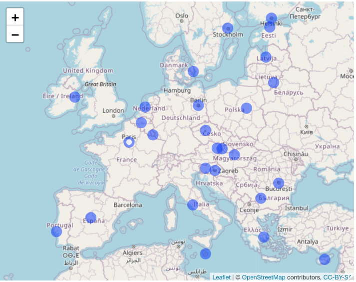
Wasserstein multivariate auto-regressive models for modeling distributional time series and its application in graph learning

Yiye JIANG

Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France
yiye.jiang@inria.fr

Multivariate distributional time series



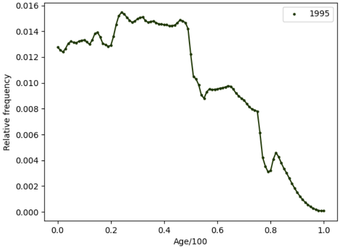


Figure 1: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).*

Multivariate distributional time series

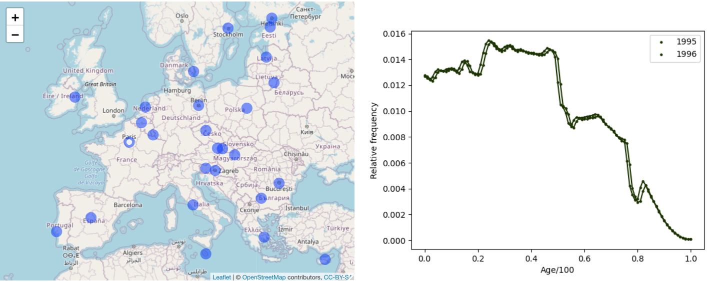


Figure 2: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).*

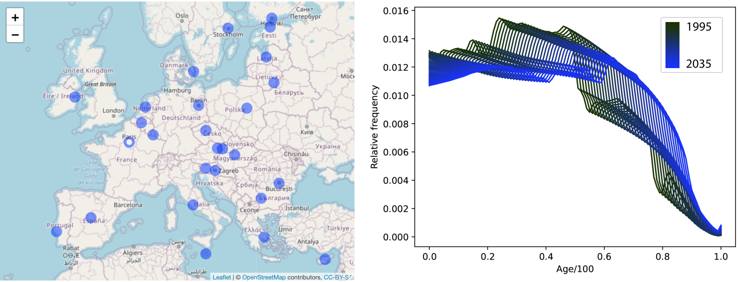
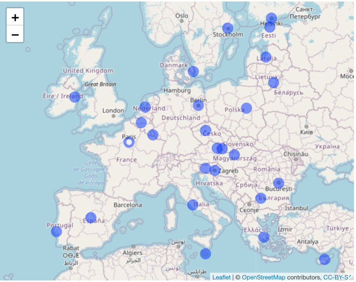


Figure 3: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).* On the right are the observations $(\boldsymbol{\mu}_{it})_t \in \mathcal{P}([0,1])$ along time recorded at $i =$ France. Lighter curves correspond to more recent years.



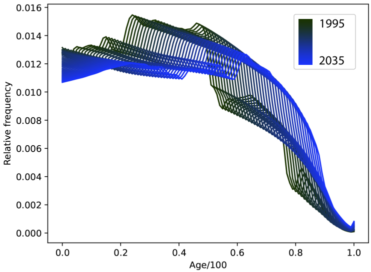
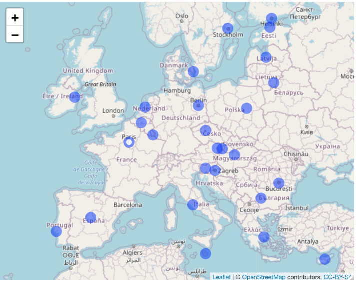


Figure 3: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).* On the right are the observations $(\boldsymbol{\mu}_{it})_t \in \mathcal{P}([0,1])$ along time recorded at i = France. Lighter curves correspond to more recent years.

Objectives:

1. A model to describe the new time series type:

$$\boldsymbol{\mu}_{it} \in \mathcal{P}(\mathbb{R}), i = 1...N, t \in \mathbb{Z}.$$



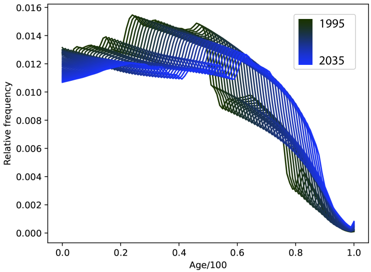


Figure 3: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).* On the right are the observations $(\boldsymbol{\mu}_{it})_t \in \mathcal{P}([0,1])$ along time recorded at i = France. Lighter curves correspond to more recent years.

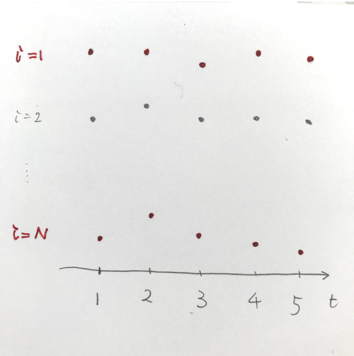
Objectives:

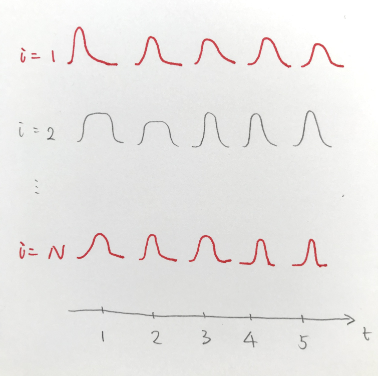
1. A model to describe the new time series type:

$$\boldsymbol{\mu}_{it} \in \mathcal{P}(\mathbb{R}), i = 1...N, t \in \mathbb{Z}.$$

2. Represent the series dependencies by a graph.

Vector auto-regressive model (VAR)





N scalar TS (data for VAR) N distributional TS (our data)

Vector auto-regressive model

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in \mathbb{R}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series.

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in \mathbb{R}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series.
Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant.

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in \mathbb{R}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise,

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in \mathbb{R}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j)$ defines the regressive dependencies.

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in \mathbb{R}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

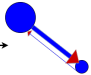
$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j)$ defines the regressive dependencies.

$A \rightarrow \mathcal{G}$

1.8	0.05
1.2	-0.4

$\longrightarrow$



Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in \mathbb{R}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

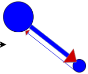
$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j)$ defines the regressive dependencies.

$A \to \mathcal{G}$

1.8	0.05
1.2	-0.4

$\longrightarrow$



$$Extension: \quad \boldsymbol{x}_{it} \in \mathbb{R} \longrightarrow \boldsymbol{\mu}_{it} \in \mathcal{W}_2(\mathbb{R}).$$

Backgrounds on statistics in $W_2(\mathbb{R})$

$$\mathcal{W}_2(\mathbb{R}) = \left\{ \mu \in \mathcal{P}(\mathbb{R}) \middle| \int_{\mathbb{R}} x^2 d\mu(x) < \infty \right\},$$

endowed with the 2-Wasserstein distance

Backgrounds on statistics in $W_2(\mathbb{R})$

$$\mathcal{W}_2(\mathbb{R}) = \left\{ \mu \in \mathcal{P}(\mathbb{R}) \middle| \int_{\mathbb{R}} x^2 d\mu(x) < \infty \right\},$$

endowed with the 2-Wasserstein distance

$$d_W(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} (x_1 - x_2)^2 d\pi(x_1, x_2)$$

Backgrounds on statistics in $W_2(\mathbb{R})$

$$\mathcal{W}_2(\mathbb{R}) = \left\{ \mu \in \mathcal{P}(\mathbb{R}) \middle| \int_{\mathbb{R}} x^2 d\mu(x) < \infty \right\},$$

endowed with the 2-Wasserstein distance

$$\begin{aligned} d_W(\mu, \nu) &= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} (x_1 - x_2)^2 d\pi(x_1, x_2) \\ &\overset{\|\cdot\|^2 + \mathbb{R}}{=} \int_0^1 \left(F_\mu^{-1}(u) - F_\nu^{-1}(u) \right)^2 du, \end{aligned}$$

where $F_\mu^{-1}(u), F_\nu^{-1}(u)$ are the quantile functions of μ and ν.

