Conditional Independence Testing for Nonstationary Time Series

Michael Wieck-Sosa

joint work with Michel F. C. Haddad and Aaditya Ramdas

Department of Statistics & Data Science
Carnegie Mellon University

IWSM 2026
Session 4: Sequential Methods for Signals and Complex Data

Overview of presentation

Overview of presentation

Part I

Introduction

Time series setting

Null hypothesis

Real data application

Overview of presentation

Part I

Introduction

Time series setting

Null hypothesis

Real data application

Part II

Testing procedure

Nonlinear regression

Gaussian approx.

Varying covariances

Overview of presentation

Part I	Part II	Part III
Introduction	**Testing procedure**	**Theory**
Time series setting	Nonlinear regression	Hardness result
Null hypothesis	Gaussian approx.	Physical dependence
Real data application	Varying covariances	Type I error control

Time series setting

- We observe X_t, $t = 1, \ldots, n$, for some sample size $n \in \mathbb{N}$

Time series setting

- We observe X_t, $t = 1, \ldots, n$, for some sample size $n \in \mathbb{N}$
- Each observation X_t takes values in $\mathbb{R}^d$ for some dimension $d \in \mathbb{N}$

Time series setting

- We observe X_t, $t = 1, \ldots, n$, for some sample size $n \in \mathbb{N}$
- Each observation X_t takes values in $\mathbb{R}^d$ for some dimension $d \in \mathbb{N}$
- **Temporal dependence:** X_t may depend on the past $X_{t-1}, X_{t-2}, \ldots$

Time series setting

- We observe X_t, $t = 1, \dots, n$, for some sample size $n \in \mathbb{N}$
- Each observation X_t takes values in $\mathbb{R}^d$ for some dimension $d \in \mathbb{N}$
- **Temporal dependence:** X_t may depend on the past $X_{t-1}, X_{t-2}, \dots$
- **Nonstationarity:** joint distribution of $X_{t_1+\tau}, \dots, X_{t_k+\tau}$ may change with τ

Conditional independence testing

- Let $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$ be nonstationary time series taking values in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, $\mathbb{R}^{d_Z}$

Conditional independence testing

- Let $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$ be nonstationary time series taking values in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, $\mathbb{R}^{d_Z}$
- X_t, Y_t, Z_t may include **any leads and lags**, but Z_t must be known at time t

Conditional independence testing

- Let $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$ be nonstationary time series taking values in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, $\mathbb{R}^{d_Z}$
- X_t, Y_t, Z_t may include **any leads and lags**, but Z_t must be known at time t
- We present a test for the null hypothesis that

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

Conditional independence testing

- Let $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$ be nonstationary time series taking values in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, $\mathbb{R}^{d_Z}$
- X_t, Y_t, Z_t may include **any leads and lags**, but Z_t must be known at time t
- We present a test for the null hypothesis that

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- **Intuitively:** knowing Z_t renders X_t irrelevant for "learning about" Y_t

Conditional independence testing

- Let $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$ be nonstationary time series taking values in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, $\mathbb{R}^{d_Z}$
- X_t, Y_t, Z_t may include **any leads and lags**, but Z_t must be known at time t
- We present a test for the null hypothesis that

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- **Intuitively:** knowing Z_t renders X_t irrelevant for "learning about" Y_t
- If X_t, Y_t, Z_t have a joint density with respect to the Lebesgue measure, then

$$\begin{aligned} X_t \perp\!\!\!\perp Y_t \mid Z_t &\iff f_{X_t,Y_t|Z_t}(x, y \mid z) = f_{X_t|Z_t}(x \mid z) f_{Y_t|Z_t}(y \mid z) \\ &\iff f_{Y_t|X_t,Z_t}(y \mid x, z) = f_{Y_t|Z_t}(y \mid z) \end{aligned}$$

Conditional independence testing

- Let $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$ be nonstationary time series taking values in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_Y}$, $\mathbb{R}^{d_Z}$
- X_t, Y_t, Z_t may include **any leads and lags**, but Z_t must be known at time t
- We present a test for the null hypothesis that

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

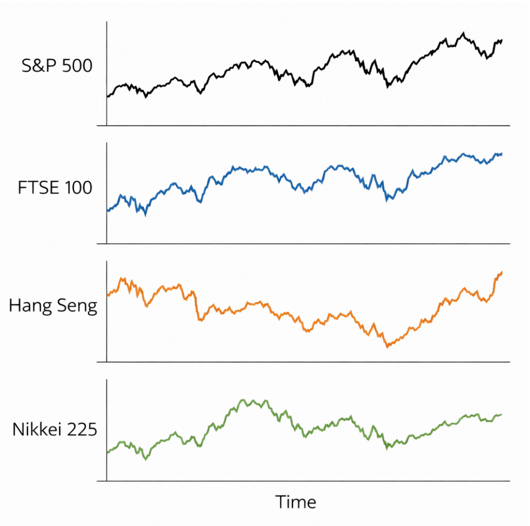
- **Intuitively:** knowing Z_t renders X_t irrelevant for "learning about" Y_t
- If X_t, Y_t, Z_t have a joint density with respect to the Lebesgue measure, then

$$\begin{aligned} X_t \perp\!\!\!\perp Y_t \mid Z_t &\iff f_{X_t,Y_t|Z_t}(x, y \mid z) = f_{X_t|Z_t}(x \mid z) f_{Y_t|Z_t}(y \mid z) \\ &\iff f_{Y_t|X_t,Z_t}(y \mid x, z) = f_{Y_t|Z_t}(y \mid z) \end{aligned}$$

