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

Yiye JIANG

Institut de Mathématiques de Bordeaux
1. Data and problems
2. Model set up
3. Existence, uniqueness and stationarity
4. Estimation
5. Experiments: Age distribution of countries
1. Data and problems

2. Model set up

3. Existence, uniqueness and stationarity

4. Estimation

5. Experiments: Age distribution of countries
**Multivariate distributional time series**

**Figure 1:** Observations of the age distributions across European union countries over years 1995 to 2035 (projected).
Figure 2: Observations of the age distributions across European union countries over years 1995 to 2035 (projected). On the right are the observations $(\mu_{it})_t \in \mathcal{P}([0,1])$ along time recorded at $i = $ France. Lighter curves correspond to more recent years.
1. Data and problems

2. Model set up

3. Existence, uniqueness and stationarity

4. Estimation

5. Experiments: Age distribution of countries
Vector auto-regressive model
Vector auto-regressive model

Let $x_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series.
Vector auto-regressive model

Let $x_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}x_{it} = u_i$ exists and time invariant.
Vector auto-regressive model

Let \( x_{it} \in IR, \ t \in \mathbb{Z}, \ i = 1, \ldots, N, \) a multivariate time series. Assume \( E x_{it} = u_i \) exists and time invariant. The VAR model of order 1 writes as

\[
x_{it} - u_i = \sum_{j=1}^{N} A_{ij}(x_{j,t-1} - u_j) + \epsilon_{it},
\]

where \( \epsilon_{it} \) is a white noise,
Vector auto-regressive model

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

$$x_{it} - u_i = \sum_{j=1}^{N} A_{ij}(x_{j,t-1} - u_j) + \epsilon_{it},$$

where $\epsilon_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(x_{j,t-1} - u_j)$ is the regression operation.
Vector auto-regressive model

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

$$x_{it} - u_i = \sum_{j=1}^{N} A_{ij}(x_{j,t-1} - u_j) + \epsilon_{it},$$

where $\epsilon_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(x_{j,t-1} - u_j)$ is the regression operation.

Extension: $x_{it} \in IR \longrightarrow \mu_{it} \in \mathcal{W}_2(IR)$. 
Let \( x_{it} \in IR, \ t \in \mathbb{Z}, \ i = 1, \ldots, N, \) a multivariate time series. Assume \( \mathbb{E}x_{it} = u_i \) exists and time invariant. The VAR model of order 1 writes as

\[
x_{it} - u_i = \sum_{j=1}^{N} A_{ij} (x_{j,t-1} - u_j) + \epsilon_{it},
\]

where \( \epsilon_{it} \) is a white noise, and \( \sum_{j=1}^{N} A_{ij} (x_{j,t-1} - u_j) \) is the regression operation.

**Extension:** \( x_{it} \in IR \rightarrow \mu_{it} \in \mathcal{W}_2(IR). \)

\[
\pi \overset{\mathcal{W}_2(IR^d)}{\leftrightarrow} T
\]

\[
\overset{d=1}{\leftrightarrow} F_2^{-1} \circ F_1 \overset{\text{fix ref}}{\leftrightarrow} \text{functions}
\]
Related work: Univariate Wasserstein AR model

Chen et al. (2021); Zhang et al. (2021); Zhu and Müller (2021) extended the univariate AR model

$$x_t - u = \alpha(x_{t-1} - u) + \epsilon_t,$$

by interpreting the regression operation from the geometric point of view.
Figure 3: *Geometric interpretation of regression dependency.*
Related work: Univariate Wasserstein AR model

Figure 4: Geometric interpretation of regression dependency. $\mu_\oplus$ is the time-invariant Fréchet mean of $\mu_t$, $T_{t-1}$ is the optimal transport map which pushforwards $\mu_\oplus$ to $\mu_{t-1}$.
Multivariate Wasserstein AR model

Construction of univariate regression operation (ignoring the noise)

\[ x_t = u + \alpha(x_{t-1} - u) \implies \mu_t = \text{Exp}_{\mu} \left( \alpha(T_{t-1} - id) \right) \]
Construction of univariate regression operation (ignoring the noise)

\[ x_t = u + \alpha(x_{t-1} - u) \implies \mu_t = \text{Exp}_{\mu} (\alpha(T_{t-1} - \text{id})) \]

