications. Examples include yearly age-at-death distributions (Mazzuco and Scarpa, 2015; Shang and Hyndman, 2017; Dubey and Müller, 2020), annual observations of compositions of energy sources for

power generation in the U.S. (Zhu and Müller, 2024), and daily Pearson correlation matrices for five cryptocurrencies (Jiang et al., 2024b), among others. One of the well-known challenges in dealing with such object-valued data is the lack of the usual vector/Hilbert space operations, such as scalar multiplication, addition/subtraction, and inner product, which may not be well defined and only the distance between two non-Euclidean data objects is available. Another challenge stems from the serial dependence, which is the norm rather than the exception for many time-ordered data.

Despite the challenge, there has been a growing literature on statistical inference for such object-valued time series. For example, change-point testing and estimation have been addressed in Dubey and Müller (2020); Jiang et al. (2024b); Zhang et al. (2026); testing for serial independence has been tackled in Jiang et al. (2024a). On the other front, new statistical models have been developed for some specific type of object-valued time series, such as distributional time series. In particular, Zhang et al. (2022) proposed to use the logarithmic map to lift the random densities into a linear tangent space, which is a Hilbert space and the autoregressive model is imposed; also see Chen et al. (2023) for related contribution along this line. Recently, Zhu and Müller (2023) developed an intrinsic approach to autoregressive modeling of one-dimensional distributional time series. The basic idea is to model the optimal transport map using newly defined scalar multiplication and addition in Wasserstein space of distributions. In particular, a random perturbation of the identity map was iteratively contracted/composed and applied to either the increments between consecutive distributions or to the deviations from the marginal Fréchet mean, to produce autoregressive transport map (ATM) models. Subsequently Jiang (2022) generalized this ATM model in Zhu and Müller (2023) to the case of vector-valued distributional time series, i.e., time-evolving vectors with

distributions as coordinates. Another important extension of Zhu and Müller (2023) was proposed by Ghodrati and Panaretos (2024) recently, who expanded the ATM model by allowing the shape of the dynamics being captured by a monotone map, modulated by a contractive parameter regulating the degree of non-degeneracy of the model.

For vector-valued time series, diagnostic checking is an integral part of the model building and has been extensively studied in econometrics and time series analysis; see Li (2003) for a book-length treatment of this topic. Generally speaking, model mis-specification can lead to biased parameter estimates, unreliable inference, and poor predictive performance, making model validation through diagnostic tools an essential step. In this article, we study diagnostic checking for the ATM models developed in Zhu and Müller (2023), a topic that appears lacking in the literature. In this paper, we establish the main asymptotic results for the ATM model of lag one, i.e. the ATM(1) specification, and discuss higher-order ATM(p) diagnostics as possible extensions in Section 4.5. Our contribution is threefold.

(a) We extend the classical definition of the autocorrelation function from vector-valued residuals to optimal transport distortion maps within the ATM framework. This enables the construction of a McLeod-type portmanteau statistic (McLeod, 1978) and the derivation of its asymptotic distribution.

(b) Motivated by the recent work of Davis and Fernandes (2025), who investigated sample splitting as a means to mitigate the estimation effect on the portmanteau statistic in univariate time series models, we develop an analogous sample-splitting adjustment tailored to distributional time series under the ATM model.

(c) Through a comprehensive simulation study, we compare the performance of the classical portmanteau test and its sample-splitting counterpart, highlighting their relative size

and power properties in finite samples.

Finally, we list a few commonly used notations that will appear throughout the paper; any additional notation will be introduced as needed in the main text. Let $\mathbb{R}$ denote the real line. For two functions $f, g : \mathbb{R} \to \mathbb{R}$, we write $f \circ g$ to denote their composition, that is, $[f \circ g](x) = f\big(g(x)\big)$, for $x \in \mathbb{R}$. For any sequence of random variables $\{X_n\}$, we use $\to_d$ and $\to_p$ to denote convergence in distribution and convergence in probability, respectively. We write $X_n = o_p(1)$ if $X_n \to_p 0$ as $n \to \infty$.

2. Background on optimal transport

Let $\Omega = [\omega_s, \omega_e]$ be a closed interval on $\mathbb{R}$ and denote by $\mathcal{W}_2(\Omega)$ the space of Borel probability measures on Ω with finite second moments. For $\mu, \nu \in \mathcal{W}_2(\Omega)$, denote their corresponding distribution functions by F_μ and F_ν, respectively. The 2-Wasserstein distance between μ and ν is defined as

$$d_W^2(\nu, \mu) = \inf_{\gamma \in \Gamma(\nu,\mu)} \int_\Omega |x - y|^2 \mathrm{d}\gamma(x, y),$$

where $\Gamma(\nu, \mu)$ denotes the set of couplings of ν and μ, i.e. Borel probability measures on $\Omega \times \Omega$ with marginal distributions ν and μ.

In the univariate case, if F_μ is absolutely continuous with respect to the Lebesgue measure, then the optimal coupling is induced by a deterministic monotone transport map

$$T(x) = F_\nu^{-1}\big(F_\mu(x)\big),$$

in which $\nu = T\#\mu$ is the pushforward measure of μ by T, and as a consequence,

$$d_W^2(\mu,\nu) = \int_\Omega |T(x)-x|^2 \,\mathrm{d}\mu(x) = \int_0^1 \left|F_\mu^{-1}(u) - F_\nu^{-1}(u)\right|^2 \mathrm{d}u.$$

In this paper, we restrict our attention to the space of monotone transport maps, i.e.

$$\mathcal{T} = \{\, T : \Omega \to \Omega \mid T \text{ is nondecreasing and } T(\omega_s) = \omega_s,\, T(\omega_e) = \omega_e \,\}.$$

Therefore, since both ω_s and ω_e are bounded, all transports in $\mathcal{T}$ (and their inverse transports) are bounded.

The ATM model is built upon a class of iterated random systems of transport maps known as "contractive compositions". A key ingredient is the notion of an α-contraction of a transport map. In particular, for a parameter $\alpha \in [-1,1]$ and any optimal transport map $T \in \mathcal{T}$, the α-contraction is defined by the operator $T \mapsto [\alpha \odot T]$, where

$$[\alpha \odot T](x) = \begin{cases} x + \alpha(T(x)-x), & 0 < \alpha \le 1, \\ x, & \alpha = 0, \\ x + \alpha\left(x - T^{-1}(x)\right), & -1 \le \alpha < 0. \end{cases}$$

Intuitively speaking, the operator $[\alpha \odot T]$ attenuates the deformation induced by T. When $0 < \alpha \le 1$, the map moves each point x only a fraction α of the way along the transport path from x to $T(x)$, thereby "contracting" the action of T toward the identity. When $\alpha \le 0$, the movement is partly reversed, shifting x in the opposite direction of the original transport and producing a reflected contraction. In short, the α-contraction generates a controlled and

smoothly scaled version of the transformation induced by T, forming the building block of the ATM dynamics.

3. ATM model and estimation

This section provides concise background information on ATM models in Zhu and Müller (2023). Section 3.1 introduces the model setup, and Section 3.2 derives the asymptotic distribution of the estimated parameters under the assumption that the underlying distributional data are fully observed.

3.1 Autoregressive Transport Map models

Let $\{\mu_i\}_{i=1}^n \subset \mathcal{W}_2(\Omega)$ be a sequence of stationary probability measures on Ω, with the Wasserstein barycenter defined as

$$\mu_F = \arg \min_{\nu \in \mathcal{W}_2(\Omega)} \mathbb{E} d_W^2(\nu, \mu_i).$$

In this paper, we focus on the associated optimal transport maps of the form

$$T_i = F_{\mu_i}^{-1} \circ F_{\mu_F}, \qquad i = 1, \ldots, n, \tag{3.1}$$

which transport the barycenter measure μ_F to each μ_i.

Following Zhu and Müller (2023), the first-order autoregressive transport map model, denoted by ATM(1), is

$$T_i = T_{\epsilon_i} \circ [\alpha \odot T_{i-1}], \qquad i = 2, \ldots, n, \tag{3.2}$$

Alternatively, one may model the incremental transports by considering $T_i = F_{\mu_i}^{-1} \circ F_{\mu_{i-1}}$, which captures the incremental evolution of the distributions over time.