- **Motivations:** causal discovery, variable selection, and *understanding structure*

Real data application: links in the global economic system

- Four stock market indices:
  1. S&P 500 (United States)
  2. FTSE 100 (United Kingdom)
  3. Hang Seng (Hong Kong)
  4. Nikkei 225 (Japan)
- Daily returns based on closing prices
- Trading days January 2022–April 2025
- Sample size: $n = 860$
- Results based on BH-adjusted p-values



Unconditional independence testing does not reveal much structure

- For each pair $X, Y \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \text{ for all times } t = 1, \ldots, n$$

Unconditional independence testing does not reveal much structure

- For each pair $X, Y \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \text{ for all times } t = 1, \ldots, n$$

- Same-day returns appear to exhibit dependence (reject each null)

Unconditional independence testing does not reveal much structure

- For each pair $X, Y \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \text{ for all times } t = 1, \ldots, n$$

- Same-day returns appear to exhibit dependence (reject each null)
- Not very interesting!

Conditional independence testing reveals more structure

- For each triplet $X, Y, Z \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

Conditional independence testing reveals more structure

- For each triplet $X, Y, Z \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- Reject nulls: US market independent of UK market given either Asian market

Conditional independence testing reveals more structure

- For each triplet $X, Y, Z \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- Reject nulls: US market independent of UK market given either Asian market
- Retain nulls: US market independent of either Asian market given UK market

Conditional independence testing reveals more structure

- For each triplet $X, Y, Z \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- Reject nulls: US market independent of UK market given either Asian market
- Retain nulls: US market independent of either Asian market given UK market
- Earliest to latest closing exchanges: Tokyo, Hong Kong, London, New York

Conditional independence testing reveals more structure

- For each triplet $X, Y, Z \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- Reject nulls: US market independent of UK market given either Asian market
- Retain nulls: US market independent of either Asian market given UK market
- Earliest to latest closing exchanges: Tokyo, Hong Kong, London, New York
- UK has info about Asian markets on the same day, but not vice versa

Conditional independence testing reveals more structure

- For each triplet $X, Y, Z \in \{\texttt{S\&P}, \texttt{FTSE}, \texttt{HangSeng}, \texttt{Nikkei}\}$, test

$$X_t \perp\!\!\!\perp Y_t \mid Z_t \text{ for all times } t = 1, \ldots, n$$

- Reject nulls: US market independent of UK market given either Asian market
- Retain nulls: US market independent of either Asian market given UK market
- Earliest to latest closing exchanges: Tokyo, Hong Kong, London, New York
- UK has info about Asian markets on the same day, but not vice versa
- **Next:** explain the testing procedure

Procedure for conditional independence test (1/2)

1. Regress X_t on Z_t, then calculate residuals $\hat{\varepsilon}_t$, $t = 1, \ldots, n$

Procedure for conditional independence test (1/2)

1. Regress X_t on Z_t, then calculate residuals $\hat{\varepsilon}_t$, $t = 1, \ldots, n$
2. Regress Y_t on Z_t, then calculate residuals $\hat{\xi}_t$, $t = 1, \ldots, n$

Procedure for conditional independence test (1/2)

1. Regress X_t on Z_t, then calculate residuals $\hat{\varepsilon}_t$, $t = 1, \dots, n$
2. Regress Y_t on Z_t, then calculate residuals $\hat{\xi}_t$, $t = 1, \dots, n$
3. Obtain residual products $\hat{R}_t = \mathrm{vec}(\hat{\varepsilon}_t \hat{\xi}_t^\top)$, $t = 1, \dots, n$

Procedure for conditional independence test (1/2)

1. Regress X_t on Z_t, then calculate residuals $\hat{\varepsilon}_t$, $t = 1, \ldots, n$
2. Regress Y_t on Z_t, then calculate residuals $\hat{\xi}_t$, $t = 1, \ldots, n$
3. Obtain residual products $\hat{R}_t = \mathrm{vec}(\hat{\varepsilon}_t \hat{\xi}_t^\top)$, $t = 1, \ldots, n$
4. Select window size L for covariance estimation based on $\hat{R}_{1:n}$ (minimum volatility)

Procedure for conditional independence test (1/2)