Multivariate regression operation

\[ x_{it} - u_i = \sum_{j=1}^{N} A_{ij} (x_{j,t-1} - u_j), \]
Construction of univariate regression operation (ignoring the noise)

\[ \mathbf{x}_t = u + \alpha(\mathbf{x}_{t-1} - u) \implies \mathbf{\mu}_t = \text{Exp}_{\mu_\oplus}(\alpha(\mathbf{T}_{t-1} - \text{id})) \]

Multivariate regression operation

\[ \mathbf{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\mathbf{x}_{j,t-1} - u_j), \]

\[ \Leftarrow \]

\[ \begin{cases} 
\text{Center} & \mathbf{\tilde{x}}_{it} = \mathbf{x}_{it} - u_i, \quad \text{ref pt} \quad \mathbb{E}\mathbf{\tilde{x}}_{it} = 0, \\
\text{Push} & \mathbf{\tilde{x}}_{it} = \sum_{j=1}^{N} A_{ij}\mathbf{\tilde{x}}_{jt}, 
\end{cases} \]
\[
\begin{align*}
\text{Center} & \quad \tilde{x}_{it} = x_{it} - u_i, \quad \implies \quad \tilde{\mu}_{it} = ? \quad \text{ref pt} \quad E_\oplus \tilde{\mu}_{it} = c \\
\text{Push} & \quad \tilde{x}_{it} = \sum_{j=1}^{N} A_{ij} \tilde{x}_{jt},
\end{align*}
\]
\[ \begin{cases} \text{Center} & \tilde{x}_{it} = x_{it} - u_i, \implies \tilde{\mu}_{it} = \text{ref pt} \implies \mathbb{E}_\oplus \tilde{\mu}_{it} = c \\ \text{Push} & \tilde{x}_{it} = \sum_{j=1}^{N} A_{ij} \tilde{x}_{jt}, \implies \tilde{\mu}_{it} = \exp_c \left( \sum_{j=1}^{N} A_{ij} (\tilde{T}_{i,t-1} - id) \right) \end{cases} \]

where \( \tilde{T}_{i,t-1} = \tilde{F}_{i,t-1}^{-1} \circ F_c \).
\[
\begin{align*}
\text{Center} & \quad \tilde{x}_{it} = x_{it} - u_i, \implies \tilde{\mu}_{it} = \text{ref} \implies \mathbb{E}_\oplus \tilde{\mu}_{it} = c \\
\text{Push} & \quad \tilde{x}_{it} = \sum_{j=1}^{N} A_{ij} \tilde{x}_{jt}, \implies \tilde{\mu}_{it} = \text{Exp}_c \left( \sum_{j=1}^{N} A_{ij} (\tilde{T}_{i,t-1} - \text{id}) \right)
\end{align*}
\]

where $\tilde{T}_{i,t-1} = \tilde{F}_{i,t-1}^{-1} \circ F_c$.

Center a random measure to Lebesgue mean

\[
\tilde{F}_{i,t}^{-1} = F_{i,t}^{-1} \oplus F_{i,\oplus}^{-1} := F_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},
\]

where $F_{i,t}^{-1}, F_{i,\oplus}^{-1}$ et $\tilde{F}_{i,t}^{-1}$ are respectively quantile functions of $\mu_{it}, \mathbb{E}_\oplus \mu_{it}$, and $\tilde{\mu}_{it}$, all $^{-1}$ are the left continuous inverse.
\[
\begin{cases}
\text{Center} & \tilde{x}_{it} = x_{it} - u_i, \quad \implies \tilde{\mu}_{it} = \text{ref pt } E_{\oplus} \tilde{\mu}_{it} = c \\
\text{Push} & \tilde{x}_{it} = \sum_{j=1}^{N} A_{ij}\tilde{x}_{jt}, \quad \implies \tilde{\mu}_{it} = \text{Exp}_{c} \left( \sum_{j=1}^{N} A_{ij}(\tilde{T}_{i,t-1} - \text{id}) \right)
\end{cases}
\]

where \( \tilde{T}_{i,t-1} = \bar{F}_{i,t-1} \circ F_c \).