Statistica Sinica: Newly accepted Paper

where $\alpha \in (-1, 1)$, and $\{T_{\epsilon_i}\}$ is a sequence of i.i.d. innovation distortion maps satisfying $\mathbb{E}\{T_{\epsilon_i}(x)\} = x$ for all $x \in \Omega$ almost surely.

The key idea of the ATM model is to mimic the classical autoregressive dynamics within the Wasserstein space by replacing vector addition and scalar multiplication with their transport-map counterparts, namely composition and α-contraction, respectively. Specifically, the map T_i evolves by composing the α-contracted previous map $[\alpha \odot T_{i-1}]$ with a random innovation map T_{ϵ_i}, thereby capturing temporal dependence in the evolution of probability measures $\{\mu_i\}$. Under appropriate contraction and moment conditions, the sequence $\{T_i\}$ admits a unique stationary distribution in $\mathcal{T}$, and the induced process $\{\mu_i\}$ is stationary in $\mathcal{W}_2(\Omega)$.

For the ATM(1) model in (3.2), for any transport map $t \in \mathcal{T}$ and $\alpha \in (-1, 1)$, define

$$\phi_\epsilon(t) = T_\epsilon \circ [\alpha \odot t], \qquad \widetilde{\phi}_{i,m}(t) = \phi_{\epsilon_i} \circ \phi_{\epsilon_{i-1}} \circ \cdots \circ \phi_{\epsilon_{i-m+1}}(t).$$

Assumption 3.1. The sequence $\{T_{\epsilon_i}\}$ consists of i.i.d. distortion maps satisfying $\mathbb{E}\{T_{\epsilon_i}(x)\} = x$ almost surely for all $x \in \Omega$. Moreover, there exist constants $C > 0$ and $r \in (0, 1)$, such that for all $t, t' \in \mathcal{T}$ and all $m \in \mathbb{N}$,

$$\mathbb{E}\left[d_1\left(\widetilde{\phi}_{i,m}(t), \widetilde{\phi}_{i,m}(t')\right)\right] \leq C r^m d_1\left(t, t'\right),$$

where $d_1(t, s) = \int_\Omega \|t(x) - s(x)\| \mathrm{d}x$ is the L_1 distance for $t, s \in \mathcal{T}$.

Assumption 3.1 imposes a moment contraction condition (Wu and Shao, 2004) on the ATM(1) dynamics, which guarantees the existence and uniqueness of a stationary solution to (3.2). Similar conditions have been adopted in Zhu and Müller (2023) and Ghodrati and

Statistica Sinica: Newly accepted Paper

Panaretos (2024); we employ the L_1 distance metric here, primarily to simplify the estimation theory.

The following proposition is a direct consequence of Theorem 1 in Zhu and Müller (2023).

Proposition 3.1. Under Assumption 3.1, for any $T_0 \in \mathcal{T}$, the limit $\widetilde{T}_i := \lim_{m\to\infty} \widetilde{\phi}_{i,m}(T_0) \in \mathcal{T}$ exists almost surely and does not depend on T_0. Moreover, $\widetilde{T}_i$ is the unique stationary solution to the ATM(1) model (3.2).

3.2 Estimation

Suppose the true model (3.2) is parameterized by $\alpha_0 \in (-1, 1)$, then it is the unique minimizer of

$$l(\alpha) = \mathbb{E} \int_\Omega \{T_{i+1}(x) - [\alpha \odot T_i](x)\}^2 \, \mathrm{d}x.$$

In other words, we have that

$$\alpha_0 = \begin{cases} N_+/D_+, & \text{if } 0 < \alpha_0 < 1, \\ N_-/D_-, & \text{if } -1 < \alpha_0 \le 0, \end{cases}$$

where

$$N_+ = \int_\Omega \mathbb{E}\{(T_{i+1}(x) - x)(T_i(x) - x)\} \, \mathrm{d}x, \quad D_+ = \int_\Omega \mathbb{E}\{(T_i(x) - x)^2\} \, \mathrm{d}x,$$
$$N_- = \int_\Omega \mathbb{E}\{(T_{i+1}(x) - x)(x - T_i^{-1}(x))\} \, \mathrm{d}x, \quad D_- = \int_\Omega \mathbb{E}\{(x - T_i^{-1}(x))^2\} \, \mathrm{d}x.$$

Therefore, following Zhu and Müller (2023), we estimate α_0 by the least square type estimator,

$$\widehat{\alpha}_n = \widehat{\alpha}_+ \mathbb{I}\{l_+(\widehat{\alpha}_+) \le l_-(\widehat{\alpha}_-)\} + \widehat{\alpha}_- \mathbb{I}\{l_+(\widehat{\alpha}_+) > l_-(\widehat{\alpha}_-)\}, \tag{3.3}$$

Statistica Sinica: Newly accepted Paper

where $\widehat{\alpha}_+ = \operatorname{argmin}_{\alpha\in(0,1)} l_+(\alpha)$, $\widehat{\alpha}_- = \operatorname{argmin}_{\alpha\in(-1,0]} l_-(\alpha)$, and

$$l_+(\alpha) = \sum_{i=2}^{n} \int_\Omega \{T_i(x) - x - \alpha(T_{i-1}(x) - x)\}^2 \, \mathrm{d}x,$$
$$l_-(\alpha) = \sum_{i=2}^{n} \int_\Omega \left\{T_i(x) - x - \alpha(x - T_{i-1}^{-1}(x))\right\}^2 \mathrm{d}x.$$

Note that the estimation step involves comparing two different objective functions corresponding to positive and negative values of the contraction parameter α, making the theoretical analysis substantially more involved than that for classical least-squares estimators. To this end, we employ the concentration inequality developed in Wu and Shao (2004) and show that, the estimation of the sign of α_0 is super-consistent, thereby ensuring that the estimator $\widehat{\alpha}_n$ correctly identifies the contraction regime with probability approaching one. This is crucial for establishing the subsequent asymptotic linearity and normality of $\widehat{\alpha}_n$.

Theorem 3.1. Let $\{T_i\}_{i=1}^n$ be a sequence of observed optimal transportation maps generated by model (3.2) satisfying $\mathbb{E}\int_\Omega (T_i(x) - x)^2 \mathrm{d}x \geq \underline{c}$ for some constant $\underline{c} > 0$. Let $\widehat{\alpha}_n$ be defined in (3.3). Under Assumption 3.1, as $n \to \infty$,

$$\sqrt{n}(\widehat{\alpha}_n - \alpha_0) = \frac{1}{\sqrt{n}} \sum_{i=2}^{n} m_i(\alpha_0) + o_p(1),$$

where for $\Xi = \mathbb{E}\int_\Omega [T_i(x) - x]^2 \mathrm{d}x$, $\widetilde{\Xi} = \mathbb{E}\int_\Omega [T_i^{-1}(x) - x]^2 \mathrm{d}x$, we define

$$m_i(\alpha) = \begin{cases} \Xi^{-1} \displaystyle\int_\Omega \{T_i(x) - [\alpha \odot T_{i-1}](x)\}(T_{i-1}(x) - x)\mathrm{d}x, & \alpha > 0, \\ \widetilde{\Xi}^{-1} \displaystyle\int_\Omega \{T_i(x) - [\alpha \odot T_{i-1}](x)\}(x - T_{i-1}^{-1}(x))\mathrm{d}x, & \alpha \leq 0. \end{cases}$$

Statistica Sinica: Newly accepted Paper

Furthermore, for the natural filtration $\mathcal{F}_i = \sigma(T_j, j \leq i)$, $\{m_i(\alpha_0); \mathcal{F}_i\}$ forms a martingale difference sequence.