1. Regress X_t on Z_t, then calculate residuals $\hat{\varepsilon}_t$, $t = 1, \dots, n$
2. Regress Y_t on Z_t, then calculate residuals $\hat{\xi}_t$, $t = 1, \dots, n$
3. Obtain residual products $\hat{R}_t = \mathrm{vec}(\hat{\varepsilon}_t \hat{\xi}_t^\top)$, $t = 1, \dots, n$
4. Select window size L for covariance estimation based on $\hat{R}_{1:n}$ (minimum volatility)
5. Calculate test statistic based on time series of residual products

$$T(\hat{R}_{L:n}) = \max_{j=L,\dots,n} \left\| \frac{1}{\sqrt{n-L+1}} \sum_{t=L}^{j} \hat{R}_t \right\|_p, \quad p \in [2, \infty]$$

Procedure for conditional independence test (2/2)

6. For each time $t = L, \ldots, n$, obtain estimates of *time-varying* covariance matrices

$$\hat{\Sigma}_t = \frac{1}{L} \left(\sum_{j=t-L+1}^{t} \hat{R}_j \right) \left(\sum_{j=t-L+1}^{t} \hat{R}_j \right)^\top$$

Procedure for conditional independence test (2/2)

6. For each time $t = L, \ldots, n$, obtain estimates of *time-varying* covariance matrices

$$\hat{\Sigma}_t = \frac{1}{L} \left(\sum_{j=t-L+1}^{t} \hat{R}_j \right) \left(\sum_{j=t-L+1}^{t} \hat{R}_j \right)^\top$$

7. For each simulation $r = 1, \ldots, s$, sample *independent* Gaussian random vectors

$$\tilde{R}_t^{(r)} \sim N(0, \hat{\Sigma}_t), \quad t = L, \ldots, n$$

Procedure for conditional independence test (2/2)

6. For each time $t = L, \ldots, n$, obtain estimates of *time-varying* covariance matrices

$$\hat{\Sigma}_t = \frac{1}{L} \left(\sum_{j=t-L+1}^{t} \hat{R}_j \right) \left(\sum_{j=t-L+1}^{t} \hat{R}_j \right)^\top$$

7. For each simulation $r = 1, \ldots, s$, sample *independent* Gaussian random vectors

$$\tilde{R}_t^{(r)} \sim N(0, \hat{\Sigma}_t), \quad t = L, \ldots, n$$

8. For each simulation $r = 1, \ldots, s$, calculate test statistic based on the Gaussians

$$T(\tilde{R}_{L:n}^{(r)}) = \max_{j=L,\ldots,n} \left\| \frac{1}{\sqrt{n-L+1}} \sum_{t=L}^{j} \tilde{R}_t^{(r)} \right\|_p, \quad p \in [2, \infty]$$

Procedure for conditional independence test (2/2)

6. For each time $t = L, \ldots, n$, obtain estimates of *time-varying* covariance matrices

$$\hat{\Sigma}_t = \frac{1}{L}\left(\sum_{j=t-L+1}^{t} \hat{R}_j\right)\left(\sum_{j=t-L+1}^{t} \hat{R}_j\right)^\top$$

7. For each simulation $r = 1, \ldots, s$, sample *independent* Gaussian random vectors

$$\tilde{R}_t^{(r)} \sim N(0, \hat{\Sigma}_t), \quad t = L, \ldots, n$$

8. For each simulation $r = 1, \ldots, s$, calculate test statistic based on the Gaussians

$$T(\tilde{R}_{L:n}^{(r)}) = \max_{j=L,\ldots,n} \left\| \frac{1}{\sqrt{n-L+1}} \sum_{t=L}^{j} \tilde{R}_t^{(r)} \right\|_p, \quad p \in [2, \infty]$$

9. Calculate the $1-\alpha$ empirical quantile $\hat{q}_{1-\alpha}^{\mathrm{boot}}$ of $T(\tilde{R}_{L:n}^{(1)}), \ldots, T(\tilde{R}_{L:n}^{(s)})$

Procedure for conditional independence test (2/2)

6. For each time $t = L, \ldots, n$, obtain estimates of *time-varying* covariance matrices

$$\hat{\Sigma}_t = \frac{1}{L}\left(\sum_{j=t-L+1}^{t} \hat{R}_j\right)\left(\sum_{j=t-L+1}^{t} \hat{R}_j\right)^\top$$

7. For each simulation $r = 1, \ldots, s$, sample *independent* Gaussian random vectors

$$\tilde{R}_t^{(r)} \sim N(0, \hat{\Sigma}_t), \quad t = L, \ldots, n$$

8. For each simulation $r = 1, \ldots, s$, calculate test statistic based on the Gaussians

$$T(\tilde{R}_{L:n}^{(r)}) = \max_{j=L,\ldots,n} \left\| \frac{1}{\sqrt{n-L+1}} \sum_{t=L}^{j} \tilde{R}_t^{(r)} \right\|_p, \quad p \in [2, \infty]$$