Center a random measure to Lebesgue mean

\[
\bar{F}_{i,t}^{-1} = F_{i,t}^{-1} \ominus F_{i,\oplus}^{-1} := F_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},
\]

where \( F_{i,t}^{-1}, F_{i,\oplus}^{-1} \) et \( \bar{F}_{i,t}^{-1} \) are respectively quantile functions of \( \mu_{it}, E_{\oplus} \mu_{it} \), and \( \tilde{\mu}_{it} \), all \(-1\) are the left continuous inverse. The difference operation \( \ominus \) is proposed in Zhu and Müller (2021).
\[
\begin{aligned}
\text{Center} & \quad \tilde{x}_{it} = x_{it} - u_i, \quad \implies \quad \tilde{\mu}_{it} = ? \quad \implies \quad E_{\oplus} \tilde{\mu}_{it} = c \\
\text{Push} & \quad \tilde{x}_{it} = \sum_{j=1}^{N} A_{ij} \tilde{x}_{jt}, \quad \implies \quad \tilde{\mu}_{it} = \text{Exp}_c \left( \sum_{j=1}^{N} A_{ij} (\tilde{T}_{i,t-1} - \text{id}) \right)
\end{aligned}
\]

where \( \tilde{T}_{i,t-1} = \tilde{F}_{i,t-1}^{-1} \circ F_c \).

Center a random measure to Lebesgue mean

\[
\tilde{F}_{i,t}^{-1} = F_{i,t}^{-1} \oplus F_{i,\oplus}^{-1} := F_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},
\]

where \( F_{i,t}^{-1}, F_{i,\oplus}^{-1} \) et \( \tilde{F}_{i,t}^{-1} \) are respectively quantile functions of \( \mu_{it}, E_{\oplus} \mu_{it} \), and \( \tilde{\mu}_{it} \), all \( ^{-1} \) are the left continuous inverse. The difference operation \( \ominus \) is proposed in Zhu and Müller (2021).

**Assumption**

All \( \mu_{it}, t \in \mathbb{Z}, i = 1, \ldots, N \) are supported on \([0, 1]\).
Regression operation (ignoring the noise)

\[ \tilde{\mu}_{it} = \text{Exp}_{\text{Leb}} \left( \sum_{j=1}^{N} A_{ij}(\tilde{F}_{i,t-1} - id) \right) \]
Regression operation (ignoring the noise)

\[
\tilde{\mu}_{it} = \text{Exp}_{Leb} \left( \sum_{j=1}^{N} A_{ij} (\tilde{F}_{i,t-1} - id) \right)
\]
Regression operation (ignoring the noise)

\[ \tilde{\mu}_{it} = \text{Exp}_{Leb} \left( \sum_{j=1}^{N} A_{ij} (\tilde{F}_{i,t-1} - \text{id}) \right) \]

A tractable in estimation:
\[ \forall \gamma \text{ a.c. } \in \mathcal{W}, \text{Exp}_\gamma |_{\text{Log}_\gamma \mathcal{W}} \text{ is an isometric homeomorphism from \ Log}_\gamma \mathcal{W} \text{ to } \mathcal{W}, \text{ with the inverse map } \text{Log}_\gamma. \]
\[ \forall g \in \text{Tan}_\gamma, g \in \text{Log}_\gamma \mathcal{W} \iff g + \text{id} \text{ is non-decreasing } \gamma\text{-a.e, } \]
Bigot et al. (2017).
Regression operation (ignoring the noise)

\[ \tilde{\mu}_{it} = \text{Exp}_{\text{Leb}} \left( \sum_{j=1}^{N} A_{ij} (\tilde{F}_{i,t-1} - \text{id}) \right) \]

A tractable in estimation:
\[ \forall \gamma \text{ a.c. } \in \mathcal{W}, \text{ Exp}_\gamma \mid_{\text{Log}_\gamma \mathcal{W}} \text{ is an isometric homeomorphism from } \text{Log}_\gamma \mathcal{W} \text{ to } \mathcal{W}, \text{ with the inverse map } \text{Log}_\gamma. \]
\[ \forall g \in \text{Tan}_\gamma, g \in \text{Log}_\gamma \mathcal{W} \iff g + \text{id} \text{ is non-decreasing } \gamma\text{-a.e}, \]
Bigot et al. (2017).