The lower bound condition $\mathbb{E}\int_\Omega (T_i(x)-x)^2 \mathrm{d}x \geq \underline{c}$ is imposed solely to exclude degenerate cases of the transport map of $T_i(x)$. As otherwise, $T_i(x) = x$ for $x \in \Omega$ a.s., which corresponds to a trivial static setting. By the boundedness of $T_i \in \mathcal{T}$, the classical central limit theorem for martingales (Hall and Heyde, 2014) and Theorem 3.1 together imply that $\widehat{\alpha}_n$ is asymptotically normal, i.e.

$$\sqrt{n}(\widehat{\alpha}_n - \alpha_0) \to_d \mathcal{N}\left(0, \mathbb{E}[m_i^2(\alpha_0)]\right), \quad \text{as} \quad n \to \infty.$$

In this paper, we assume the data are fully observed, whereas Zhu and Müller (2023) allow for discretely observed distributions, where the transport maps can be estimated from empirical grid observations. Compared with the convergence rate result in Zhu and Müller (2023), Theorem 3.1 further establishes the asymptotic distribution of the estimator $\widehat{\alpha}_n$, thereby enabling valid statistical inference for the contraction parameter α. A similar fully observed setting is also considered in Ghodrati and Panaretos (2024), where an additional "intercept" term is introduced to capture the mean shape of the transport maps. However, extending the asymptotic normality result to the case of estimated or partially observed maps is considerably more challenging, as the estimation error of the empirical transport maps propagates nonlinearly through the composition and contraction operations. In Section 4.4, we discuss how to estimate the transport maps under discrete observations and outline how the procedure can be adapted. In this sense, the fully observed setting considered in this paper can be viewed as an ideal dense-observation regime.

4. Diagnostic Tests

In this section, we develop diagnostic procedures to evaluate the adequacy of the ATM(1) model, for which formal asymptotic theory can be established. Section 4.1 formalizes residual transport maps and their autocorrelation functions, which serve as the primary tools for assessing model fit. Section 4.2 introduces McLeod (McLeod, 1978) type portmanteau tests to detect serial dependence in the residuals, and Section 4.3 develops the sample-splitting version of the diagnostic test. Section 4.4 collects the detailed implementation procedure. Finally, section 4.5 discusses extensions to ATM(p) models, explains why the sample-splitting test may fail for higher-order models.

4.1 Residuals and auto-correlation functions

To assess the adequacy of the fitted ATM model, we define residual maps that quantify the deviation between the observed and fitted transport maps. By model (3.2), with $\{T_i\}$ and model parameter α_0, we can recover the distortion map as

$$T_i \circ [\alpha_0 \odot T_{i-1}]^{-1} = T_{\epsilon_i} \circ [\alpha_0 \odot T_{i-1}] \circ [\alpha_0 \odot T_{i-1}]^{-1} = T_{\epsilon_i}.$$

Therefore, given the estimated contraction parameter $\widehat{\alpha}_n$, the residual is naturally defined as

$$\widehat{T}_{\epsilon_i} = T_i \circ [\widehat{\alpha}_n \odot T_{i-1}]^{-1}, \quad i = 2, \ldots, n. \tag{4.4}$$

Under the correctly specified model, the estimated innovation maps $\{\widehat{T}_{\epsilon_i}\}$ should mimic the behavior of the true innovations $\{T_{\epsilon_i}\}$. Similar to the classical Euclidean autoregressive model, the temporal dependence of $\{T_{\epsilon_i}\}$ can be assessed via the autocorrelation function

Statistica Sinica: Newly accepted Paper

(ACF).

Definition 4.1. The theoretical ACF at lag $k \geq 1$ for the distortion map process $\{T_{\epsilon_i}\}$ is given by

$$\rho(k) = \frac{\int_\Omega \mathbb{E}\left[\left(T_{\epsilon_{i+k}}(x) - x\right)\left(T_{\epsilon_i}(x) - x\right)\right] \mathrm{d}x}{\int_\Omega \mathbb{E}\left[\left(T_{\epsilon_i}(x) - x\right)^2\right] \mathrm{d}x}.$$

Here, the definition of ACF projects the distortion maps onto $L^2(\Omega)$ space and assesses temporal dependence by integrating the pointwise autocorrelations of $T_{\epsilon_i}(x)$ over Ω. While possibly the pointwise correlations may cancel out during integration, the resulting global autocorrelation measure still provides a meaningful overall measure of temporal dependence. It serves as the basis for constructing the portmanteau statistic that follows.

4.2 Portmanteau Test

Given estimated residual transports $\widehat{T}_{\epsilon_2}, \cdots, \widehat{T}_{\epsilon_n}$ defined in (4.4), the corresponding sample ACF at lag k is

$$\widehat{\rho}_n(k) = \frac{\int_\Omega \sum_{i=2}^{n-k} \left[\left(\widehat{T}_{\epsilon_{i+k}}(x) - x\right)\left(\widehat{T}_{\epsilon_i}(x) - x\right)\right] \mathrm{d}x}{\int_\Omega \sum_{i=2}^{n} \left[\left(\widehat{T}_{\epsilon_i}(x) - x\right)^2\right] \mathrm{d}x}. \tag{4.5}$$

If $\widehat{\alpha}_n$ consistently estimates the true parameter α, then the estimated residual maps $\widehat{T}_{\epsilon_i}$ should approximate the true distortion maps T_{ϵ_i}, thereby preserving their statistical properties. Hence, under the correctly specified model, the sample autocorrelations $\widehat{\rho}_n(k)$ in (4.5) are expected to be close to zero for all $k \geq 1$, reflecting the absence of serial dependence among the residual transports. Significant deviations from zero would indicate potential model mis-specification, motivating the use of portmanteau-type tests.

While similar ideas have been widely used in the classical Euclidean time-series literature—see e.g. Ljung and Box (1978) and McLeod (1978), the extension to the Wasserstein

Statistica Sinica: Newly accepted Paper

setting requires substantial additional work due to the nonlinear structure of transport maps. Before presenting the asymptotic distribution of $\widehat{\rho}_n(k)$, we introduce additional notations that will be used in the subsequent analysis.

Define $G_i(\alpha, x) = T_i \circ [\alpha \odot T_{i-1}]^{-1}(x)$, and let $g_i(\alpha, x)$ be the derivative of $G_i(\alpha, x)$ with respect to α, where by equation (31) in Ghodrati and Panaretos (2024),

$$
\begin{aligned}
g_i(\alpha, x) &= \frac{\partial G_i(\alpha, x)}{\partial \alpha} \\
&= T_i'(z) \times \begin{cases} \left(z - T_{i-1}(z)\right) \times \frac{1}{\alpha\left(T_{i-1}'(z)-1\right)+1}\Big|_{z=[\alpha\odot T_{i-1}]^{-1}(x)}, & 0 \le \alpha < 1, \\ \left(T_{i-1}^{-1}(z) - z\right) \times \frac{1}{\alpha\left(1-\left(T_{i-1}^{-1}\right)'(z)\right)+1}\Big|_{z=[\alpha\odot T_{i-1}]^{-1}(x)}, & -1 < \alpha \le 0. \end{cases}
\end{aligned} \tag{4.6}
$$

Fix an integer $K \ge 1$, for $1 \le k \le K$, define $M_1 = (\mu_{11}, \cdots, \mu_{1K})^\top$, $M_2 = (\mu_{21}, \cdots, \mu_{2K})^\top$, where

$$
\begin{aligned}
\mu_{1k} &= \mathbb{E}\left[\int_\Omega (T_{\epsilon_0}(x) - x) g_k(\alpha_0, x) \mathrm{d}x\right], \\
\mu_{2k} &= \mathbb{E}\left[\int_\Omega \left[(T_{\epsilon_0}(x) - x)\left(T_{\epsilon_k}(x) - x\right)\right] \mathrm{d}x \cdot m_k(\alpha_0)\right].
\end{aligned}
$$

Furthermore, let

$$
\begin{aligned}
\sigma_1^2 &= \mathbb{E}\left[\int_\Omega \{T_{\epsilon_i}(x) - x\}^2 \, \mathrm{d}x\right], \\
\sigma_2^4 &= \mathbb{E}\left[\left\{\int_\Omega [T_{\epsilon_1}(x) - x][T_{\epsilon_2}(x) - x] \, \mathrm{d}x\right\}^2\right].
\end{aligned}
$$

Statistica Sinica: Newly accepted Paper

Assumption 4.1. There exists a constant $c > 0$ such that

$$\sup_{\alpha\in[-1,1]} \mathbb{E}\int_{\Omega}[g(\alpha,x)]^2\mathrm{dx} \le c < \infty.$$

Assumption 4.1 imposes a mild second moment condition on the derivative function. It is a regularity requirement on the local sensitivity of the residual map with respect to the contraction parameter. Under compact support and smooth innovation maps, the transport maps and their inverses are bounded, so the derivative $g_i(\alpha, x)$ in (4.6) is expected to be uniformly square integrable; this covers the smooth innovation-map families used in the paper.