9. Calculate the $1-\alpha$ empirical quantile $\hat{q}_{1-\alpha}^{\mathrm{boot}}$ of $T(\tilde{R}_{L:n}^{(1)}), \ldots, T(\tilde{R}_{L:n}^{(s)})$
10. Reject null hypothesis at significance level α if $T(\hat{R}_{L:n}) > \hat{q}_{1-\alpha}^{\mathrm{boot}}$, else retain

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

- Any CI test with size α has **no power** (power at most α against every alternative)

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

- Any CI test with size α has **no power** (power at most α against every alternative)
- If a CI test has power β against *some* alternative distribution, then there exists a null distribution s.t. *the test will reject with probability at least* β

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

- Any CI test with size α has **no power** (power at most α against every alternative)
- If a CI test has power β against *some* alternative distribution, then there exists a null distribution s.t. *the test will reject with probability at least* β

- Can only hope to achieve Type I error control on subset of null distributions $\mathcal{P}_n^{\mathrm{CI}}$

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

- Any CI test with size α has **no power** (power at most α against every alternative)
- If a CI test has power β against *some* alternative distribution, then there exists a null distribution s.t. *the test will reject with probability at least* β

- Can only hope to achieve Type I error control on subset of null distributions $\mathcal{P}_n^{\mathrm{CI}}$
- Need to restrict the null hypothesis to make the CI testing problem feasible

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

- Any CI test with size α has **no power** (power at most α against every alternative)
- If a CI test has power β against *some* alternative distribution, then there exists a null distribution s.t. *the test will reject with probability at least* β

- Can only hope to achieve Type I error control on subset of null distributions $\mathcal{P}_n^{\mathrm{CI}}$
- Need to restrict the null hypothesis to make the CI testing problem feasible
- This result has been extended to the time series setting (Bodik and Pasche 2024)

The hardness of conditional independence testing

No-free-lunch in conditional independence (CI) testing (Shah and Peters 2020)

- Any CI test with size α has **no power** (power at most α against every alternative)
- If a CI test has power β against *some* alternative distribution, then there exists a null distribution s.t. *the test will reject with probability at least* β

- Can only hope to achieve Type I error control on subset of null distributions $\mathcal{P}_n^{\mathrm{CI}}$
- Need to restrict the null hypothesis to make the CI testing problem feasible
- This result has been extended to the time series setting (Bodik and Pasche 2024)
- **Next:** explain how to restrict null hypothesis via assumptions on $X_{1:n}$, $Y_{1:n}$, $Z_{1:n}$

Assumption 1: representation in terms of noise inputs (Wu 2005)

- Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be an iid sequence of random variables and denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$

Assumption 1: representation in terms of noise inputs (Wu 2005)

- Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be an iid sequence of random variables and denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$
- Let Θ be a parameter space, which can be infinite-dimensional

Assumption 1: representation in terms of noise inputs (Wu 2005)

- Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be an iid sequence of random variables and denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \dots)$
- Let Θ be a parameter space, which can be infinite-dimensional
- For each $n \in \mathbb{N}$, let $G_t^{(n)} : \mathbb{R}^\infty \times \Theta \to \mathbb{R}^d$, $t = 1, \dots, n$, define a *generative model*

Assumption 1: representation in terms of noise inputs (Wu 2005)

- Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be an iid sequence of random variables and denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$
- Let Θ be a parameter space, which can be infinite-dimensional
- For each $n \in \mathbb{N}$, let $G_t^{(n)} : \mathbb{R}^\infty \times \Theta \to \mathbb{R}^d$, $t = 1, \ldots, n$, define a *generative model*

Assumption

The time series $X_{1:n}$ is generated by nature as

$$X_t = G_t^{(n)}(\boldsymbol{\varepsilon}_t, \theta), \quad t = 1, \ldots, n$$

for some parameter $\theta \in \Theta$ and noise inputs $\boldsymbol{\varepsilon}_t$

Assumption 1: representation in terms of noise inputs (Wu 2005)

- Let $(\varepsilon_i)_{i\in\mathbb{Z}}$ be an iid sequence of random variables and denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$
- Let Θ be a parameter space, which can be infinite-dimensional
- For each $n \in \mathbb{N}$, let $G_t^{(n)} : \mathbb{R}^\infty \times \Theta \to \mathbb{R}^d$, $t = 1, \ldots, n$, define a *generative model*

Assumption

The time series $X_{1:n}$ is generated by nature as

$$X_t = G_t^{(n)}(\boldsymbol{\varepsilon}_t, \theta), \quad t = 1, \ldots, n$$

for some parameter $\theta \in \Theta$ and noise inputs $\boldsymbol{\varepsilon}_t$

- **We do not know** the generative model $G_t^{(n)}$, noise inputs $\boldsymbol{\varepsilon}_t$, or parameter θ

Assumption 2: decaying influence of noise inputs (Wu 2005)

- For all $j \in \mathbb{N}$, denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ with ε_{t-j} replaced by an iid copy ε^*_{t-j} as

$$\tilde{\boldsymbol{\varepsilon}}_{t,j} = \left(\varepsilon_t, \ldots, \varepsilon_{t-j+1}, \varepsilon^*_{t-j}, \varepsilon_{t-j-1}, \ldots\right)$$

