Multiscale inference for nonparametric time trends

Marina Khismatullina   Michael Vogt

August 24, 2021

HCM Symposium

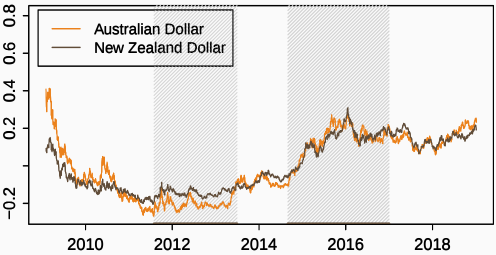
Table of contents

Introduction

Motivation

Aim of the paper

To develop new inference methods that allow to *identify* and *locate* differences between nonparametric trend curves with dependent errors.



Research question: Out of many given intervals, how to find those where the trends are significantly different?

Motivation

Why is it relevant?

Finding systematic differences between trends = basis for further research

Why is it difficult?

Testing many hypotheses at the same time = multiple testing problem

$\Rightarrow$ large probability of one true null hypothesis being rejected.

Is it limited to one application?

No! Our method = general method for comparing nonparametric trends

$\Rightarrow$ new statistical test for equality of nonparametric trend curves.

Literature

Comparison of deterministic trends:

- Park et al. (2009), Degras et al. (2012), Zhang et al. (2012), Hidalgo and Lee (2014), Chen and Wu (2019).

Multiscale tests:

- Chaudhuri and Marron (1999, 2000), Hall and Heckman (2000), Dümbgen and Spokoiny (2001), Park et al. (2009).

Comparison of volatility trends:

- Nyblom and Harvey (2000), ...

Model

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

We assume the following model:

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

where

- m_i are unknown trend functions on $[0, 1]$;
- $\boldsymbol{\beta}_i$ is $d \times 1$ vector of unknown parameters;
- α_i are so-called fixed effect error terms;
- $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary and causal error process.

Model, part 2

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

If we knew α_i and $\boldsymbol{\beta}_i$, then the model becomes much simpler:

$$\begin{aligned} Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} &=: Y_{it}^\circ \\ &= m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}. \end{aligned}$$

In reality the variables Y_{it}° are **not** observed.

But given $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$, we can consider

$$\widehat{Y}_{it} := Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \boldsymbol{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \big(\alpha_i - \widehat{\alpha}_i\big) + \varepsilon_{it}.$$

Model, part 3

1. We estimate $\boldsymbol{\beta}_i$:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^{\top} \Big)^{-1} \sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta Y_{it}$$

Theorem

Under certain regularity assumptions, $\widehat{\boldsymbol{\beta}}_i$ is a consistent estimator of $\boldsymbol{\beta}_i$ with the property $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P(T^{-1/2})$.

2. We estimate the fixed effects α_i:

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^{\top} \boldsymbol{X}_{it} \big)$$

We then work with the augmented time series $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^{\top} \boldsymbol{X}_{it}$.

Testing procedure

Testing problem

$$H_0 : m_1 = m_2 = \ldots = m_n$$

Question: if we reject the global null, how to locate the differences between the trends?

Consider a grid $\mathcal{G}_T = \{(u,h) : [u-h, u+h] \subseteq [0,1]\}$ of location-bandwidth parameters. For each pair (i,j) and for each interval $[u-h, u+h]$ we consider the null hypothesis

$$H_0^{[i,j]}(u,h) : m_i(w) = m_j(w) \text{ for all } w \in [u-h, u+h].$$

Then the global null $H_0 : m_1 = m_2 = \ldots = m_n$ can be reformulated as

$$H_0 : \text{The hypotheses } H_0^{[i,j]}(u,h) \text{ hold true for all intervals}$$
$$[u-h, u+h], (u,h) \in \mathcal{G}_T, \text{ and for all } 1 \leq i < j \leq n.$$

Test statistic

For a given location $u \in [0,1]$ and bandwidth h and a given pair (i,j) we construct the kernel averages

$$\widehat{\psi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big(\widehat{Y}_{it} - \widehat{Y}_{jt}\big),$$

where

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^{T} \Lambda_{t,T}^2(u,h)\}^{1/2}},$$

$$\Lambda_{t,T}(u,h) = K\Big(\frac{t/T-u}{h}\Big)\Big[S_{T,2}(u,h) - S_{T,1}(u,h)\Big(\frac{t/T-u}{h}\Big)\Big],$$

$$S_{T,\ell}(u,h) = \frac{1}{Th}\sum_{t=1}^{T} K\Big(\frac{t/T-u}{h}\Big)\Big(\frac{t/T-u}{h}\Big)^{\ell}$$

for $\ell = 1,2$ and K is a kernel function.