Theorem 4.1. Under the correct model specification (3.2) with parameter $\alpha_0 \in (-1, 1)$, let $\{\widehat{T}_{\epsilon_i}\}_{i=2}^n$ be the residual maps defined in (4.4). For any fixed integer $K \ge 1$, under Assumption 3.1 and 4.1, as $n \to \infty$, we have that

$$\sqrt{n}\widehat{\rho}_n^{\top} = \sqrt{n}\left(\widehat{\rho}_n(1), \cdots, \widehat{\rho}_n(K)\right)^{\top} \to_d \mathcal{N}(\mathbf{0}, \Sigma_K),$$

where

$$\Sigma_K := \sigma_1^{-4}\left\{\sigma_2^4 I_K + (M_1M_2' + M_2M_1' + M_1\mathbb{E}[m_i^2(\alpha_0)]M_1^{\top})\right\}.$$

Theorem 4.1 establishes the joint asymptotic normality of the sample autocorrelations of the residual transport maps. The presence of M_1 and M_2 reflects the nonlinear dependence of the residuals on the estimated contraction parameter α_0 via the transport-map derivatives g_i. The additional term $M_1\,\mathbb{E}[m_i^2(\alpha_0)]\,M_1^{\top}$ captures the accumulated effect of parameter estimation on the limiting distribution in Theorem 3.1, analogous to the correction introduced

Statistica Sinica: Newly accepted Paper

by McLeod (1978); Li and McLeod (1981) for finite-sample adjustment to the classical Ljung-Box test in Ljung and Box (1978).

In practice, we estimate the covariance Σ_K by its sample counterpart

$$\widehat{\Sigma}_K = \widehat{\sigma}_1^{-4}\left\{\widehat{\sigma}_2^4 I_K + \left(\widehat{M_1}\widehat{M_2}^\top + \widehat{M_2}\widehat{M_1}^\top + \widehat{M_1}\widehat{v}_m\widehat{M_1}^\top\right)\right\},$$

where $\widehat{M_1} = (\widehat{\mu}_{11}, \cdots, \widehat{\mu}_{1K})^\top$ and $\widehat{M_2} = (\widehat{\mu}_{21}, \cdots, \widehat{\mu}_{2K})^\top$ are K-vectors, and $\widehat{v}_m = n^{-1}\sum_{i=2}^n \widehat{m}_i^2(\widehat{\alpha}_n)$. For $1 \le k \le K$,

$$\begin{aligned}
\widehat{\mu}_{1k} =& \frac{1}{n-k-1}\sum_{i=2}^{n-k}\left[\int_\Omega (\widehat{T}_{\epsilon_i}(x) - x) g_{i+k}(\widehat{\alpha}_n, x)\mathrm{d}x\right], \\
\widehat{\mu}_{2k} =& \frac{1}{n-k-1}\sum_{i=2}^{n-k}\left[\int_\Omega \left[\left(\widehat{T}_{\epsilon_i}(x) - x\right)\left(\widehat{T}_{\epsilon_{i+k}}(x) - x\right)\right]\mathrm{d}x \cdot \widehat{m}_{i+k}(\widehat{\alpha}_n)\right],
\end{aligned}$$

where for $i = 2, \ldots, n$, the empirical influence function is

$$\widehat{m}_i(\alpha) = \begin{cases} \widehat{\Xi}^{-1}\displaystyle\int_\Omega \{T_i(x) - [\alpha \odot T_{i-1}](x)\}(T_{i-1}(x) - x)\mathrm{d}x, & \alpha \ge 0, \\ \widehat{\widehat{\Xi}}^{-1}\displaystyle\int_\Omega \{T_i(x) - [\alpha \odot T_{i-1}](x)\}(x - T_{i-1}^{-1}(x))\mathrm{d}x, & \alpha < 0, \end{cases}$$

with $\widehat{\Xi} = n^{-1}\sum_{j=1}^{n-1}\int_\Omega [T_j(x) - x]^2\,\mathrm{d}x$, and $\widehat{\widehat{\Xi}} = n^{-1}\sum_{j=1}^{n-1}\int_\Omega [T_j^{-1}(x) - x]^2\,\mathrm{d}x$.

Finally,

$$\begin{aligned}
\widehat{\sigma}_1^2 &= \frac{1}{n-1}\sum_{i=2}^{n}\int_\Omega [\widehat{T}_{\epsilon_i}(x) - x]^2\,\mathrm{d}x, \\
\widehat{\sigma}_2^4 &= \frac{1}{n-2}\sum_{i=2}^{n-1}\left\{\int_\Omega [\widehat{T}_{\epsilon_i}(x) - x][\widehat{T}_{\epsilon_{i+1}}(x) - x]\,\mathrm{d}x\right\}^2.
\end{aligned} \tag{4.7}$$

Statistica Sinica: Newly accepted Paper

Then, the McLeod type portmanteau test statistic is formulated as

$$Q_n = n\,\widehat{\rho}_n^{\top}\widehat{\Sigma}_K^{-1}\,\widehat{\rho}_n,$$

and we reject the null hypothesis of correct model specification if

$$Q_n > \chi^2_K(1-\beta), \tag{4.8}$$

the $(1-\beta)$ upper quantile of the chi-square distribution with K degrees of freedom.

Remark 4.1. In classical univariate ARMA models, portmanteau test statistics are typically constructed as

$$Q_n^* = n\sum_{k=1}^{K} w_n(k)\,\widehat{\rho}_n^2(k), \tag{4.9}$$

where $w_n(k) = 1$ corresponds to the Box–Pierce statistic (Box and Pierce, 1970), and $w_n(k) = (n+2)/(n-k)$ yields the Ljung–Box statistic (Ljung and Box, 1978). To improve the finite-sample approximation, Li and McLeod (1981) further proposed a modified statistic of the form

$$Q_n^{\dagger} = Q_n^* + \frac{K(K+1)}{2n}. \tag{4.10}$$

Then, the null hypothesis is rejected when Q_n^* (or $Q_n^{\dagger}$) exceeds the $(1-\alpha)$ quantile of a chi-squared distribution with $K-m$ degrees of freedom, where m denotes the total number of estimated ARMA parameters. This chi-square reference relies on an *increasing-lag* asymptotic framework, that is, $K = K(n) \to \infty$ as $n \to \infty$ with $n \gg K \gg 1$, see Chapter 2 in Li (2003) for a detailed discussion. In this diverging-K regime, the effective degrees of freedom reduce from K to $K-m$ because the covariance matrix of residual autocorrelations becomes

Statistica Sinica: Newly accepted Paper

increasingly singular (and thus not directly invertible) as K grows. Consequently, the classical portmanteau tests of the form (4.9) or (4.10) can be implemented without inverting the covariance matrix, and instead rely on weighting schemes or small-sample corrections to improve the chi-square approximation.

However, in this paper (and arguably in typical empirical applications), we treat K as a fixed integer. In this fixed-K regime, the asymptotic covariance matrix of the residual ACF is non-singular, and following the recommendations of McLeod (1978), Ansley and Newbold (1979), and Newbold (1980), we explicitly estimate this covariance matrix to account for the parameter estimation effect. This kind of treatment has been also used in diagnostic checks for nonlinear time series models, see e.g. Li (1992) and Li and Mak (1994).

4.3 Sample Splitting Approach

Despite the finite-sample correction for estimation effects provided in Theorem 4.1, estimating quantities such as M_1, M_2, and $\mathbb{E}[m_i^2(\alpha_0)]$ may be challenging in practice, as they involve expectations of nonlinear functionals of transport maps and their derivatives, especially in moderate sample sizes. This motivates the development of alternative diagnostic procedures that avoid direct estimation of these higher-order terms while still accounting for the impact of parameter estimation.