Assumption 2: decaying influence of noise inputs (Wu 2005)

- For all $j \in \mathbb{N}$, denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ with ε_{t-j} replaced by an iid copy ε^*_{t-j} as

$$\tilde{\boldsymbol{\varepsilon}}_{t,j} = \left(\varepsilon_t, \ldots, \varepsilon_{t-j+1}, \varepsilon^*_{t-j}, \varepsilon_{t-j-1}, \ldots\right)$$

- We assume the influence of the j-th noise input in the past *decays* as j grows

Assumption 2: decaying influence of noise inputs (Wu 2005)

- For all $j \in \mathbb{N}$, denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$ with ε_{t-j} replaced by an iid copy ε^*_{t-j} as

$$\tilde{\boldsymbol{\varepsilon}}_{t,j} = \left(\varepsilon_t, \ldots, \varepsilon_{t-j+1}, \varepsilon^*_{t-j}, \varepsilon_{t-j-1}, \ldots\right)$$

- We assume the influence of the j-th noise input in the past *decays* as j grows

Assumption

There exist constants $C > 0$, $\rho > 1$ such that, for some $q > 2$, we have

$$\sup_{n,t,\theta} \left\| G_t^{(n)}(\boldsymbol{\varepsilon}_t, \theta) - G_t^{(n)}(\tilde{\boldsymbol{\varepsilon}}_{t,j}, \theta) \right\|_{L^q(\theta)} \leq C\, j^{-\rho}, \quad j \in \mathbb{N}$$

where $\|\cdot\|_{L^q(\theta)} = \mathbb{E}_{P_\theta} \left(\|\cdot\|_2^q\right)^{1/q}$

Asymptotic Type I error control, uniformly over large collection of nulls

- **Assumption 3:** convergence rates of regression estimators (double robustness)

Asymptotic Type I error control, uniformly over large collection of nulls

- **Assumption 3:** convergence rates of regression estimators (double robustness)
- $\forall n \in \mathbb{N}$, let $\mathcal{P}_n^{\mathrm{CI}}$ be collection of distributions s.t. $X_t \perp\!\!\!\perp Y_t \mid Z_t$ for all $t = 1, \ldots, n$

Asymptotic Type I error control, uniformly over large collection of nulls

- **Assumption 3:** convergence rates of regression estimators (double robustness)
- $\forall n \in \mathbb{N}$, let $\mathcal{P}_n^{\mathrm{CI}}$ be collection of distributions s.t. $X_t \perp\!\!\!\perp Y_t \mid Z_t$ for all $t = 1, \ldots, n$

Theorem (Informal)

Suppose Assumptions 1-3 hold for all distributions $P \in \mathcal{P}_n^* \subset \mathcal{P}_n^{\mathrm{CI}}$ for every $n \in \mathbb{N}$. Then we have uniformly asymptotic Type I error control

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}_n^*} \mathbb{P}_P \left(T(\hat{R}_{L:n}) > \hat{q}_{1-\alpha}^{\mathrm{boot}} \right) \leq \alpha$$

where $T(\hat{R}_{L:n})$ is the test statistic and $\hat{q}_{1-\alpha}^{\mathrm{boot}}$ is the estimated quantile.

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series
- **Key technical ingredients:** (1) distribution-uniform strong Gaussian approximation, and (2) nonlinear regression, both for nonstationary time series

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series
- **Key technical ingredients:** (1) distribution-uniform strong Gaussian approximation, and (2) nonlinear regression, both for nonstationary time series
- Waudby-Smith and Ramdas (2023) extend GCM test to the iid sequential setting. Anytime-valid test, requires distribution-uniform strong (a.s.) Gaussian approx.

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series
- **Key technical ingredients:** (1) distribution-uniform strong Gaussian approximation, and (2) nonlinear regression, both for nonstationary time series
- Waudby-Smith and Ramdas (2023) extend GCM test to the iid sequential setting. Anytime-valid test, requires distribution-uniform strong (a.s.) Gaussian approx.
  - Builds on previous work by Waudby-Smith, Arbour, et al. (2024) on asymptotic confidence sequences (AsympCSs)

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series
- **Key technical ingredients:** (1) distribution-uniform strong Gaussian approximation, and (2) nonlinear regression, both for nonstationary time series
- Waudby-Smith and Ramdas (2023) extend GCM test to the iid sequential setting. Anytime-valid test, requires distribution-uniform strong (a.s.) Gaussian approx.
  - Builds on previous work by Waudby-Smith, Arbour, et al. (2024) on asymptotic confidence sequences (AsympCSs)
- **Question 1:** can distribution-uniform anytime-valid inference be extended to nonstationary time series? Sequential GCM for nonstationary time series?