Backgrounds on statistics in $W_2(\mathbb{R})$

$$\mathcal{W}_2(\mathbb{R}) = \left\{ \mu \in \mathcal{P}(\mathbb{R}) \middle| \int_{\mathbb{R}} x^2 d\mu(x) < \infty \right\},$$

endowed with the 2-Wasserstein distance

$$\begin{aligned} d_W(\mu,\nu) &= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} (x_1 - x_2)^2 d\pi(x_1,x_2) \\ &\overset{\|\cdot\|^2+\mathbb{R}}{=} \int_0^1 \left(F_\mu^{-1}(u) - F_\nu^{-1}(u)\right)^2 du, \end{aligned}$$

where $F_\mu^{-1}(u), F_\nu^{-1}(u)$ are the quantile functions of μ and ν.

Extension of VAR models: $\mathcal{W}_2 := \mathcal{W}_2(\mathbb{R})$ is not linear.

Enable again linear methods - Tangent space

Ambrosio et al. (2008); Bigot et al. (2017); Zemel and Panaretos (2019) generalized basic concepts of Riemannian manifold to $\mathcal{W}_2$, e.g. Tangent space.

Enable again linear methods - Tangent space

Ambrosio et al. (2008); Bigot et al. (2017); Zemel and Panaretos (2019) generalized basic concepts of Riemannian manifold to $\mathcal{W}_2$, e.g. Tangent space.

Let $\gamma \in \mathcal{W}_2$ be an atomless measure (it possesses a continuous cdf F_γ), the tangent space at γ is defined as

$$\mathrm{Tan}_\gamma = \overline{\{t(T_\gamma^\mu - i) : \mu \in \mathcal{W}_2,\ t > 0\}}^{\mathcal{L}_\gamma^2},$$

where $T_\gamma^\mu = F_\mu^{-1} \circ F_\gamma$ is the optimal transport map, that pushes γ forward to μ.

Enable again linear methods - Tangent space

Ambrosio et al. (2008); Bigot et al. (2017); Zemel and Panaretos (2019) generalized basic concepts of Riemannian manifold to $\mathcal{W}_2$, e.g. Tangent space.

Let $\gamma \in \mathcal{W}_2$ be an atomless measure (it possesses a continuous cdf F_γ), the tangent space at γ is defined as

$$\mathrm{Tan}_\gamma = \overline{\{t(T_\gamma^\mu - i) : \mu \in \mathcal{W}_2,\ t > 0\}}^{\mathcal{L}_\gamma^2},$$

where $T_\gamma^\mu = F_\mu^{-1} \circ F_\gamma$ is the optimal transport map, that pushes γ forward to μ. Tan_γ is endowed with the inner product $\langle \cdot, \cdot \rangle_\gamma$ defined by

$$\langle f, g \rangle_\gamma := \int_{\mathbb{R}} f(x) g(x)\, d\gamma(x),\ f, g \in \mathcal{L}_\gamma^2(\mathbb{R}),$$

and the induced norm $\| \cdot \|_\gamma$.

Enable again linear methods - Tangent space

$$\mathrm{Tan}_\gamma = \overline{\{t(T_\gamma^\mu - i) : \mu \in \mathcal{W}_2,\ t > 0\}}^{\mathcal{L}_\gamma^2},$$

where $T_\gamma^\mu = F_\mu^{-1} \circ F_\gamma$ is the optimal map, that pushes γ forward to μ.

Definition

The logarithmic map $\mathrm{Log}_\gamma : \mathcal{W}_2 \to \mathrm{Tan}_\gamma$ is defined as

$$\mathrm{Log}_\gamma \mu = T_\gamma^\mu - i.$$

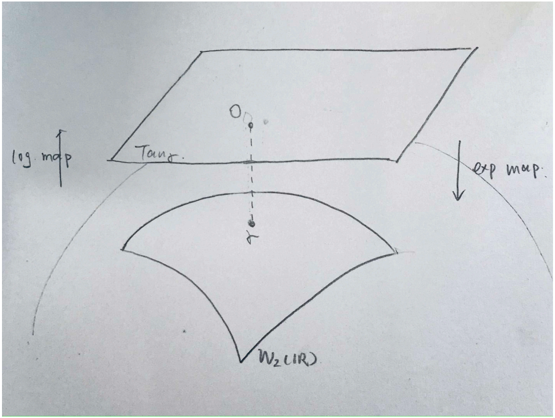
The exponential map $\mathrm{Exp}_\gamma : \mathrm{Tan}_\gamma \to \mathcal{W}_2$ is defined as

$$\mathrm{Exp}_\gamma g = (g + id)\#\gamma,$$

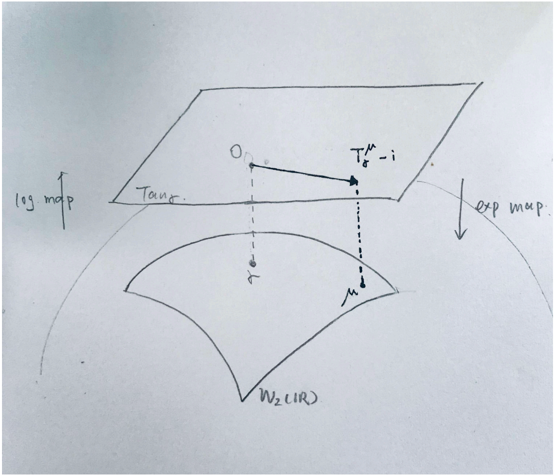
where $T\#\mu$ is the measure pushforwarded by function T, defined as $[T\#\mu](A) = \mu(\{x : T(x) \in A\})$.

Tangent space and Geodesic

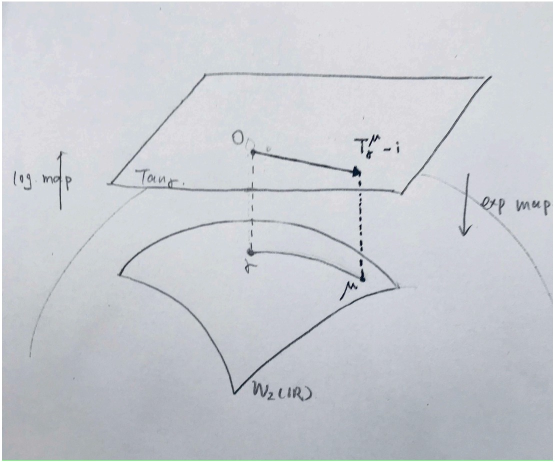
Tangent space and Geodesic



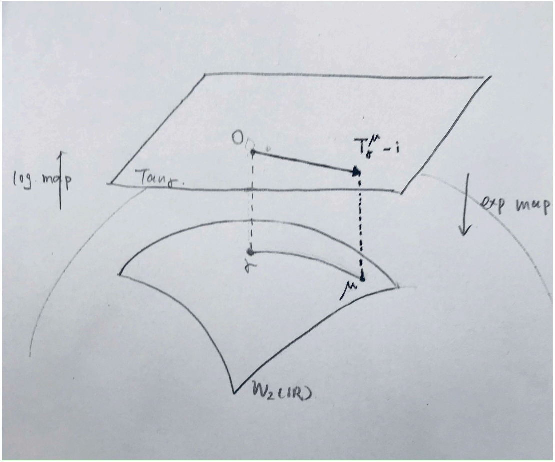
Tangent space and Geodesic



A geodesic in W is defined by a geodesic in Tan



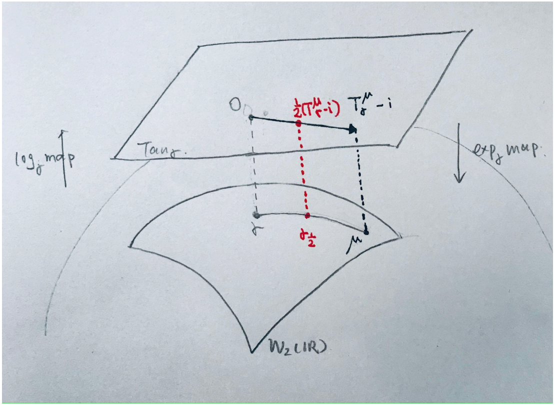
A geodesic in W is defined by a geodesic in Tan