Sample splitting has recently emerged as an effective tool for time-series inference, with theoretical guarantees established in several distinct settings; see, for example, Lunde (2019), Chang et al. (2022), Zhang and Shao (2024), and Gao et al. (2025). Of particular relevance is the work of Davis and Fernandes (2025), who employed sample splitting to remove estimation effects in portmanteau tests applied to residuals of parametric time-series models, such as ARMA and GARCH models. Their methodology estimates the model parameters using

Statistica Sinica: Newly accepted Paper

only the first f_n observations, and then computes the autocorrelation function (ACF) of the residuals based solely on the last l_n observations. They show that as long as the relative overlap between the two subsamples converges to 1/2, the resulting ACF-based tests retain the same limiting distribution as if the underlying residuals were i.i.d. Consequently, the classical portmanteau test admits its convenient chi-square limiting distribution when sample splitting with appropriately chosen (f_n, l_n) is employed—for instance, $f_n = n/2$ together with $l_n = n$.

Here, we extend the idea of Davis and Fernandes (2025) to the non-Euclidean setting of distributional data. In particular, let $\widehat{\alpha}_{f_n}$ be the sample-splitting analogue of (3.3) using first f_n observations $\{T_i\}_{i=1}^{f_n}$, and define the residual maps by

$$\widehat{T}_{\epsilon_i}^{f_n} = T_i \circ [\widehat{\alpha}_{f_n} \odot T_{i-1}]^{-1}, \quad i = n - l_n + 2, \ldots, n, \tag{4.11}$$

which reduces to the standard case in (4.4) when $f_n = l_n = n$. Similarly, for $1 \le k \le K$, we define the lag-k sample autocorrelation function as

$$\widehat{\rho}_{l_n}(k) = \frac{\int_\Omega \sum_{i=n-l_n+2}^{n-k} \left[\left(\widehat{T}_{\epsilon_{i+k}}^{f_n}(x) - x \right) \left(\widehat{T}_{\epsilon_i}^{f_n}(x) - x \right) \right] \mathrm{d}x}{\int_\Omega \sum_{i=n-l_n+2}^{n} \left[\left(\widehat{T}_{\epsilon_i}^{f_n}(x) - x \right)^2 \right] \mathrm{d}x}. \tag{4.12}$$

The next result generalizes Theorem 4.1 to the sample-splitting case.

Theorem 4.2. Under the correct model specification (3.2) with parameter $\alpha_0 \in (-1, 1)$, let $\{\widehat{T}_{\epsilon_i}^{f_n}\}_{i=n-l_n+2}^{n}$ be the residual maps defined in (4.11). Suppose the split sequence (f_n, l_n)

Statistica Sinica: Newly accepted Paper

satisfies that $\min\{f_n, l_n\} \to \infty$ as $n \to \infty$. Let

$$k_{ra} = \lim_{n\to\infty} \frac{l_n}{f_n} < \infty \quad \text{and} \quad k_{ov} = \lim_{n\to\infty} \frac{\max\{0,\ f_n + l_n - n\}}{f_n} < \infty,$$

denote the limiting ratio and limiting overlap coefficient, respectively. Then, for any fixed integer $K \geq 1$, under Assumption 3.1 and 4.1, as $n \to \infty$,

$$\sqrt{l_n}\widehat{\rho}_{l_n}^{\top} = \sqrt{l_n}(\widehat{\rho}_{l_n}(1), \cdots, \widehat{\rho}_{l_n}(K))^{\top} \to_d \mathcal{N}(\mathbf{0}, \Sigma_K^S),$$

where

$$\Sigma_K^S := \sigma_1^{-4}\left\{\sigma_2^4 I_K + (k_{ov} M_1 M_2^{\top} + k_{ov} M_2 M_1^{\top} + k_{ra} M_1 \mathbb{E}[m_i^2(\alpha_0)] M_1^{\top})\right\}.$$

Furthermore, if

$$M_2 = -M_1 \mathbb{E}[m_i^2(\alpha_0)], \tag{4.13}$$

then Σ_K^S simplifies to

$$\Sigma_K^S = \sigma_1^{-4}\left\{\sigma_2^4 I_K + (k_{ra} - 2k_{ov}) M_1 \mathbb{E}[m_i^2(\alpha_0)] M_1^{\top}\right\}.$$

The result in Theorem 4.2 mirrors that of Theorem 4.1, with the key difference being the appearance of the split sequence (f_n, l_n) and its limiting behavior. In particular, when $f_n = l_n = n$, Theorem 4.2 reduces exactly to Theorem 4.1.

In Theorem 4.2, the condition (4.13) that $M_2 = -M_1 \mathbb{E}[m_i^2(\alpha_0)]$ is motivated by the findings of Davis and Fernandes (2025), who showed that it holds for classical ARMA and GARCH models. Establishing this identity analytically for the ATM model is substantially

Statistica Sinica: Newly accepted Paper

more difficult, as it involves functional derivatives such as those appeared in (4.6). To assess its validity in the ATM setting, we therefore conduct a numerical investigation under the ATM(1) specification.

We consider three different forms of the innovation map $T_\epsilon(x)$: trigonometric, power, and polynomial.

$$e_1(x) = x + \frac{\sin(Y_1 \pi x)}{|Y_1|\pi},$$
$$e_2(x) = x + Y_2(x^2 - x),$$
$$e_3(x) = x + Y_3 x^2 (1-x)^2,$$

where Y_1 is a random variable generated from a discrete uniform distribution over $\{\pm 5, \pm 6, \cdots, \pm 15\}$, and Y_2, Y_3 are random variables generated from a uniform distribution over $[-1, 1]$.

We numerically evaluate condition (4.13) at four values of the contraction parameter, -0.4, -0.2, 0.2, and 0.5, with $K = 12$. The verification for the case $K = 3, 6, 9$ is included in Section S.6 and Table S.7 of the Supplement, and the results are similar. Specifically, we generate $n = 5000$ observations from model (3.2) and compute empirical estimates of the two sides of (4.13). Table 1 reports the mean L^1 and L^2 distances between them over 1000 Monte Carlo repetitions, with standard deviations shown in parentheses. The discrepancies are uniformly close to zero across signs and magnitudes of α_0. The variation is more visible across innovation families than across coefficient signs, with the power-type innovation e_2 producing the largest, but still very small, discrepancies.

Another direct consequence of condition (4.13) is its practical implication: when the splitting sequence satisfies $k_{ra} = 2k_{ov}$, the resulting covariance matrix Σ_K^S becomes nearly

Statistica Sinica: Newly accepted Paper

Table 1: Numerical validation of condition (4.13) with $K = 12$ based on 1000 Monte Carlo repetitions. Each entry is mean (standard deviation).

Function	Difference	Mean (Standard Deviation)			
		$\alpha = 0.5$	$\alpha = 0.2$	$\alpha = -0.2$	$\alpha = -0.4$
$e_1(x)$	L^1	$2.14(0.38) \times 10^{-6}$	$1.89(0.44) \times 10^{-6}$	$2.58(0.44) \times 10^{-6}$	$2.14(0.42) \times 10^{-6}$
	L^2	$2.83(0.46) \times 10^{-6}$	$2.33(0.53) \times 10^{-6}$	$3.59(0.43) \times 10^{-6}$	$2.65(0.49) \times 10^{-6}$
$e_2(x)$	L^1	$1.31(0.19) \times 10^{-4}$	$1.07(0.23) \times 10^{-4}$	$1.28(0.22) \times 10^{-4}$	$9.84(2.15) \times 10^{-5}$
	L^2	$1.67(0.16) \times 10^{-4}$	$1.31(0.28) \times 10^{-4}$	$1.63(0.21) \times 10^{-4}$	$1.21(0.26) \times 10^{-4}$
$e_3(x)$	L^1	$4.28(0.96) \times 10^{-6}$	$4.90(1.15) \times 10^{-6}$	$4.88(1.12) \times 10^{-6}$	$4.49(1.01) \times 10^{-6}$
	L^2	$5.31(1.13) \times 10^{-6}$	$6.15(1.36) \times 10^{-6}$	$6.08(1.32) \times 10^{-6}$	$5.60(1.20) \times 10^{-6}$

pivotal, up to nuisance parameters σ_1 and σ_2.

In particular, if we choose $f_n = n/2$ and $l_n = n$, then $k_{ra} = 2$ and $k_{ov} = 1$, so that

$$\Sigma_K^S = \frac{\sigma_2^4}{\sigma_1^4} I_K.$$

This split is also attractive in finite samples. The first half of the sample is used to obtain a stable parameter estimate, while the full sample is retained for residual autocorrelation estimation, avoiding an unnecessary loss of information in the diagnostic step. Other choices of (f_n, l_n) are possible but they may result in further efficiency loss, and hence power.