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series
- **Key technical ingredients:** (1) distribution-uniform strong Gaussian approximation, and (2) nonlinear regression, both for nonstationary time series
- Waudby-Smith and Ramdas (2023) extend GCM test to the iid sequential setting. Anytime-valid test, requires distribution-uniform strong (a.s.) Gaussian approx.
  - Builds on previous work by Waudby-Smith, Arbour, et al. (2024) on asymptotic confidence sequences (AsympCSs)
- **Question 1:** can distribution-uniform anytime-valid inference be extended to nonstationary time series? Sequential GCM for nonstationary time series?
- **Question 2:** can we *sequentially* improve estimates of conditional functionals with nonstationary time series? Under what conditions? Can these be tested?

Sequential extensions, open questions, and future work

- Wieck-Sosa, Haddad, and Ramdas (2025) extend the *generalized covariance measure* (GCM) test from Shah and Peters (2020) to nonstationary time series
- **Key technical ingredients:** (1) distribution-uniform strong Gaussian approximation, and (2) nonlinear regression, both for nonstationary time series
- Waudby-Smith and Ramdas (2023) extend GCM test to the iid sequential setting. Anytime-valid test, requires distribution-uniform strong (a.s.) Gaussian approx.
  - Builds on previous work by Waudby-Smith, Arbour, et al. (2024) on asymptotic confidence sequences (AsympCSs)
- **Question 1:** can distribution-uniform anytime-valid inference be extended to nonstationary time series? Sequential GCM for nonstationary time series?
- **Question 2:** can we *sequentially* improve estimates of conditional functionals with nonstationary time series? Under what conditions? Can these be tested?
  - Ongoing work with A. Ramdas on invariant prediction

Thank you!

Questions and comments are welcome

You can reach me at: mwiecksosa@cmu.edu

Michael Wieck-Sosa
Department of Statistics & Data Science
Carnegie Mellon University

Example of nonstationary time series

- tvAR(1) process (time-varying autoregressive, one lag)
- Given parameter curves $\beta : [0,1] \to (-1,1)$ and $\sigma : [0,1] \to \mathbb{R}$ define

$$X_t^{(n)} = \beta\left(\frac{t}{n}\right) X_{t-1}^{(n)} + \sigma\left(\frac{t}{n}\right) \varepsilon_t, \quad t = 1, \ldots, n$$

  where $(\varepsilon_t)_t$ are iid random variables
- Extend β, σ to domain $(-\infty, 1]$ by setting $\beta(u) = \beta(0)$ and $\sigma(u) = \sigma(0)$ for $u < 0$
- tvMA(∞) representation (time-varying moving average, infinitely many lags)

$$X_t^{(n)} = \sum_{j=0}^{\infty} \phi_{t,j}^{(n)} \varepsilon_{t-j},$$

  where $\phi_{t,0}^{(n)} = \sigma\left(\frac{t}{n}\right)$ and for $j \geq 1$

$$\phi_{t,j}^{(n)} = \left(\prod_{r=0}^{j-1} \beta\left(\frac{t-r}{n}\right)\right) \sigma\left(\frac{t-j}{n}\right)$$

Nonstationarity and infill asymptotics

- When there is strong nonstationarity (i.e. not asymptotically stationary), we introduce rescaled time $u \in [0, 1]$ and use infill asymptotics so observations at t/n
- **Intuition:** more observations at each local structure of the process as n grows
- Recall: $(\varepsilon_i)_{i\in\mathbb{Z}}$ is an iid sequence of random variables, we denote $\boldsymbol{\varepsilon}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots)$, and Θ is a parameter space
- For each $n \in \mathbb{N}$, let $G_t^{(n)} : \mathbb{R}^\infty \times \Theta \to \mathbb{R}^d$, $t = 1, \ldots, n$, and $\tilde{G}_u^{(n)} : \mathbb{R}^\infty \times \Theta \to \mathbb{R}^d$, $u \in [0, 1]$, define a *generative model*

Assumption

The time series $X_{1:n}$ is generated by nature as

$$X_t = G_t^{(n)}(\boldsymbol{\varepsilon}_t, \theta) = \tilde{G}_{t/n}^{(n)}(\boldsymbol{\varepsilon}_t, \theta), \quad t = 1, \ldots, n$$

for some parameter $\theta \in \Theta$ and noise inputs $\boldsymbol{\varepsilon}_t$

Control of nonstationarity

- We measure the nonstationarity using a p-variation-type quantity, define

$$\left\| \left(\tilde{G}_u^{(i)}(\varepsilon_0, \theta) \right)_u \right\|_{p\text{-var}, \mathcal{L}^q(\theta)} = \sup_{\substack{0=u_0<\ldots<u_\ell=1 \\ \ell\in\mathbb{N}}} \left(\sum_{j=1}^{\ell} \left\| \tilde{G}_{u_j}^{(i)}(\varepsilon_0, \theta) - \tilde{G}_{u_{j-1}}^{(i)}(\varepsilon_0, \theta) \right\|_{\mathcal{L}^q(\theta)}^p \right)^{\frac{1}{p}}$$