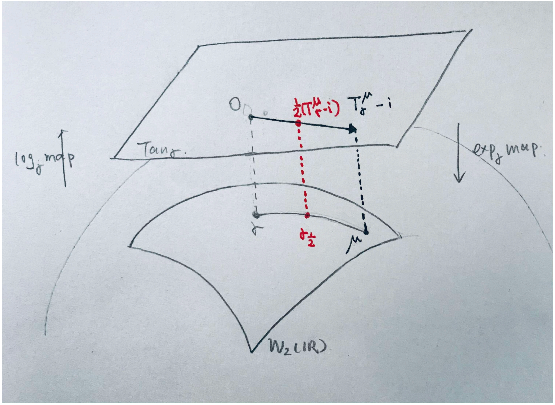
The geodesic (McCann's interpolant) between γ and μ

$$\mathrm{Exp}_\gamma[\alpha(T_\gamma^\mu - i)],\ \alpha : 0 \to 1,$$

Constant-speed Geodesic

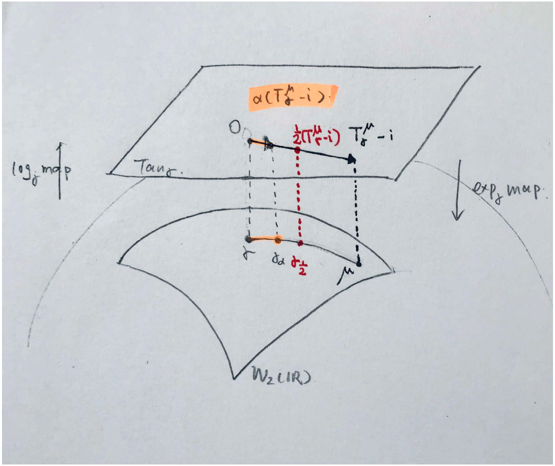


Constant-speed Geodesic

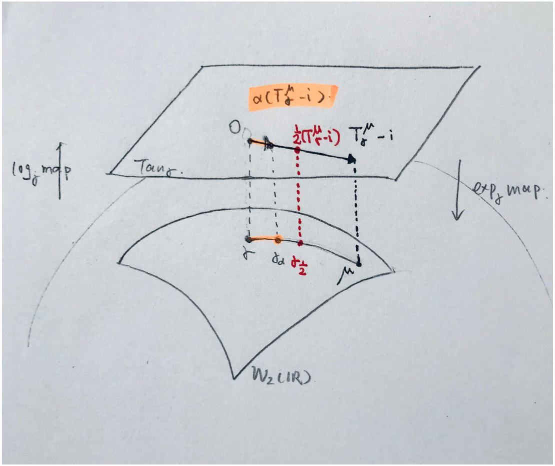


$$d_W(\gamma, \gamma_{\frac{1}{2}}) = \frac{1}{2} d_W(\gamma, \mu)$$

Constant-speed Geodesic



Constant-speed Geodesic



$$d_W(\gamma, \gamma_\alpha) = \alpha d_W(\gamma, \mu)$$

Related work: Univariate Wasserstein AR model

Vector AR models with $N = 1$:

$$\boldsymbol{x}_t - u = \alpha(\boldsymbol{x}_{t-1} - u) + \boldsymbol{\epsilon}_t,$$

Related work: Univariate Wasserstein AR model

Vector AR models with $N = 1$:

$$\boldsymbol{x}_t - u = \alpha(\boldsymbol{x}_{t-1} - u) + \boldsymbol{\epsilon}_t,$$

$$\overset{\text{ignore the noise}}{\Longleftrightarrow} \quad \mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u + \alpha(\boldsymbol{x}_{t-1} - u).$$

Related work: Univariate Wasserstein AR model

Vector AR models with $N = 1$:

$$\boldsymbol{x}_t - u = \alpha(\boldsymbol{x}_{t-1} - u) + \boldsymbol{\epsilon}_t,$$

$$\overset{\text{ignore the noise}}{\Longleftrightarrow} \quad \mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u + \alpha(\boldsymbol{x}_{t-1} - u).$$

Equation above defines the **regressive dependency**, in other words, the **prediction of x_t given x_{t-1}**.

Related work: Univariate Wasserstein AR model

Vector AR models with $N = 1$:

$$\boldsymbol{x}_t - u = \alpha(\boldsymbol{x}_{t-1} - u) + \boldsymbol{\epsilon}_t,$$

$$\overset{\text{ignore the noise}}{\iff} \quad \mathbb{E}\boldsymbol{x}_t | \boldsymbol{x}_{t-1} = u + \alpha(\boldsymbol{x}_{t-1} - u).$$

Equation above defines the **regressive dependency**, in other words, the **prediction of** x_t **given** x_{t-1}.

Chen et al. (2021); Zhang et al. (2021); Zhu and Müller (2021) extended the univariate AR model by **interpreting the regressive dependency from a geometric point of view.**

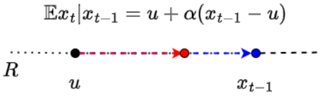


Figure 4: *Geometric interpretation of regressive dependency of AR models.*

$$\mathbb{E}x_t|x_{t-1} = u + \alpha(x_{t-1} - u)$$

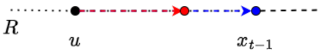
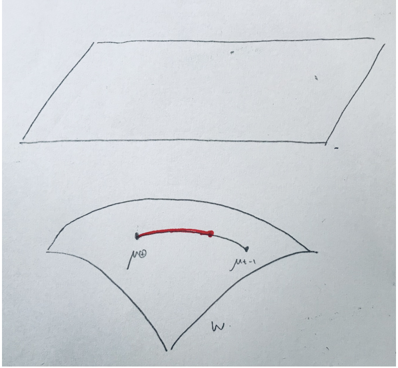
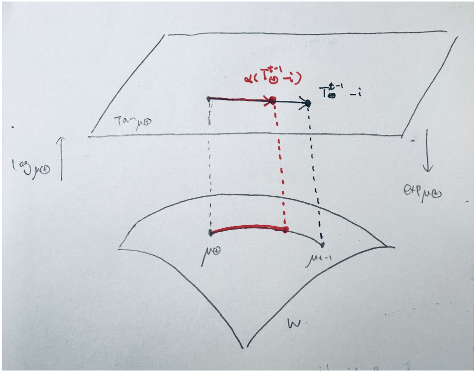


Figure 4: *Geometric interpretation of regressive dependency of AR models.*

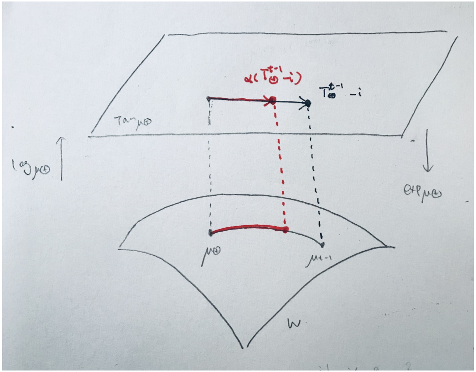


$$\mathbb{E}x_t|x_{t-1} = u + \alpha(x_{t-1} - u)$$

Figure 5: *Geometric interpretation of regressive dependency of AR models.*



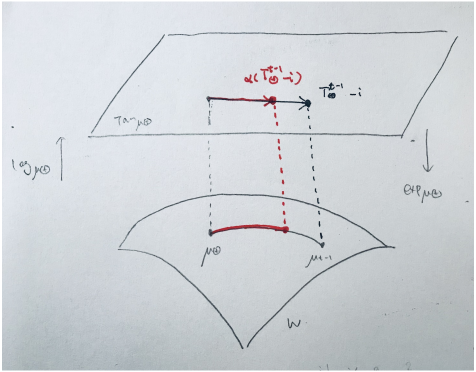
$$\mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u + \alpha(\boldsymbol{x}_{t-1} - u) \Longrightarrow \mathbb{E}_{\oplus}\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1} = \mathrm{Exp}_{\mu_\oplus}\left\{\alpha(\boldsymbol{T}_{\oplus}^{t-1} - i)\right\}$$



$$\mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u+\alpha(\boldsymbol{x}_{t-1}-u) \Longrightarrow \mathbb{E}_{\oplus}\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1} = \mathrm{Exp}_{\mu_{\oplus}}\left\{\alpha(\boldsymbol{T}_{\oplus}^{t-1} - i)\right\}$$