Assumption

\[ \sum_{j=1}^{N} A_{ij} \leq 1 \text{ and } 0 \leq A_{ij} \leq 1. \]
Wasserstein multivariate AR Model

\[ \tilde{\mu}_{it} = \epsilon_{it} \# \text{Exp}_{\text{Leb}} \left( \sum_{j=1}^{N} A_{ij} (\tilde{F}_{i,t-1} - id) \right), \]

where \( \{\epsilon_{it}\}_{i,t} \) are i.i.d. random increasing functions, \( \epsilon_{it} \) is almost surely independent of \( \mu_{j,t-1}, i, j = 1, \ldots, N \), for all \( t \in \mathbb{Z} \), and

\[ \mathbb{E} [\epsilon_{it}(x)] = x, \ x \in [0, 1]. \]

Assumption

\[ \sum_{j=1}^{N} A_{ij} \leq 1 \text{ and } 0 \leq A_{ij} \leq 1. \]
Wasserstein multivariate AR Model

\[ \tilde{\mu}_{it} = \epsilon_{it} \# \Exp_{\text{Leb}} \left( \sum_{j=1}^{N} A_{ij} (\tilde{F}_{i,t-1} - id) \right), \]

where \( \{\epsilon_{it}\} \) are i.i.d. random increasing functions, \( \epsilon_{it} \) is almost surely independent of \( \mu_{j,t-1}, i,j = 1, \ldots, N \), for all \( t \in \mathbb{Z} \), and

\[ \mathbb{E}[\epsilon_{it}(x)] = x, \ x \in [0, 1]. \]

Assumption

\[ \sum_{j=1}^{N} A_{ij} \leq 1 \text{ and } 0 \leq A_{ij} \leq 1. \]

Quantile function representation

\[ \tilde{F}_{i,t}^{-1} = \epsilon_{i,t} \circ \left[ \sum_{j=1}^{N} A_{ij} \left( \tilde{F}_{j,t-1}^{-1} - id \right) + id \right], \]
Wasserstein multivariate AR Model

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

where \( \{\epsilon_{it}\}_{i,t} \) are i.i.d. random increasing functions, \( \epsilon_{it} \) is almost surely independent of \( \mu_{j,t-1}, i,j = 1, \ldots, N, \) for all \( t \in \mathbb{Z}, \) and

$$\mathbb{E} [\epsilon_{it}(x)] = x, \ x \in [0, 1].$$

Assumption

$$\sum_{j=1}^{N} A_{ij} \leq 1 \ \text{and} \ 0 \leq A_{ij} \leq 1.$$ 

Quantile function representation

$$\tilde{F}_{i,t}^{-1} = \epsilon_{i,t} \circ \left[ \sum_{j=1}^{N} A_{ij} \left( \tilde{F}_{j,t-1}^{-1} - id \right) + id \right], \quad A \iff \mathcal{G}$$
1. Data and problems

2. Model set up

3. Existence, uniqueness and stationarity

4. Estimation

5. Experiments: Age distribution of countries
Iterated random function (IRF) system

\[ \tilde{F}_{i,t}^{-1} = \epsilon_{i,t} \circ \left[ \sum_{j=1}^{N} A_{ij} \left( \tilde{F}_{j,t-1}^{-1} - id \right) + id \right], \quad (1) \]

Admissible as a TS model: existence, uniqueness and stationarity (additionally Hilbert space).
Iterated random function (IRF) system

\[ \hat{F}_{i,t}^{-1} = \epsilon_{i,t} \circ \left[ \sum_{j=1}^{N} A_{ij} \left( \hat{F}_{j,t-1}^{-1} - \text{id} \right) + \text{id} \right], \]

Admissible as a TS model: existence, uniqueness and stationarity (additionally Hilbert space).