- We impose the following regularity condition to control the nonstationarity

Assumption

For some $q > 2$, some $p \in [1, 2)$, and all $i \in \mathbb{N}$, there exists a constant $\Lambda > 0$ such that

$$\sup_{\theta\in\Theta} \sup_{u\in[0,1]} \left\| \tilde{G}_u^{(i)}(\varepsilon_0, \theta) \right\|_{\mathcal{L}^q(\theta)} + \sup_{\theta\in\Theta} \left\| \left(\tilde{G}_u^{(i)}(\varepsilon_0, \theta) \right)_u \right\|_{p\text{-var}, \mathcal{L}^q(\theta)} \leq \Lambda$$

Replication and Rosenblatt transform

- We can take the random variables $(\varepsilon_\ell)_{\ell \in \mathbb{Z}}$ to be $U[0,1]$
- Standard results in probability theory imply that the causal representations based on measurable functions of sequences of iid $U[0,1]$ are already sufficiently general
- First, by Kallenberg (2021) Lemma 4.21, Lemma 4.22, and the surrounding discussion, we can replicate each of the iid $U[0,1]$ random variables by taking each $G_t^{(n)}$ to include compositions with measurable functions to replicate the original $U[0,1]$, which yields d iid $U[0,1]$ random variables $U_1, \ldots, U_d$
- Second, we can use the Rosenblatt transform (i.e. for each $j \in [d]$, sample $X_{t,j}$ from the conditional quantile given $X_{t,1}, \ldots, X_{t,j-1}$ by transforming U_j), see Section 4 of Wu and Mielniczuk (2010)

Regression functions, errors, and residuals

- For each time t, we can always decompose the nonstationary time series as

$$X_t = f_t(Z_t) + \varepsilon_t, \quad Y_t = g_t(Z_t) + \xi_t$$

where $f_t(z) = \mathbb{E}(X_t \mid Z_t = z)$, $g_t(z) = \mathbb{E}(Y_t \mid Z_t = z)$ are regression functions at time t, and the errors $(\varepsilon_t)_t$, $(\xi)_t$ can be dependent and nonstationary

- Denote the vectorized outer product of the errors at time t by $R_t = \mathrm{vec}(\varepsilon_t \xi_t^\top)$
- Let $\hat{f}_t$ and $\hat{g}_t$ be estimates of f_t and g_t, respectively, and denote the residuals by

$$\hat{\varepsilon}_t = X_t - \hat{f}_t(Z_t), \quad \hat{\xi}_t = Y_t - \hat{g}_t(Z_t)$$

- Denote the vectorized outer product of the residuals at time t by $\hat{R}_t = \mathrm{vec}(\hat{\varepsilon}_t \hat{\xi}_t^\top)$

Regression estimators for nonstationary time series

- Recall: $\mathcal{P}_n^* \subset \mathcal{P}_n^{\mathrm{CI}}$ is collection of null distributions $\mathcal{P}_n^{\mathrm{CI}}$ s.t. *all assumptions hold*
- For all distributions $P \in \mathcal{P}_n^*$, we require the estimators of the regression functions to satisfy the following conditions:
  1. The **product** of $L^2(P)$ norms of the estimation errors must be $o\left(1/\sqrt{n}\ \mathrm{polylog}(n)\right)$
  2. Each $L^2(P)$ norm only needs to be $o\left(1/\mathrm{polylog}(n)\right)$
- The exact notation is given on the next slide
- Many regression estimators for nonstationary time series have been developed
- **Examples:** linear regression, kernel smoothing, sieve, boosting, neural networks
- Rescaled time $t/n \in [0, 1]$ is included as an argument along with the covariates

Required convergence rates for regression estimators

- Denote the estimation errors at time t by

$$\Delta^f_{P,t,i,a} = f_{P,t,i,a}(Z_t) - \hat{f}_{t,i,a}(Z_t), \quad \Delta^g_{P,t,j,b} = g_{P,t,j,b}(Z_t) - \hat{g}_{t,j,b}(Z_t)$$

- For each combination (i, j, a, b) of dimensions $i \in [d_X]$, $j \in [d_Y]$ and time-offsets $a \in A_i$, $b \in B_j$, under consideration, we have

$$\sup_{P \in \mathcal{P}^*_n} \sup_{i \in [d_X], a \in A_i} \sup_{t \in [n]} \mathbb{E}_P \left(\left| \Delta^f_{P,t,i,a} \right|^2 \right)^{1/2} = o\left(1/\mathrm{polylog}(n)\right)$$

$$\sup_{P \in \mathcal{P}^*_n} \sup_{j \in [d_Y], b \in B_j} \sup_{t \in [n]} \mathbb{E}_P \left(\left| \Delta^g_{P,t,j,b} \right|^2 \right)^{1/2} = o\left(1/\mathrm{polylog}(n)\right)$$

$$\sup_{P \in \mathcal{P}^*_n} \sup_{(i,j,a,b)} \sup_{t \in [n]} \mathbb{E}_P \left(\left| \Delta^f_{P,t,i,a} \right|^2 \right)^{1/2} \mathbb{E}_P \left(\left| \Delta^g_{P,t,j,b} \right|^2 \right)^{1/2} = o\left(1/\sqrt{n}\, \mathrm{polylog}(n)\right)$$