Accordingly, we estimate Σ_K^S by $\widehat{\Sigma}_K^S = \widehat{\sigma}_{2,S}^4 / \widehat{\sigma}_{1,S}^4 \cdot I_K$, where

$$\begin{aligned} \widehat{\sigma}_{1,S}^2 &= \frac{1}{l_n - 1} \sum_{i=n-l_n+2}^{n} \int_\Omega [\widehat{T}_{\epsilon_i}^{f_n}(x) - x]^2 \, \mathrm{d}x, \\ \widehat{\sigma}_{2,S}^4 &= \frac{1}{l_n - 2} \sum_{i=n-l_n+2}^{n-1} \left\{ \int_\Omega [\widehat{T}_{\epsilon_i}^{f_n}(x) - x][\widehat{T}_{\epsilon_{i+1}}^{f_n}(x) - x] \, \mathrm{d}x \right\}^2. \end{aligned} \tag{4.14}$$

Therefore, the sample-splitting-based test statistic is

$$Q_{l_n}^S = l_n \widehat{\rho}_{l_n}^\top (\widehat{\Sigma}_K^S)^{-1} \widehat{\rho}_{l_n}, \tag{4.15}$$

Statistica Sinica: Newly accepted Paper

and we reject the null hypothesis of correct model specification if $Q_{l_n}^S > \chi_K^2(1-\beta)$, with significance level $\beta \in (0,1)$.

4.4 Practical Implementation

For the ATM(1) model, the proposed tests can be implemented as follows.

(i) *Estimation.* For the McLeod-type test, estimate α by $\widehat{\alpha}_n$ in (3.3) using the full observations $\{T_i\}_{i=1}^n$; for the sample-splitting test, estimate α by $\widehat{\alpha}_{f_n}$ using the first f_n observations.

(ii) *Construct residual maps.* For the McLeod-type test, let $\widehat{T}_{\epsilon_i} = T_i \circ [\widehat{\alpha}_n \odot T_{i-1}]^{-1}$; for the sample-splitting test, compute $\widehat{T}_{\epsilon_i}^{f_n} = T_i \circ [\widehat{\alpha}_{f_n} \odot T_{i-1}]^{-1}$ for the last l_n residuals.

(iii) *Form the test statistic.* For the McLeod-type test, compute $\widehat{\rho}_n(k)$ in (4.5), $k = 1, \ldots, K$, and form Q_n in (4.8); for the sample-splitting test, compute $\widehat{\rho}_{l_n}(k)$, $k = 1, \ldots, K$, in (4.12) and form $Q_{l_n}^S$ in (4.15).

(iv) *Draw conclusion.* Given a significance level $\beta \in (0,1)$, compare either Q_n or $Q_{l_n}^S$ with the critical value $\chi_K^2(1-\beta)$, and reject the ATM(1) specification when the statistic exceeds this critical value.

Although we assume that the transport maps T_i are fully observed in the theoretical development, they generally need to be estimated in practice. Following Zhu and Müller (2023), if each distribution μ_i is observed through raw samples $X_{i1}, \ldots, X_{iN_i}$, $N_i > 0$, we first estimate its distribution function F_i or quantile function Q_i. For empirical quantiles, sort the observations and compute $\widehat{Q}_i(u_j) = X_{i(\lceil N_i u_j \rceil)}$ on a common grid $0 < u_1 < \cdots < u_J < 1$, where J is chosen sufficiently large, say $J = 1000$. Set $\widehat{Q}_F(u_j) = n^{-1}\sum_{i=1}^n \widehat{Q}_i(u_j)$, let $\widehat{F}_F = \widehat{Q}_F^{-1}$, and estimate the transport map from the barycenter to μ_i by $\widehat{T}_i = \widehat{Q}_i \circ \widehat{F}_F$, equivalently

Statistica Sinica: Newly accepted Paper

$\widehat{T}_i(\widehat{Q}_F(u_j)) = \widehat{Q}_i(u_j)$ on the grid. Values between adjacent grid points, as well as inverse transport maps, are obtained by monotone interpolation. All integrals involved in parameter estimation and test-statistic construction are approximated numerically on this common grid using the classical Riemann rule.

The number of observations N_i at each time point can affect the diagnostic analysis. It controls the accuracy of the estimated quantile functions, and the empirical transport maps; this error then propagates through the inverse-composition step used to construct the residual maps. Indeed, Theorem 2 in Zhu and Müller (2023) show that the convergence rate of the model parameter is of order $O_p(\tau + n^{-1/2})$, where $\tau = \sup_i \mathbb{E}d_1(\widehat{T}_i, T_i)$, is the uniform estimation error of the transport maps. When N_i is small, the empirical quantile functions may be noisy, particularly near the tails. Letting $N = \min_{1\le i\le n} N_i$, then by the discussions in Zhu and Müller (2023), the approximation error ranges from $\tau \sim N^{-1/4}$ to $\tau \sim N^{-1/3}$. Therefore, under sufficiently dense sampling, for example $N \gg n^2$, the transport-map estimation error is asymptotically negligible relative to $n^{-1/2}$. This justifies treating the transport maps as fully observed, as assumed in the present paper.

4.5 Scope of Extensions

Following Zhu and Müller (2023), a higher-order ATM(p) model is

$$T_i = T_{\epsilon_i} \circ [\alpha_1 \odot T_{i-1}] \circ \cdots \circ [\alpha_p \odot T_{i-p}], \qquad i = p+1, \ldots, n, \tag{4.16}$$

under suitable assumptions on $\theta = (\alpha_1, \ldots, \alpha_p)^\top$.

Although the proposed testing framework for ATM(1) can be conceptually extended to assess fitted ATM(p) models, the pivotal limiting distribution established for the sample-

Statistica Sinica: Newly accepted Paper

splitting test in the ATM(1) case does not carry over directly. In particular, for the ATM(1) model, the fitted residual is $\widehat{T}_{\epsilon_i} = T_i \circ [\widehat{\alpha}_n \odot T_{i-1}]^{-1}$, so its first-order expansion contains only one estimation error, $\widehat{\alpha}_n - \alpha_0$. The sample-splitting statistic becomes nearly pivotal when the effect of estimation cancels through condition (4.13).

However, the above argument does not extend directly to a general ATM(p) model. Consider, for example, the ATM(2) model, for which the residual map becomes $\widehat{T}_{\epsilon_i} = T_i \circ [\widehat{\alpha}_{2,n} \odot T_{i-2}]^{-1} \circ [\widehat{\alpha}_{1,n} \odot T_{i-1}]^{-1}$. Its expansion around T_{ϵ_i} contains separate terms involving $\widehat{\alpha}_{1,n} - \alpha_1$ and $\widehat{\alpha}_{2,n} - \alpha_2$. Consequently, the scalar cancellation condition in (4.13), which underlies the ATM(1) sample-splitting test, is replaced by a matrix-type condition.

The main difficulty is that, unlike in conventional ARMA models in Davis and Fernandes (2025) or the ATM(1) model in Section 4.3, the first-order expansion of the ATM(2) estimator does not involve a simple linear function of the lagged observations. Although its influence function remains linear in the one-step prediction error, the corresponding loading involves nested transport compositions and their derivatives, and therefore depends nonlinearly on both T_{i-1} and T_{i-2}. As a result, the estimation effect in the residual autocorrelations is no longer automatically aligned with the influence function of the estimator. Section S.3 of the Supplement provides the corresponding first-order expansion for estimators of the ATM(2) model, and presents numerical evidence showing that the discrepancy from the required matrix condition as (4.13) is generally non-negligible.

Therefore, we conjecture that extending the proposed tests to general ATM(p) models, ARMA-type transport dynamics, or other object spaces will require new model specifications, residual constructions, or even different geometric arguments. We leave these extensions for future research.