Consider the product metric space

\[ (\mathcal{X}, d) := (\mathcal{T}, \| \cdot \|_{\text{Leb}})^{\otimes N}, \]

where \( \mathcal{T} := \text{Log}_{\text{Leb}} \mathcal{W} + \text{id} \) is the space of all quantile functions of \( \mathcal{W} \), equipped with the norm \( \| \cdot \|_{\text{Leb}} \) in the tangent space at the Lebesgue measure. Thus, for any \( X = (X_i)_{i=1}^{N}, Y = (Y_i)_{i=1}^{N} \in \mathcal{X} \)

\[ d(X, Y) := \sqrt{\sum_{i=1}^{N} \| X_i - Y_i \|_{\text{Leb}}^2}. \]
By Wu and Shao (2004), IRF system in a complete, separable metric space

\[
\text{exp decay rate} \rightarrow \text{stability} \rightarrow \text{existence} \rightarrow \text{stationarity.}
\]
Existence, uniqueness and stationarity

By Wu and Shao (2004), IRF system in a complete, separable metric space

\[ \exp \text{ decay rate} \rightarrow \text{stability} \rightarrow \text{existence} \xrightarrow{\text{add str}} \text{stationarity.} \]

Assumption

Contraction of the regression operator (at exp decay rate)

1. \( \mathbb{E} [\epsilon_{i,t}(x) - \epsilon_{i,t}(y)]^2 \leq L^2(x - y)^2, \forall x, y \in [0, 1], t \in \mathbb{Z}, i = 1, \ldots, N, \)

2. \( \|A\|_2 < \frac{1}{L}. \)
By Wu and Shao (2004), IRF system in a complete, separable metric space

\[ \text{exp decay rate} \rightarrow \text{stability} \rightarrow \text{existence} \xrightarrow{\text{add str}} \text{stationarity}. \]

**Assumption**

*Contraction of the regression operator (at exp decay rate)*

1. \[ E[\epsilon_{i,t}(x) - \epsilon_{i,t}(y)]^2 \leq L^2 (x - y)^2, \quad \forall x, y \in [0, 1], \quad t \in \mathbb{Z}, \quad i = 1, \ldots, N, \]
2. \[ \|A\|_2 < \frac{1}{L}. \]

**Theorem**

*Under Assumptions N-simplex and contraction, the IRF system (1) almost surely admits a solution \( X_t, \quad t \in \mathbb{Z} \), with \( X_t \overset{d}{=} \pi, \quad \forall \ t \in \mathbb{Z} \). Moreover, if there exists another solution \( S_t, \quad t \in \mathbb{Z} \), then for all \( t \in \mathbb{Z} \)

\[ X_t \overset{d}{=} S_t, \quad \text{almost surely}. \]
Existence, uniqueness and stationarity

$(X, d)$ with $d$ the induced metric of inner product

$$\langle X, Y \rangle = \sum_{i=1}^{N} \langle X_i, Y_i \rangle_{\text{Leb}}.$$
\((X, d)\) with \(d\) the induced metric of inner product

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

A random process \(\{V_t\}_t\) in a separable Hilbert space \(\mathcal{H}, \langle \cdot, \cdot \rangle\) is said to be stationary if the following properties are satisfied.

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

\[
G_{t, t-h}(V) := \mathbb{E} \langle V_t - U, V \rangle (V_{t-h} - U), \quad V \in \mathcal{H},
\]

do not depend on \(t\), that is \(G_{t, t-h}(V) = G_{0, t-h}(V)\) for all \(t\).
Existence, uniqueness and stationarity

$(X, d)$ with $d$ the induced metric of inner product

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

A random process $\{V_t\}_t$ in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is said to be stationary if the following properties are satisfied.

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

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

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

**Theorem**

The unique solution given in Theorem 1 is stationary as a random process in $(X, \langle \cdot, \cdot \rangle)$ in the sense of Definition above.
1. Data and problems

2. Model set up

3. Existence, uniqueness and stationarity

4. Estimation

5. Experiments: Age distribution of countries
Constrained least-square estimation

\[ \widetilde{A}_i : = \arg \min_{A_i : \in B^1_+} \frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{F}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left( \widetilde{F}_{j,t-1}^{-1} - id \right) - id \right\|_{Leb}^2, \]

where \( B^1_+ \) is the constraint set of \( N \)-simplex.
Constrained least-square estimation

\[ \begin{align*}
\widetilde{A}_i &= \arg \min_{A_i \in B^1_+} \frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{F}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left( \tilde{F}_{j,t-1}^{-1} - id \right) - id \right\|^2_{Leb}, \\
\end{align*} \]

where \( B^1_+ \) is the constraint set of \( N \)-simplex.