Main ideas of distribution-uniform strong Gaussian approximation

- Based on strong Gaussian approximation from Mies and Steland (2023)
- Suppose the aforementioned assumptions hold for some collection of distributions $\mathcal{P}_n^*$ for the time series $(W_t)_{t \in [n]}$
  1. Representation in terms of noise inputs
  2. Decay of physical dependence measure
  3. Control of nonstationarity
- **Main idea:** Uniformly over a collection of distributions, we can closely couple $\sqrt{n}$-scaled partial sums of $(W_t)_t$ and a nonstationary Gaussian process
- A more precise statement is given on the next slide, for when the constants for the decay of the physical dependence measure satisfy $\rho \geq 3$ and $q \gg 2$ (q large), and when the dimension $d \in \mathbb{N}$ is fixed

More details for distribution-uniform strong Gaussian approximation

- On a potentially enriched collection of probability spaces $(\Omega, \mathcal{F}, \mathbb{P}_P)_{P \in \mathcal{P}_n^*}$ indexed by the pushforward measures $P \in \mathcal{P}_n^*$, there exist random vectors $(W_t')_{t \in [n]}$ with the same distribution as $(W_t)_{t \in [n]}$ for each $P \in \mathcal{P}_n^*$, and *independent* Gaussian random vectors $(V_t')_{t \in [n]}$ where $V_t' \sim N(0, \Sigma_{P,t})$ for all $P \in \mathcal{P}_n^*$ and $t \in [n]$, s.t.

$$\sup_{P \in \mathcal{P}_n^*} \left(\mathbb{E}_P \max_{T \in [n]} \left\| \frac{1}{\sqrt{n}} \sum_{t \leq T} (W_t - V_t') \right\|^2 \right)^{\frac{1}{2}} = o\left(\frac{\mathrm{polylog}(n)}{n^{\frac{1}{6}}} \right)$$

where $\Sigma_{P,t}$ is the *local long-run* covariance matrix at time t defined by

$$\Sigma_{P,t} = \Sigma_{P_\theta,t}^{(n)} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}_{P_\theta}\left(G_t^{(n)}(\varepsilon_0, \theta), G_t^{(n)}(\varepsilon_h, \theta) \right)$$

Distributions of the stochastic system

- For all $\theta \in \Theta$ and $n \in \mathbb{N}$, denote by $P_{\theta,n}$ the distribution of $(G_t^{(n)}(\varepsilon_\ell, \theta))_{t\in[n],\ell\in\mathbb{Z}}$
- Distribution on $(\mathbb{R}^d)^{[n]\times\mathbb{Z}}$, i.e. the space of $\mathbb{R}^d$-valued functions of $[n] \times \mathbb{Z}$
- We observe the nonstationary time series $(G_t^{(n)}(\varepsilon_t, \theta))_{t\in[n]}$ at $\theta = \theta_0$
- At each fixed $t \in [n]$ and $\theta \in \Theta$, the process $(G_t^{(n)}(\varepsilon_\ell, \theta))_{\ell\in\mathbb{Z}}$ is stationary, for an example, see the time-varying autoregressive process example

References I

- Bodik, Juraj and Olivier C. Pasche (2024). "Granger causality in extremes". arXiv preprint arXiv: 2407.09632.
- Kallenberg, Olav (2021). *Foundations of modern probability*. Third edition. Springer.
- Mies, Fabian and Ansgar Steland (2023). "Sequential Gaussian approximation for nonstationary time series in high dimensions". In: *Bernoulli* 29.4, pp. 3114–3140.
- Shah, Rajen D. and Jonas Peters (2020). "The hardness of conditional independence testing and the generalised covariance measure". In: *Annals of Statistics* 48.3, pp. 1514–1538.
- Waudby-Smith, Ian, David Arbour, et al. (2024). "Time-uniform central limit theory and asymptotic confidence sequences". In: *Annals of Statistics* 52.6, pp. 2613–2640.
- Waudby-Smith, Ian and Aaditya Ramdas (2023). "Distribution-uniform anytime-valid inference". arXiv preprint arXiv:2311.03343.

References II

Wieck-Sosa, Michael, Michel F. C. Haddad, and Aaditya Ramdas (2025). "Conditional independence testing with a single realization of a multivariate nonstationary nonlinear time series". arXiv preprint arXiv:2504.21647.

Wu, Wei Biao (2005). "Nonlinear system theory: another look at dependence". In: *Proceedings of the National Academy of Sciences* 102.40, pp. 14150–14154.

Wu, Wei Biao and Jan Mielniczuk (2010). "A new look at measuring dependence". In: *Dependence in Probability and Statistics*, pp. 123–142.