Let $\boldsymbol{\mu}, \boldsymbol{\gamma}$ be two random measures from $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathcal{W}_2$,

$$\mathbb{E}_{\oplus}\boldsymbol{\mu} = \underset{\nu\in\mathcal{W}_2}{\arg\min}\,\mathbb{E}\left[d_W^2(\boldsymbol{\mu},\nu)\right], \; \mathbb{E}_{\oplus}\boldsymbol{\mu}|\boldsymbol{\gamma} := \underset{\nu\in\mathcal{W}_2}{\arg\min}\,\mathbb{E}\left[d_W^2(\boldsymbol{\mu},\nu)|\boldsymbol{\gamma}\right]$$



Multivariate Wasserstein AR model

Extension of the univariate AR models:

$$\mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u+\alpha(\boldsymbol{x}_{t-1}-u) \Longrightarrow \mathbb{E}_{\oplus}\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1} = \mathrm{Exp}_{\mu_\oplus}\left(\alpha(\boldsymbol{T}_{\oplus}^{t-1} - i)\right)$$

Multivariate AR models:

$$\mathbb{E}\boldsymbol{x}_{it}|\boldsymbol{x}_{j,t-1} = u_i+\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1}-u_j)$$

Multivariate Wasserstein AR model

Extension of the univariate AR models:

$$\mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u+\alpha(\boldsymbol{x}_{t-1}-u) \Longrightarrow \mathbb{E}_{\oplus}\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1} = \mathrm{Exp}_{\mu_{\oplus}}\left(\alpha(\boldsymbol{T}_{\oplus}^{t-1} - i)\right)$$

Multivariate AR models:

$$\mathbb{E}\boldsymbol{x}_{it}|\boldsymbol{x}_{j,t-1} = u_i+\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1}-u_j) \overset{\times!}{\Rightarrow} \begin{cases} \boldsymbol{T}_{1,\oplus}^{1,t-1} - i & \in \mathrm{Tan}_{\mu_{1,\oplus}} \\ \boldsymbol{T}_{2,\oplus}^{2,t-1} - i & \in \mathrm{Tan}_{\mu_{2,\oplus}} \\ & \vdots \end{cases}$$

Multivariate Wasserstein AR model

Extension of the univariate AR models:

$$\mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u+\alpha(\boldsymbol{x}_{t-1}-u) \Longrightarrow \mathbb{E}_{\oplus}\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1} = \mathrm{Exp}_{\mu_{\oplus}}\left(\alpha(\boldsymbol{T}_{\oplus}^{t-1} - i)\right)$$

Multivariate AR models:

$$\mathbb{E}\boldsymbol{x}_{it}|\boldsymbol{x}_{j,t-1} = u_i+\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1}-u_j) \overset{\times !}{\Rightarrow} \begin{cases} \boldsymbol{T}_{1,\oplus}^{1,t-1} - i & \in \mathrm{Tan}_{\mu_{1,\oplus}} \\ \boldsymbol{T}_{2,\oplus}^{2,t-1} - i & \in \mathrm{Tan}_{\mu_{2,\oplus}} \\ & \vdots \end{cases}$$

Center: $\boldsymbol{\mu}_{it}, \mathbb{E}_{\oplus}\boldsymbol{\mu}_{it} = u_{i,\oplus} \longrightarrow \tilde{\boldsymbol{\mu}}_{it}, \mathbb{E}_{\oplus}\boldsymbol{\mu}_{it} \equiv c.$

$$\Longrightarrow \mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}}_{it}|\tilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_c\left(\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{T}}_c^{j,t-1} - i)\right)$$

Multivariate Wasserstein AR model

Extension of the univariate AR models:

$$\mathbb{E}\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = u+\alpha(\boldsymbol{x}_{t-1}-u) \Longrightarrow \mathbb{E}_{\oplus}\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1} = \mathrm{Exp}_{\mu_{\oplus}}\left(\alpha(\boldsymbol{T}_{\oplus}^{t-1} - i)\right)$$

Multivariate AR models:

$$\mathbb{E}\boldsymbol{x}_{it}|\boldsymbol{x}_{j,t-1} = u_i+\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1}-u_j) \overset{\times!}{\Rightarrow} \begin{cases} \boldsymbol{T}_{1,\oplus}^{1,t-1} - i & \in \mathrm{Tan}_{\mu_{1,\oplus}} \\ \boldsymbol{T}_{2,\oplus}^{2,t-1} - i & \in \mathrm{Tan}_{\mu_{2,\oplus}} \\ & \vdots \end{cases}$$

Center: $\boldsymbol{\mu}_{it}, \mathbb{E}_{\oplus}\boldsymbol{\mu}_{it} = u_{i,\oplus} \longrightarrow \tilde{\boldsymbol{\mu}}_{it}, \mathbb{E}_{\oplus}\boldsymbol{\mu}_{it} \equiv c.$

$$\Longrightarrow \mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}}_{it}|\tilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_c\left(\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{T}}_c^{j,t-1} - i)\right)$$

We proposed a centering for random measures so that the centered measures always have $U[0,1]$ as population Fréchet mean.

Multivariate Wasserstein AR model

For a multivariate distributional time series $\boldsymbol{\mu}_{it}$, $i = 1, \ldots, N, t \in \mathbb{Z}$, we define the regressive dependency on their centered versions $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ as

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i)\right),$$

where $\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i$ is the tangent vector of $\widetilde{\boldsymbol{\mu}}_{j,t-1}$.

Multivariate Wasserstein AR model

For a multivariate distributional time series $\boldsymbol{\mu}_{it}$, $i = 1, \ldots, N, t \in \mathbb{Z}$, we define the regressive dependency on their centered versions $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ as

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i)\right),$$

where $\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i$ is the tangent vector of $\widetilde{\boldsymbol{\mu}}_{j,t-1}$.

Objectives obtained:

1. A new model for the new time series type:

$$\boldsymbol{\mu}_{it} \in \mathcal{P}(\mathbb{R}), i = 1...N, t \in \mathbb{Z}.$$

Multivariate Wasserstein AR model

For a multivariate distributional time series $\boldsymbol{\mu}_{it}$, $i = 1, \ldots, N, t \in \mathbb{Z}$, we define the regressive dependency on their centered versions $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ as

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i)\right),$$

where $\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i$ is the tangent vector of $\widetilde{\boldsymbol{\mu}}_{j,t-1}$.

Objectives obtained:

1. A new model for the new time series type:
$$\boldsymbol{\mu}_{it} \in \mathcal{P}(\mathbb{R}), i = 1 \ldots N, t \in \mathbb{Z}.$$
2. Graph learning. $A \to \mathcal{G}$.

Multivariate Wasserstein AR model

For a multivariate distributional time series $\boldsymbol{\mu}_{it},\ i = 1, \ldots, N, t \in \mathbb{Z}$, we define the regressive dependency on their centered versions $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ as

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i)\right),$$

where $\widetilde{\boldsymbol{T}}_{Leb}^{j,t-1} - i$ is the tangent vector of $\widetilde{\boldsymbol{\mu}}_{j,t-1}$.

Objectives obtained:

1. A new model for the new time series type:
$$\boldsymbol{\mu}_{it} \in \mathcal{P}(\mathbb{R}), i = 1...N, t \in \mathbb{Z}.$$
2. Graph learning. $A \to \mathcal{G}$.
3. Theoretically, we proved the stationarity under assumptions.

Constrained least-square estimation

Given the observations $\boldsymbol{\mu}_{it}, t = 0, ..., T, i = 1, ..., N$ (thus the centered observations $\widetilde{\boldsymbol{\mu}}_{it}$), we propose the estimator of A as

$$\widetilde{\boldsymbol{A}} = \underset{\substack{A \text{ satisfies the} \\ \text{model assumptions}}}{\arg\min}{}^{1} \; \frac{1}{T} \sum_{t=1}^{T} d_{\mathcal{W}}^{2} \left(\widetilde{\boldsymbol{\mu}}_{it}, \; \mathbb{E}_{\oplus} \widetilde{\boldsymbol{\mu}}_{it} | \widetilde{\boldsymbol{\mu}}_{j,t-1} \right),$$

[1] The details see Jiang (2022).