Population mean is also an unknown parameter, we estimate as

\[ \begin{align*}
F_{\mu_i}^{-1} &= \frac{1}{T} \sum_{t=1}^{T} F_{\mu_{i,t}}^{-1}, \\
\end{align*} \]

and center \( \mu_{i,t} \) by \( F_{\mu_i}^{-1} \) with difference \( \ominus \)

\[ \begin{align*}
\hat{F}_{i,t}^{-1} := F_{i,t}^{-1} \ominus F_{\mu_i}^{-1} = F_{i,t}^{-1} \circ F_{\mu_i}. \\
\end{align*} \]
Constrained least-square estimation

\[ \hat{A}_i = \arg \min_{A_i \in B^1_+} \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left( \hat{F}_{j,t-1}^{-1} - id \right) - id \right\|_2^2 \]  

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

\[ \hat{A}_i: = \arg\min_{A_i: \in B_1^+} \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left( \hat{F}_{j,t-1}^{-1} - id \right) - id \right\|_2^2, \]

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 \( \hat{A} \).
Theorem

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

\[ \hat{A} - A \overset{p}{\to} 0. \]

\(^a\)Complete statement of theorem sees Jiang (2022)
Experiments: Age distribution of countries

1. Data and problems
2. Model set up
3. Existence, uniqueness and stationarity
4. Estimation
5. Experiments: Age distribution of countries
Figure 5: *Inferred age structure graph*. The non-zero coefficients $A_{ij}$ are represented by the weighted directed edges from node $j$ to node $i$. Thicker arrow corresponds to larger weights. The blue circles around nodes represent the weights of self-loop.
Figure 6: Evolution of age structure from 1996 to 2036 (projected). Estonia (top left), Latvia (top right), Sweden (bottom left) versus Norway (bottom right).
Figure 7: *Evolution of age structure from 1996 to 2036 (projected) of France (left) versus Italy (right).*
Table 1: Top 5 edges with the largest weights excluding all the self-loops, based on the data from 1996 to 2036 (projected).

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Estonia</td>
<td>Latvia</td>
</tr>
<tr>
<td>2 Sweden</td>
<td>Norway</td>
</tr>
<tr>
<td>3 Belgium</td>
<td>Germany</td>
</tr>
<tr>
<td>4 Finland</td>
<td>Netherlands</td>
</tr>
<tr>
<td>5 France</td>
<td>Greece</td>
</tr>
</tbody>
</table>


The space $\mathcal{W} := \mathcal{W}_2(\mathbb{R})$ has a pseudo-Riemannian structure (Ambrosio et al., 2008). Let $\gamma \in \mathcal{W}$ be an absolutely continuous measure, the tangent space at $\gamma$ is defined as

$$\text{Tan}_\gamma = \{ t(F^{-1}_\mu \circ F_\gamma - \text{id}) : \mu \in \mathcal{W}, \ t > 0 \}^{L^2(\mathbb{R})},$$

\textbf{Definition}

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

$$\text{Exp}_\gamma g = (g + \text{id}) \# \gamma.$$  

\textbf{Definition}

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

$$\text{Log}_\gamma \mu = F^{-1}_\mu \circ F_\gamma - \text{id}.$$
Related work: Univariate Wasserstein AR model

Describe this regression relationship with

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

\[ T_{t+1} = \epsilon_t \circ (\alpha (T_t - id) + id), \quad 0 < \alpha < 1 \]

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

\[ T_{t+1} - id = \alpha (T_t - id) + \epsilon_t, \quad 0 < |\alpha| < 1, \]

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

\[ T_{t+1} - id = \Gamma (T_t - id) + \epsilon_t, \quad \Gamma : Log_{\mu}(\mathcal{W}) \to Log_{\mu}(\mathcal{W}) \]

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