5. Simulation Results

This section investigates the finite-sample performance of the proposed portmanteau tests. All empirical rejection rates reported below are computed at the nominal 5% significance level. Following Ghodrati and Panaretos (2024), the distortion maps are generated according to

$$T_{\epsilon_i}(x) = x + \frac{\sin(Y\pi x)}{|Y|\pi}, \tag{5.17}$$

where Y is a random variable generated from a discrete uniform distribution over $\{\pm 5, \pm 6, \cdots, \pm 15\}$. Here we focus on the trigonometric innovation maps, the results of the power and polynomial innovation maps considered in Table 1 are similar, and hence relegated to Section S.4 of the Supplement.

To evaluate the size performance under the ATM(1) null, we vary $\alpha_0 \in \{-0.4, -0.2, 0.2, 0.5\}$, sample sizes $n \in \{100, 200, 500, 1000\}$, and lags $K \in \{3, 6, 9\}$. Table 2 reports empirical rejection rates for the covariance-corrected statistic Q_n in (4.8) and the sample-splitting statistic Q_n^S, based on 1000 Monte Carlo repetitions at the nominal 5% level.

Table 2 shows that finite-sample size distortion is visible when n is small, especially for larger lag orders. At $n = 100$, the sample-splitting statistic Q_n^S over-rejects more than Q_n in all configurations reported here. Its empirical size improves as n increases and is generally closer to the nominal level than that of Q_n at $n = 500, 1000$.

We then examine the power performance by setting ATM(1)-based diagnostics under ATM(2) alternatives in (4.16). We consider four parameter configurations, $(\alpha_1, \alpha_2) = \pm(0.2, 0.1)$ and $(\alpha_1, \alpha_2) = \pm(0.5, 0.2)$. The results are collected in Table 3. The rejection rates generally increase with the sample size, especially for the sample-splitting statistic Q_n^S. For a fixed

Table 2: Size performance under the ATM(1) null at the 5% nominal level based on 1000 Monte Carlo repetitions.

K	Method	$\alpha = 0.2$				$\alpha = 0.5$			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
3	Q_n	0.084	0.065	0.058	0.070	0.081	0.089	0.090	0.102
	Q_n^S	0.112	0.083	0.066	0.050	0.123	0.079	0.051	0.045
6	Q_n	0.120	0.091	0.074	0.082	0.130	0.119	0.105	0.084
	Q_n^S	0.160	0.121	0.071	0.054	0.156	0.116	0.074	0.053
9	Q_n	0.153	0.096	0.067	0.075	0.167	0.133	0.116	0.087
	Q_n^S	0.183	0.127	0.080	0.058	0.181	0.128	0.078	0.059
K	Method	$\alpha = -0.2$				$\alpha = -0.4$			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
3	Q_n	0.097	0.117	0.107	0.100	0.066	0.072	0.080	0.079
	Q_n^S	0.104	0.081	0.059	0.045	0.110	0.084	0.067	0.047
6	Q_n	0.121	0.116	0.127	0.124	0.116	0.104	0.080	0.090
	Q_n^S	0.162	0.113	0.065	0.054	0.162	0.111	0.074	0.056
9	Q_n	0.144	0.111	0.136	0.146	0.157	0.122	0.097	0.089
	Q_n^S	0.184	0.123	0.078	0.064	0.178	0.123	0.084	0.061

sample size, smaller lag orders often give higher power than larger choices of K, suggesting that including large lags could diminish the signal in finite samples.

Table 3: Power performance under ATM(2) alternatives at the 5% nominal level based on 1000 Monte Carlo repetitions.

K	Method	$(\alpha_1,\alpha_2)=(0.2,0.1)$				$(\alpha_1,\alpha_2)=(0.5,0.2)$			
		$n=100$	$n=200$	$n=500$	$n=1000$	$n=100$	$n=200$	$n=500$	$n=1000$
3	Q_n	0.473	0.765	0.962	0.963	0.949	0.981	0.999	1.000
	Q_n^S	0.604	0.872	1.000	1.000	0.992	1.000	1.000	1.000
6	Q_n	0.386	0.654	0.939	0.969	0.910	0.953	0.990	0.998
	Q_n^S	0.499	0.793	1.000	1.000	0.978	1.000	1.000	1.000
9	Q_n	0.367	0.584	0.899	0.949	0.889	0.950	0.981	0.995
	Q_n^S	0.452	0.703	0.997	1.000	0.957	1.000	1.000	1.000
K	Method	$(\alpha_1,\alpha_2)=-(0.2,0.1)$				$(\alpha_1,\alpha_2)=-(0.5,0.2)$			
		$n=100$	$n=200$	$n=500$	$n=1000$	$n=100$	$n=200$	$n=500$	$n=1000$
3	Q_n	0.442	0.666	0.878	0.919	0.802	0.844	0.825	0.846
	Q_n^S	0.521	0.817	0.997	1.000	0.992	1.000	1.000	1.000
6	Q_n	0.349	0.570	0.853	0.904	0.808	0.861	0.832	0.825
	Q_n^S	0.431	0.713	0.987	1.000	0.977	1.000	1.000	1.000
9	Q_n	0.324	0.503	0.848	0.902	0.825	0.892	0.849	0.825
	Q_n^S	0.407	0.647	0.976	1.000	0.952	0.999	1.000	1.000

Overall, we find that the split cancellation condition (4.13) in Table 1is strongly supported for the ATM(1) model. In small and moderate samples, the McLeod-type statistic Q_n, which explicitly estimates the residual-ACF covariance matrix, can exhibit noticeable finite-sample size distortion. The sample-splitting statistic Q_n^S avoids this covariance-matrix inversion and shows improving size as n increases. The ATM(2) alternative experiments further suggest that the tests exhibit non-trivial power against omitted higher-order transport dependence.

6. Illustrative Data Analysis

In this section, we analyze the distributional data of summer temperatures from three major U.S. airports: Ted Stevens Anchorage International Airport (Alaska), San Francisco International Airport (California), and John F. Kennedy International Airport (New York). These locations represent distinct climate regimes—subarctic, cool-summer Mediterranean, and humid subtropical—providing a diverse set of distributional dynamics for evaluating the ATM model. This dataset has been analyzed in both Zhu and Müller (2023) and Ghodrati and Panaretos (2024). For reproducibility, we clarify that the raw daily minimum temperature records are obtained from the Global Historical Climatology Network Daily through the NOAA Climate Data Online portal, and annual empirical distributions are formed by pooling observations from June 1 to September 30 within each year; see https://www.ncei.noaa.gov/cdo-web/search?datasetid=GHCND.

For each airport, the dataset contains annual distributions of daily minimum temperatures from June 1 to September 30 for the years 1960–2018; see Figure 1 for a visual illustration. Thus, each airport series contains $n = 59$ annual distributions. We see that all three distribution series tend to shift toward higher temperatures over time, with earlier-year curves appearing mostly on the left and more recent curves progressively moving to the right. This systematic "rightward shift" can be naturally captured and quantified through the Wasserstein distance and the ATM modeling framework. Because the annual distributions exhibit a certain long-term shift and the sample size is modest, a deterministic trend or other nonstationary component may induce residual dependence and therefore affect the finite-sample performance of the portmanteau statistics. Nonetheless, the following analysis can still be interpreted as an illustrative diagnostic exercise based on a stationary ATM(1)

working approximation.

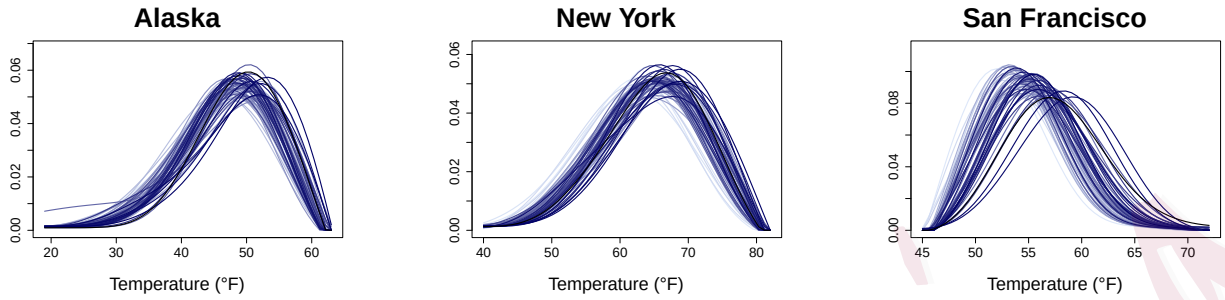


Figure 1: Time series of distributions of daily minimum temperatures during summer (June–September) for 1960–2018 at three U.S. airports: Ted Stevens Anchorage (Alaska), San Francisco (California), and John F. Kennedy (New York). The color gradient of the distribution curves encodes the temporal order, with lighter shades denoting earlier years.