Constrained least-square estimation

Given the observations $\boldsymbol{\mu}_{it}, t = 0, ..., T, i = 1, ..., N$ (thus the centered observations $\widetilde{\boldsymbol{\mu}}_{it}$), we propose the estimator of A as

$$\widetilde{\boldsymbol{A}} = \underset{\substack{A \text{ satisfies the} \\ \text{model assumptions}^{1}}}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} d^2_{\mathcal{W}} \left(\widetilde{\boldsymbol{\mu}}_{it}, \ \mathbb{E}_{\oplus} \widetilde{\boldsymbol{\mu}}_{it} | \widetilde{\boldsymbol{\mu}}_{j,t-1} \right),$$

$\widetilde{\boldsymbol{A}}$ is sparse and $0 \leqslant \widetilde{\boldsymbol{A}}_{ij} \leqslant 1$.

[1] The details see Jiang (2022).

Constrained least-square estimation

Given the observations $\boldsymbol{\mu}_{it}, t = 0, ..., T, i = 1, ..., N$ (thus the centered observations $\widetilde{\boldsymbol{\mu}}_{it}$), we propose the estimator of A as

$$\widetilde{\boldsymbol{A}} = \underset{\substack{A \text{ satisfies the} \\ \text{model assumptions}^{1}}}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} d_{\mathcal{W}}^2 \left(\widetilde{\boldsymbol{\mu}}_{it}, \ \mathbb{E}_{\oplus} \widetilde{\boldsymbol{\mu}}_{it} | \widetilde{\boldsymbol{\mu}}_{j,t-1} \right),$$

$\widetilde{\boldsymbol{A}}$ is sparse and $0 \leqslant \widetilde{\boldsymbol{A}}_{ij} \leqslant 1$.

We proved that $\widetilde{\boldsymbol{A}}$ is consistent.

[1] The details see Jiang (2022).

Age distributions of countries

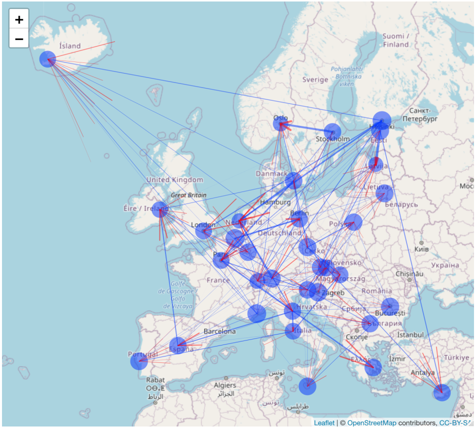


Figure 6: *Visualization of* $\widetilde{\boldsymbol{A}}$. $\widetilde{\boldsymbol{A}}_{ij}$ are represented by the weighted directed edges from node j to node i.

Age distributions of countries

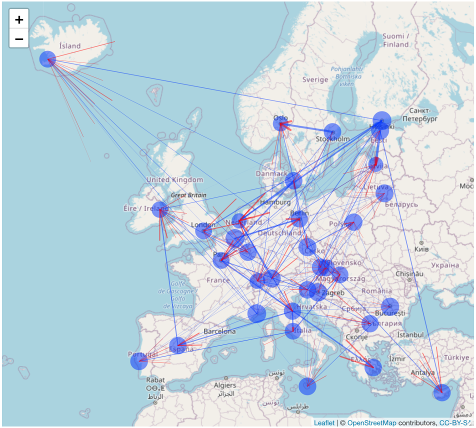


Figure 7: *Visualization of $\widetilde{\boldsymbol{A}}$.* $\widetilde{\boldsymbol{A}}_{ij}$ are represented by the weighted directed edges from node j to node i. Thicker edges correspond to larger values.

Age distributions of countries

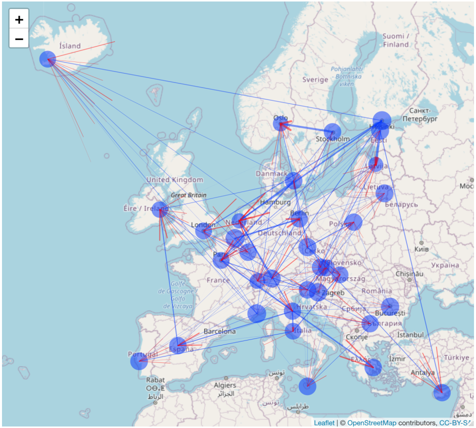


Figure 8: *Visualization of $\widetilde{\boldsymbol{A}}$.* $\widetilde{\boldsymbol{A}}_{ij}$ are represented by the weighted directed edges from node j to node i. Thicker edges correspond to larger values. The blue circles around nodes represent $\widetilde{\boldsymbol{A}}_{ii}$.

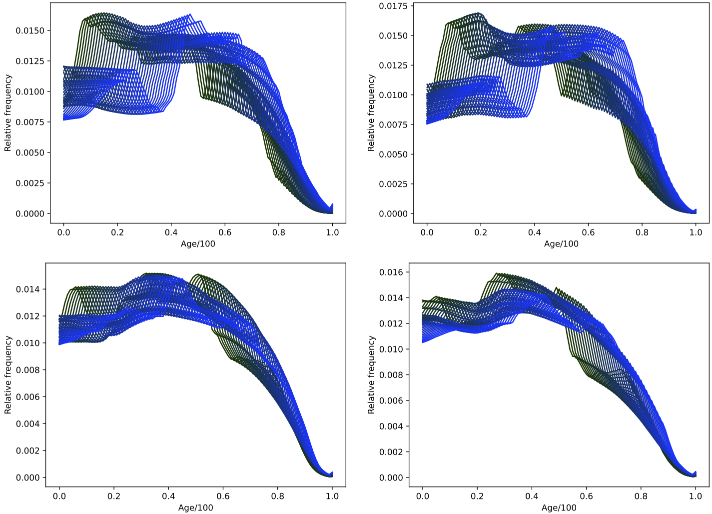


Figure 9: *Evolution of age structure from* 1996 *to* 2036 *(projected). Estonia (top left), Latvia(top right), Sweden (bottom left) versus Norway (bottom right).*

Age distributions of countries

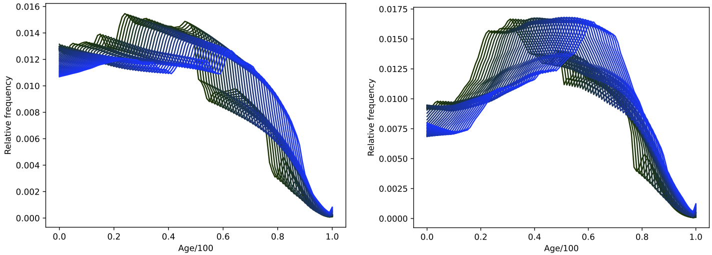


Figure 10: *Evolution of age structure from* 1996 *to* 2036 *(projected) of France (left) versus Italy (right).*

Thanks for your attention !

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

J. Bigot, R. Gouet, T. Klein, and A. López. Geodesic pca in the wasserstein space by convex pca. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):1–26, 2017.

Y. Chen, Z. Lin, and H.-G. Müller. Wasserstein regression. *Journal of the American Statistical Association*, pages 1–14, 2021.

Y. Jiang. Wasserstein multivariate auto-regressive models for modeling distributional time series and its application in graph learning. *stat*, 1050:12, 2022.

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

J. Thai, C. Wu, A. Pozdnukhov, and A. Bayen. Projected sub-gradient with ℓ_1 or simplex constraints via isotonic regression. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 2031–2036. IEEE, 2015.