Table 4 reports the fitted ATM(1) parameters and the diagnostic p-values obtained from both the McLeod-type and sample-splitting tests. From the table, we see that all three parameter estimates are positive, indicating positive first-order persistence in the transport-map dynamics, with New York exhibiting the strongest persistence. The McLeod-type test does not reject the ATM(1) specification for any of the three locations at the 5% significance level. The sample-splitting test also provides only weak evidence against the model. For Alaska, the sample-splitting p-values are 0.055, 0.045, and 0.122 for $K = 3, 6, 9$, respectively. The two tests therefore give different decisions only at $K = 6$, where the McLeod-type p-value is 0.322. Although the sample-splitting test rejects at this lag, both tests fail to reject at $K = 3, 9$. With only $n = 59$ annual observations, we interpret the evidence against ATM(1) as limited and sensitive to the test and lag choice. In contrast, the ATM(1) model is not rejected for New York or San Francisco at any of the reported lag orders. Overall, given the relatively short annual samples, the results suggest that the ATM(1) model provides a more adequate fit for New York and San Francisco.

Table 4: Estimated ATM(1) contraction parameters using full samples, and p-values for the McLeod type and sample-splitting portmanteau tests.

		McLeod Type			**Sample-Splitting Type**		
Location	$\widehat{\alpha}_n$	$K=3$	$K=6$	$K=9$	$K=3$	$K=6$	$K=9$
Alaska	0.2008	0.143	0.322	0.540	0.055	0.045	0.122
New York	0.4426	0.737	0.817	0.960	0.375	0.707	0.915
San Francisco	0.3411	0.486	0.794	0.272	0.412	0.690	0.280

Additional prediction curves for 2018 and rolling Wasserstein distances for 2010–2018 are reported in Section S.5 of the Supplement. They give the same qualitative message as the residual diagnostics: Alaska is the most difficult of the three series to capture with a first-order ATM model, whereas New York and San Francisco are relatively closer to the fitted ATM(1) specification.

7. Conclusion

In this paper, we propose two portmanteau diagnostic tests for assessing the adequacy of fitted ATM(1) models. Both tests are based on residual ACFs of the fitted transport maps. The first test is a McLeod-type statistic that corrects for the effect of parameter estimation through the limiting covariance matrix, while the second uses sample splitting to obtain a simpler chi-square calibration. For both procedures, we establish the corresponding limiting distributions under the correctly specified ATM(1) model.

For empirical use, we suggest the sample-splitting test with $f_n = n/2$ and $l_n = n$ when condition (4.13) is a reasonable working assumption under moderate sample size. It avoids estimating and inverting the full residual-ACF covariance matrix, and its empirical size improves with sample size in our simulations. The McLeod-type test is an alternative when one does not wish to rely on (4.13), provided that the covariance estimate can be reliably

computed.

Several directions remain open. First, it would be useful to incorporate an "intercept" term, as in Ghodrati and Panaretos (2024), and develop the corresponding diagnostic theory. Second, extending the tests to partially observed distributions or estimated transport maps requires accounting for map-estimation error. Third, a natural direction is to develop diagnostic tools for higher-order or higher-dimensional transport autoregressive models. Our analysis suggests that the simple extension of the present ATM(1) sample-splitting argument is not sufficient for such models. More general settings may therefore require new definitions of residual autocorrelation, or even different model and geometric structures. Fourth, given that neither test is uniformly preferable in small samples, developing bootstrap-based tests with improved finite-sample performance is a useful direction for future research. Finally, portmanteau-type diagnostics for other non-Euclidean time series, such as covariance matrices, networks, or other object-valued data, remain an interesting direction.

Acknowledgments

We would like to thank Prof. Xiaofeng Shao for his contributions to an earlier version of this work. We are also grateful to the editor, associate editor, and two referees for their insightful comments and suggestions, which have substantially improved the presentation and quality of this paper. Jiang is supported in part by the National Key R&D Program of China (No. 2024YFA1015700), and by the NSFC (Nos. 72522009, 12331009 and 72432002). Li is supported by the NSFC (No.72471127).

Statistica Sinica: Newly accepted Paper

References

Ansley, C. F. and Newbold, P. (1979). On the finite sample distribution of residual autocorrelations in autoregressive-moving average models. *Biometrika*, 66(3):547–553.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526.

Chang, J., Cheng, G., and Yao, Q. (2022). Testing for unit roots based on sample autocovariances. *Biometrika*, 109(2):543–550.

Chen, Y., Lin, Z., and Müller, H.-G. (2023). Wasserstein regression. *Journal of the American Statistical Association*, 118(542):869–882.

Davis, R. A. and Fernandes, L. (2025). Sample splitting and assessing goodness-of-fit of time series. *Biometrika*, 112(2):Paper No. asaf017, 21.

Dubey, P. and Müller, H.-G. (2020). Fréchet change-point detection. *The Annals of Statistics*, 48(6):3312–3335.

Gao, H., Wang, R., and Shao, X. (2025). Dimension-agnostic change point detection. *Journal of Econometrics*, 250:106012.

Ghodrati, L. and Panaretos, V. M. (2024). On distributional autoregression and iterated transportation. *Journal of Time Series Analysis*, 45(5):739–770.

Hall, P. and Heyde, C. C. (2014). *Martingale Limit Theory and Its Application*. Academic press.

Statistica Sinica: Newly accepted Paper

Jiang, F., Gao, H., and Shao, X. (2024a). Testing serial independence of object-valued time series. *Biometrika*, 111(3):925–944.

Jiang, F., Zhu, C., and Shao, X. (2024b). Two-sample and change-point inference for non-euclidean valued time series. *Electronic Journal of Statistics*, 18(1):848–894.

Jiang, Y. (2022). Wasserstein multivariate auto-regressive models for modeling distributional time series and its application in graph learning. *arXiv:2207.05442*.

Li, W. K. (1992). On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling. *Biometrika*, 79(2):435–437.

Li, W. K. (2003). *Diagnostic Checks in Time Series*. Chapman & Hall/CRC.

Li, W. K. and Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15(6):627–636.

Li, W. K. and McLeod, A. I. (1981). Distribution of the residual autocorrelations in multivariate arma time series models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 43(2):231–239.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Lunde, R. (2019). Sample splitting and weak assumption inference for time series. *https://arxiv.org/abs/1902.07425*.

Mazzuco, S. and Scarpa, B. (2015). Fitting age-specific fertility rates by a flexible generalized skew normal probability density function. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 178(1):187–203.

Statistica Sinica: Newly accepted Paper

McLeod, A. (1978). On the distribution of residual autocorrelations in box–jenkins models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3):296–302.

Newbold, P. (1980). The equivalence of two tests of time series model adequacy. *Biometrika*, 67(2):463–465.

Shang, H. L. and Hyndman, R. J. (2017). Grouped functional time series forecasting: an application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2):330–343.

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436.

Zhang, C., Kokoszka, P., and Petersen, A. (2022). Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 43(1):30–52.

Zhang, Y. and Shao, X. (2024). Another look at bandwidth-free inference: A sample splitting approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):246–272.

Zhang, Y., Zhu, C., and Shao, X. (2026). Change-point detection for object-valued time series. *Journal of Business & Economic Statistics*, 44(1):255–269.

Zhu, C. and Müller, H.-G. (2023). Autoregressive optimal transport models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):1012–1033.

Zhu, C. and Müller, H.-G. (2024). Spherical autoregressive models, with application to distributional and compositional time series. *Journal of Econometrics*, 239:105389.

Statistica Sinica: Newly accepted Paper

Chenxiao Dai

Yau Mathematical Sciences Center, Tsinghua University

E-mail: daicx22@mails.tsinghua.edu.cn

Feiyu Jiang

Department of Statistics and Data Science, School of Management, Fudan University

E-mail: jiangfy@fudan.edu.cn

Dong Li

Department of Statistics and Data Science, Tsinghua University

E-mail: malidong@tsinghua.edu.cn

Statistica Sinica