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

Y. Zemel and V. M. Panaretos. Fréchet means and procrustes analysis in wasserstein space. 2019.

C. Zhang, P. Kokoszka, and A. Petersen. Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 2021.

C. Zhu and H.-G. Müller. Autoregressive optimal transport models. *arXiv preprint arXiv:2105.05439*, 2021.

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

$$\boldsymbol{\mu} \leftrightarrow \boldsymbol{F}^{-1}$$

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

$$\boldsymbol{\mu} \leftrightarrow \boldsymbol{F}^{-1}$$
$$\mu_{\oplus} \leftrightarrow F_{\oplus}^{-1}$$

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

$$\left.\begin{aligned} \boldsymbol{\mu} &\leftrightarrow \boldsymbol{F}^{-1} \\ \mu_{\oplus} &\leftrightarrow F_{\oplus}^{-1} \end{aligned}\right. \quad \Rightarrow F_{\oplus}^{-1} \overset{elem.wise}{=} \mathbb{E}\boldsymbol{F}^{-1}.$$

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

$$\begin{aligned} \boldsymbol{\mu} &\leftrightarrow \boldsymbol{F}^{-1} \\ \mu_{\oplus} &\leftrightarrow F_{\oplus}^{-1} \end{aligned} \quad \Rightarrow F_{\oplus}^{-1} \overset{elem.wise}{=} \mathbb{E}\boldsymbol{F}^{-1}.$$

After certain centralization

$$\tilde{\boldsymbol{\mu}} \leftrightarrow \tilde{\boldsymbol{F}}^{-1}$$

$$\tilde{\mu}_{\oplus} \leftrightarrow id = \mathbb{E}\tilde{\boldsymbol{F}}^{-1}$$

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

$$\begin{aligned} \boldsymbol{\mu} &\leftrightarrow \boldsymbol{F}^{-1} \\ \mu_{\oplus} &\leftrightarrow F_{\oplus}^{-1} \end{aligned} \quad \Rightarrow F_{\oplus}^{-1} \overset{elem.wise}{=} \mathbb{E}\boldsymbol{F}^{-1}.$$

After certain centralization

$$\tilde{\boldsymbol{\mu}} \leftrightarrow \tilde{\boldsymbol{F}}^{-1} := \boldsymbol{F}^{-1} \circ [F_{\oplus}^{-1}]^{-1}$$

$$\tilde{\mu}_{\oplus} \leftrightarrow id = \mathbb{E}\tilde{\boldsymbol{F}}^{-1} = \mathbb{E}\boldsymbol{F}^{-1} \circ [F_{\oplus}^{-1}]^{-1}$$

Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$

$$\begin{matrix} \boldsymbol{\mu} \leftrightarrow \boldsymbol{F}^{-1} \\ \mu_{\oplus} \leftrightarrow F_{\oplus}^{-1} \end{matrix} \quad \Rightarrow F_{\oplus}^{-1} \overset{elem.wise}{=} \mathbb{E}\boldsymbol{F}^{-1}.$$

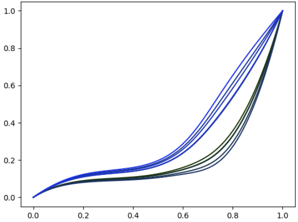
After certain centralization

$$\tilde{\boldsymbol{\mu}} \leftrightarrow \tilde{\boldsymbol{F}}^{-1} := \boldsymbol{F}^{-1} \circ [F_{\oplus}^{-1}]^{-1}$$
$$\tilde{\mu}_{\oplus} \leftrightarrow id = \mathbb{E}\tilde{\boldsymbol{F}}^{-1} = \mathbb{E}\boldsymbol{F}^{-1} \circ [F_{\oplus}^{-1}]^{-1}$$

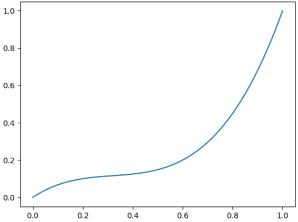
Assumption

$\boldsymbol{\mu}$ is supported on $[0,1]$.

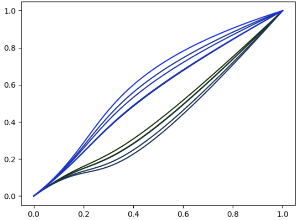
Center a random measure $\boldsymbol{\mu}$, s.t. $\mathbb{E}_{\oplus}\tilde{\boldsymbol{\mu}} = U(0,1)$



$\ominus$



$=$



Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

1. identifiable: $\begin{cases} A \rightarrow \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \\ A' \rightarrow \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \end{cases}$

Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

1. identifiable: $\begin{cases} A \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \\ A' \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \end{cases}$ $\Longleftarrow$ Exp is not injective.

Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

1. identifiable: $\begin{cases} A \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \\ A' \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \end{cases}$ $\Longleftarrow$ Exp is not injective.

$\mathrm{Exp}_{\gamma}|_{\mathrm{Log}_{\gamma}\mathcal{W}}$ is an isometric homeomorphism from $\mathrm{Log}_{\gamma}\mathcal{W}_2$ to $\mathcal{W}_2$, with the inverse map Log_{γ}(Bigot et al., 2017).

Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

1. identifiable: $\begin{cases} A \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \\ A' \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \end{cases} \Longleftarrow$ Exp is not injective.

$\mathrm{Exp}_\gamma|_{\mathrm{Log}_\gamma \mathcal{W}}$ is an isometric homeomorphism from $\mathrm{Log}_\gamma \mathcal{W}_2$ to $\mathcal{W}_2$, with the inverse map Log_γ(Bigot et al., 2017).

2. A not tractable in estimation.

Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

1. identifiable: $\begin{cases} A \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \\ A' \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \end{cases}$ $\Longleftarrow$ Exp is not injective.

$\mathrm{Exp}_{\gamma}|_{\mathrm{Log}_{\gamma}\mathcal{W}}$ is an isometric homeomorphism from $\mathrm{Log}_{\gamma}\mathcal{W}_2$ to $\mathcal{W}_2$, with the inverse map Log_{γ}(Bigot et al., 2017).

2. A not tractable in estimation.

Thus, $1 + 2 \to$

$$\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i) \in \mathrm{Log}_{\gamma}\mathcal{W}_2$$

Multivariate Wasserstein AR model

$$\mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

1. identifiable: $\begin{cases} A \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \\ A' \to \mathbb{E}_{\oplus}\widetilde{\boldsymbol{\mu}}_{it}|\widetilde{\boldsymbol{\mu}}_{j,t-1} \end{cases}$ $\Longleftarrow$ Exp is not injective.

$\mathrm{Exp}_{\gamma}|_{\mathrm{Log}_{\gamma}\mathcal{W}}$ is an isometric homeomorphism from $\mathrm{Log}_{\gamma}\mathcal{W}_2$ to $\mathcal{W}_2$, with the inverse map Log_{γ}(Bigot et al., 2017).

2. A not tractable in estimation.

Thus, $1 + 2 \to$

$$\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - i) \in \mathrm{Log}_{\gamma}\mathcal{W}_2$$

$\forall g \in \mathrm{Tan}_{\gamma}$, $g \in \mathrm{Log}_{\gamma}\mathcal{W}_2 \iff g + id$ is non-decreasing γ-a.e, (Bigot et al., 2017).

Multivariate Wasserstein AR model

$$\tilde{\boldsymbol{\mu}}_{it} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{F}}_{i,t-1} - i)\right)$$

Thus, $1 + 2 \rightarrow$

$$\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{F}}_{i,t-1} - i) \in \mathrm{Log}_{\gamma}\, \mathcal{W}_2$$

$\forall g \in \mathrm{Tan}_{\gamma}$, $g \in \mathrm{Log}_{\gamma}\, \mathcal{W}_2 \iff g + id$ is non-decreasing γ-a.e, Bigot et al. (2017).

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

For $\boldsymbol{\mu}_{it}$, $t \in \mathbb{Z}, i = 1, \ldots, N$, we propose the Wasserstein multivariate AR Model

$$\widetilde{\boldsymbol{\mu}}_{it} = \boldsymbol{\epsilon}_{it} \# \operatorname{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij} (\widetilde{\boldsymbol{F}}_{i,t-1} - i)\right),$$

where $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ are the centralizations, and $\{\boldsymbol{\epsilon}_{it}\}_{i,t}$ are i.i.d. random increasing functions, $\boldsymbol{\epsilon}_{it}$ is almost surely independent of $\boldsymbol{\mu}_{j,t-1}$, $i, j = 1, \ldots, N$, for all $t \in \mathbb{Z}$, and

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{it}(x)\right] = x,\ x \in [0, 1].$$

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

For $\boldsymbol{\mu}_{it}$, $t \in \mathbb{Z}, i = 1, \ldots, N$, we propose the Wasserstein multivariate AR Model

$$\widetilde{\boldsymbol{\mu}}_{it} = \boldsymbol{\epsilon}_{it} \# \operatorname{Exp}_{Leb} \left(\sum_{j=1}^{N} A_{ij} (\widetilde{\boldsymbol{F}}_{i,t-1} - i) \right),$$

where $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ are the centralizations, and $\{\boldsymbol{\epsilon}_{it}\}_{i,t}$ are i.i.d. random increasing functions, $\boldsymbol{\epsilon}_{it}$ is almost surely independent of $\boldsymbol{\mu}_{j,t-1}$, $i, j = 1, \ldots, N$, for all $t \in \mathbb{Z}$, and

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{it}(x)\right] = x, \, x \in [0, 1].$$

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

Quantile function representation

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right],$$

For $\boldsymbol{\mu}_{it}$, $t \in \mathbb{Z}, i = 1, \ldots, N$, we propose the Wasserstein multivariate AR Model

$$\widetilde{\boldsymbol{\mu}}_{it} = \boldsymbol{\epsilon}_{it} \# \operatorname{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij} (\widetilde{\boldsymbol{F}}_{i,t-1} - i) \right),$$

where $\tilde{\boldsymbol{\mu}}_{it} \leftrightarrow \tilde{\boldsymbol{F}}_{it}^{-1} := \boldsymbol{F}_{it}^{-1} \circ [F_{i,\oplus}^{-1}]^{-1}$ are the centralizations, and $\{\boldsymbol{\epsilon}_{it}\}_{i,t}$ are i.i.d. random increasing functions, $\boldsymbol{\epsilon}_{it}$ is almost surely independent of $\boldsymbol{\mu}_{j,t-1}$, $i, j = 1, \ldots, N$, for all $t \in \mathbb{Z}$, and

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{it}(x)\right] = x,\ x \in [0, 1].$$

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

Quantile function representation

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \quad A \iff \mathcal{G}$$

Iterated random function system: TS analysis in metric space

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \tag{1}$$

Admissible as a TS model: existence, uniqueness and stationarity of solutions $\widetilde{\boldsymbol{F}}_{i,t}^{-1}, i = 1, \ldots N, t \in \mathbb{Z}$.

Wu and Shao (2004), IRF system in a complete, separable metric space $(\mathcal{X}, d)$, and ϵ_t i.i.d. :

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_{\epsilon_t}(\boldsymbol{X}_{t-1}), \quad \boldsymbol{X}_t \in (\mathcal{X}, d)$$

Iterated random function system: TS analysis in metric space

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right],$$

Admissible as a TS model: existence, uniqueness and stationarity of solutions $\widetilde{\boldsymbol{F}}_{i,t}^{-1}, i = 1, \ldots N, t \in \mathbb{Z}$.

Wu and Shao (2004), IRF system in a complete, separable metric space $(\mathcal{X}, d)$, and ϵ_t i.i.d. :

$$\boldsymbol{X}_t = \boldsymbol{\Phi}_{\epsilon_t}(\boldsymbol{X}_{t-1}), \quad \boldsymbol{X}_t \in (\mathcal{X}, d)$$

$\boldsymbol{\Phi}_{\epsilon_t}$ contractive at exp decay rate in expectation $\rightarrow$ *stability*

$$\rightarrow existence \xrightarrow{\text{add str}} stationarity.$$

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \quad i = 1, \ldots, N$$
$$\iff \boldsymbol{X}_t = \boldsymbol{\Phi}_{\epsilon_t}(\boldsymbol{X}_{t-1}), \quad \boldsymbol{X}_t \in (\mathcal{X}, d)$$

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \quad i = 1, \dots, N$$

$$\iff \boldsymbol{X}_t = \boldsymbol{\Phi}_{\epsilon_t}(\boldsymbol{X}_{t-1}), \quad \boldsymbol{X}_t \in (\mathcal{X}, d)$$

$$\boldsymbol{X}_t = \begin{pmatrix} \widetilde{\boldsymbol{F}}_{1,t}^{-1} \\ \widetilde{\boldsymbol{F}}_{2,t}^{-1} \\ \dots \\ \widetilde{\boldsymbol{F}}_{N,t}^{-1} \end{pmatrix}$$

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \quad i = 1, \ldots, N$$

$$\iff \boldsymbol{X}_t = \boldsymbol{\Phi}_{\epsilon_t}(\boldsymbol{X}_{t-1}), \quad \boldsymbol{X}_t \in (\mathcal{X}, d)$$

$$\boldsymbol{X}_t = \begin{pmatrix} \widetilde{\boldsymbol{F}}_{1,t}^{-1} \\ \widetilde{\boldsymbol{F}}_{2,t}^{-1} \\ \ldots \\ \widetilde{\boldsymbol{F}}_{N,t}^{-1} \end{pmatrix} \Rightarrow \widetilde{\boldsymbol{F}}_{j,t}^{-1} \in (\mathcal{T}, \|\cdot\|_{Leb}),$$

where $\mathcal{T} = \mathrm{Log}_{Leb} + id$ is the space of all quantile functions of $\mathcal{W}_2$.

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \quad i = 1, \ldots, N$$

$$\iff \boldsymbol{X}_t = \boldsymbol{\Phi}_{\epsilon_t}(\boldsymbol{X}_{t-1}), \quad \boldsymbol{X}_t \in (\mathcal{X}, d)$$

$$\boldsymbol{X}_t = \begin{pmatrix} \widetilde{\boldsymbol{F}}_{1,t}^{-1} \\ \widetilde{\boldsymbol{F}}_{2,t}^{-1} \\ \cdots \\ \widetilde{\boldsymbol{F}}_{N,t}^{-1} \end{pmatrix} \Rightarrow \widetilde{\boldsymbol{F}}_{j,t}^{-1} \in (\mathcal{T}, \|\cdot\|_{Leb}),$$

where $\mathcal{T} = \mathrm{Log}_{Leb} + id$ is the space of all quantile functions of $\mathcal{W}_2$.

$$\Rightarrow (\mathcal{X}, d) := (\mathcal{T}, \|\cdot\|_{Leb})^{\otimes N},$$

for any $\mathbf{X} = (\mathbf{X}_i)_{i=1}^N, \mathbf{Y} = (\mathbf{Y}_i)_{i=1}^N \in (\mathcal{T}, \|\cdot\|_{Leb})^{\otimes N}$

$$d(\mathbf{X}, \mathbf{Y}) := \sqrt{\sum_{i=1}^{N} \|\mathbf{X}_i - \mathbf{Y}_i\|_{Leb}^2}.$$

Existence, uniqueness

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right], \quad i = 1, \ldots, N. \quad (1)$$

Contraction of system (at exp decay rate)

1. $\mathbb{E} \left[\boldsymbol{\epsilon}_{i,t}(x) - \boldsymbol{\epsilon}_{i,t}(y) \right]^2 \leqslant L^2 (x-y)^2, \ \forall x, y \in [0,1], t \in \mathbb{Z}, \ i = 1, \ldots, N,$
2. $\|A\|_2 < \frac{1}{L}$.

Theorem

Under the assumptions above, the IRF system (1) almost surely admits a solution $\boldsymbol{X}_t$, $t \in \mathbb{Z}$, with $\boldsymbol{X}_t \overset{d}{=} \pi$, $\forall\, t \in \mathbb{Z}$. Moreover, if there exists another solution $\boldsymbol{S}_t$, $t \in \mathbb{Z}$, then for all $t \in \mathbb{Z}$

$$\boldsymbol{X}_t \overset{d}{=} \boldsymbol{S}_t, \quad \textit{almost surely.}$$

Stationarity

Stationarity

Definition: A random process $\{\mathbf{V}_t\}_t$ in a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is said to be stationary if

1. $\mathbb{E}\left\|\mathbf{V}_t\right\|^2 < \infty$
2. The Hilbert mean $U := \mathbb{E}\left[\mathbf{V}_t\right]$ does not depend on t.
3. The auto-covariance operators defined as

$$\mathcal{G}_{t,t-h}(V) := \mathbb{E}\left\langle \mathbf{V}_t - U, V\right\rangle \left(\mathbf{V}_{t-h} - U\right), \quad V \in \mathcal{H},$$

do not depend on t, that is $\mathcal{G}_{t,t-h}(V) = \mathcal{G}_{0,-h}(V)$ for all t.

Stationarity

Definition: A random process $\{\mathbf{V}_t\}_t$ in a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is said to be stationary if

1. $\mathbb{E}\|\mathbf{V}_t\|^2 < \infty$
2. The Hilbert mean $U := \mathbb{E}[\mathbf{V}_t]$ does not depend on t.
3. The auto-covariance operators defined as

$$\mathcal{G}_{t,t-h}(V) := \mathbb{E}\langle \mathbf{V}_t - U, V\rangle (\mathbf{V}_{t-h} - U), \quad V \in \mathcal{H},$$

   do not depend on t, that is $\mathcal{G}_{t,t-h}(V) = \mathcal{G}_{0,-h}(V)$ for all t.

$d(X,Y) = \sqrt{\sum_{i=1}^{N} \|X_i - Y_i\|_{Leb}^2}$ is induced by the inner product:

$$\langle X, Y\rangle = \sum_{i=1}^{N} \langle X_i, Y_i\rangle_{Leb}.$$

Stationarity

Definition: A random process $\{\mathbf{V}_t\}_t$ in a separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is said to be stationary if

1. $\mathbb{E}\left\|\mathbf{V}_t\right\|^2 < \infty$
2. The Hilbert mean $U := \mathbb{E}\left[\mathbf{V}_t\right]$ does not depend on t.
3. The auto-covariance operators defined as

$$\mathcal{G}_{t,t-h}(V) := \mathbb{E}\left\langle \mathbf{V}_t - U, V\right\rangle\left(\mathbf{V}_{t-h} - U\right), \quad V \in \mathcal{H},$$

   do not depend on t, that is $\mathcal{G}_{t,t-h}(V) = \mathcal{G}_{0,-h}(V)$ for all t.

$d(X,Y) = \sqrt{\sum_{i=1}^{N} \left\|X_i - Y_i\right\|_{Leb}^2}$ is induced by the inner product:

$$\langle X, Y\rangle = \sum_{i=1}^{N}\langle X_i, Y_i\rangle_{Leb}. \quad \longrightarrow \quad \text{Our system } \in (\mathcal{X}, \langle\cdot,\cdot\rangle)$$

Stationarity

Theorem

The unique solution given in Theorem 2 is stationary as a random process in $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ *in the sense of Definition above.*

Stationarity

Theorem

The unique solution given in Theorem 2 is stationary as a random process in $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ *in the sense of Definition above.*

The model we propose is justified, thus it is a valid TS model.

Constrained least-square estimation

For auto-regressive model

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) + id \right],$$

Given the centered observations $\widetilde{\boldsymbol{F}}_{t}^{-1}$, $t = 0, 1, \ldots, T$, we propose

$$\widetilde{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_{+}^{1}}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) - i \right\|_{Leb}^{2},$$

where B_{+}^{1} is the constraint set of N-simplex.

Constrained least-square estimation

Given the centered observations $\widetilde{\boldsymbol{F}}_t^{-1}$, $t = 0, 1, \ldots, T$, we propose

$$\widetilde{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) - i \right\|_{Leb}^2 ,$$

where B_+^1 is the constraint set of N-simplex.

Constrained least-square estimation

Given the centered observations $\widetilde{\boldsymbol{F}}_t^{-1}$, $t = 0, 1, \ldots, T$, we propose

$$\widetilde{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) - i \right\|_{Leb}^2,$$

where B_+^1 is the constraint set of N-simplex.

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},$$

where the population Fréchet mean $\mu_{i,\oplus}$ is also an unknown parameter so is $F_{i,\oplus}^{-1}$, we estimate by the empirical Fréchet mean

$$\bar{\boldsymbol{\mu}}_i := \underset{\nu \in \mathcal{W}_2}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} d_W^2(\boldsymbol{\mu}_{i,t}, \nu), \quad \text{with } \boldsymbol{F}_{\bar{\mu}_i}^{-1} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{F}_{\mu_{i,t}}^{-1}$$

Constrained least-square estimation

In practice, we center $\boldsymbol{\mu}_{i,t}$ by $\boldsymbol{F}_{\bar{\mu}_i}^{-1}$

$$\widehat{\boldsymbol{F}}_{i,t}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ [\boldsymbol{F}_{\bar{\mu}_i}^{-1}]^{-1}.$$

Constrained least-square estimation

In practice, we center $\boldsymbol{\mu}_{i,t}$ by $\boldsymbol{F}_{\bar{\mu}_i}^{-1}$

$$\widehat{\boldsymbol{F}}_{i,t}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ [\boldsymbol{F}_{\bar{\mu}_i}^{-1}]^{-1}.$$

$$\widehat{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widehat{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) - i \right\|_{Leb}^{2},$$

Constrained least-square estimation

In practice, we center $\boldsymbol{\mu}_{i,t}$ by $\boldsymbol{F}_{\bar{\mu}_i}^{-1}$

$$\widehat{\boldsymbol{F}}_{i,t}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ [\boldsymbol{F}_{\bar{\mu}_i}^{-1}]^{-1}.$$

$$\widehat{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widehat{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) - i \right\|_{Leb}^2,$$

The optimization problem (1) can be solved by the accelerated projected gradient descent (Parikh and Boyd, 2014, Chapter 4.3). The projection onto B_+^1 is given in Thai et al. (2015).

Constrained least-square estimation

In practice, we center $\boldsymbol{\mu}_{i,t}$ by $\boldsymbol{F}_{\bar{\mu}_i}^{-1}$

$$\widehat{\boldsymbol{F}}_{i,t}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ [\boldsymbol{F}_{\bar{\mu}_i}^{-1}]^{-1}.$$

$$\widehat{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widehat{\boldsymbol{F}}_{j,t-1}^{-1} - i \right) - i \right\|_{Leb}^2 ,$$

The optimization problem (1) can be solved by the accelerated projected gradient descent (Parikh and Boyd, 2014, Chapter 4.3). The projection onto B_+^1 is given in Thai et al. (2015).

Note that the N-simplex constraint promotes the **sparsity** in $\widehat{\boldsymbol{A}}$.

Constrained least-square estimation

Theorem

Assume[a] *the transformed sequence* $\widetilde{\boldsymbol{F}}_t^{-1}, t = 0, 1, \ldots, T$ *checks Model* (1) *with Assumption* N*-simplex true. Suppose additionally* $\widetilde{\boldsymbol{F}}_0^{-1} \stackrel{d}{=} \pi$ *with* π *the stationary distribution defined in Theorem 2. Given Assumption contraction of regression operation holds true. Then given the true coefficient* A *satisfies Assumption* N*-simplex, we have*

$$\widehat{\boldsymbol{A}} - A \xrightarrow{p} 0.$$

[a] The complete assumption sees Jiang (2022).

Related work: Univariate Wasserstein AR model

Describe this regressive dependencyship with

AR model of optimal transport (Zhu and Müller, 2021):

$$\boldsymbol{T}_{t+1} = \boldsymbol{\epsilon}_t \circ (\alpha(\boldsymbol{T}_t - i) + id), \quad 0 < \alpha < 1$$

AR model of tangent vector (Zhang et al., 2021):

$$\boldsymbol{T}_{t+1} - i = \alpha(\boldsymbol{T}_t - i) + \boldsymbol{\epsilon}_t, \qquad 0 < |\alpha| < 1,$$

Tangent vector with regression operator (Chen et al., 2021)

$$\boldsymbol{T}_{t+1} - i = \Gamma(\boldsymbol{T}_t - i) + \boldsymbol{\epsilon}_t, \quad \Gamma : Log_{\mu_\oplus}(\mathcal{W}) \rightarrow Log_{\mu_\oplus}(\mathcal{W})$$

the model in tangent space than is the ordinary AR model for functional TS in Hilbert space, expect the log image issue