Test statistic, part 2

The kernel averages $\widehat{\psi}_{ij,T}(u,h)$ measure the distance between two trend curves m_i and m_j on $[u-h, u+h]$.

Instead with working directly with $\widehat{\psi}_{ij,T}(u,h)$, we replace them by

$$\widehat{\psi}^0_{ij,T}(u,h) = \left\{ \left| \frac{\widehat{\psi}_{ij,T}(u,h)}{\left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\widehat{\sigma}_i^2$ is an appropriate estimator of the long-run variance σ_i^2;
- $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ is an additive correction term (Dümbgen and Spokoiny (2001)). Explanation

Test statistic, part 3

To test the global null, we aggregate the individual test statistics for all (i, j) and all location-bandwidth pairs $(u, h) \in \mathcal{G}_T$:

$$\widehat{\Psi}_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\psi}^0_{ij,T}(u, h).$$

Note

Under certain conditions and under the null, $\widehat{\psi}^0_{ij,T}(u, h)$ and $\widehat{\Psi}_{n,T}$ can be approximated by the corresponding Gaussian versions of the test statistics.

Test procedure

Gaussian version of the individual test statistics:

$$\phi^0_{ij,T}(u,h) = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_T(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\phi_T(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \left\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \right\}$;
- Z_{it} are independent standard normal random variables;
- $\bar{Z}_i$ is the empirical average of $Z_{i1}, \ldots, Z_{iT}$.

Aggregated Gaussian test statistics:

$$\Phi_{n,T} = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \phi^0_{ij,T}(u,h).$$

Test procedure, part 2

1. Consider the Gaussian test statistic

$$\Phi_{n,T} = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \phi^0_{ij,T}(u,h),$$

   where ϕ^0_{ijk} are weighted averages of the differences of standard normal random variables.

2. Compute a $(1-\alpha)$-quantile $q_{n,T}(\alpha)$ of $\Phi_{n,T}$ by Monte Carlo simulations.
3. Perform the test for the global hypothesis H_0: reject H_0 if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$.
4. For each i, j, and each $(u,h) \in \mathcal{G}_T$, carry out the test for the local null hypothesis $H_0^{[i,j]}(u,h)$: reject $H_0^{[i,j]}(u,h)$ if $\widehat{\psi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)$.

Theoretical properties

Assumptions

$\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

$\mathcal{C}2$ For each i the variables ε_{it} are weakly dependent. Details

$\mathcal{C}3$ For each i we have that $\boldsymbol{X}_{it}$ is stationary and causal with all the necessary moments and no asymptotic multicollinearity.

$\mathcal{C}4$ For each i the variables $\boldsymbol{X}_{it}$ are weakly dependent. Details

$\mathcal{C}5$ $\boldsymbol{X}_{it}$ (elementwise) and ε_{is} are uncorrelated for each t, s.

$\mathcal{C}6$ All of the variables in the model are short-range dependent. Details

Assumptions, part 2

$\mathcal{C}7$ Standard assumptions on the kernel function K.

$\mathcal{C}8$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

$$\mathcal{G}_T = \big\{(u,h) : u = t/T \text{ for some } 1 \le t \le T \text{ and } h \in [h_{\min}, h_{\max}] \\ \text{with } h = t/T \text{ for some } 1 \le t \le T\big\},$$

$\mathcal{C}9$ $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$ and $h_{\max} < 1/2$.

$\mathcal{C}10$ Assume that $\sigma_i^2 = \sigma_j^2$ for all i,j and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$.

Theoretical properties

Proposition

Let $\mathcal{M}_0$ be the set of triplets (i,j,k) for which $H_0^{(ijk)}$ holds true. Then under $\mathcal{C}1-\mathcal{C}6$, it holds that

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk}| \leq c_{\text{Gauss}}(\alpha, h_k)\Big) \geq 1-\alpha+o(1)$$

Proposition

Consider a sequence of functions $\lambda_i = \lambda_{i,T}$, $\lambda_j = \lambda_{j,T}$ such that

$$\exists \mathcal{I}_k : \lambda_i(w) - \lambda_j(w) \geq c_T \sqrt{\log T/(Th_k)}\ \forall w \in \mathcal{I}_k, \tag{1}$$

and $c_T \to \infty$ faster than $\frac{\sqrt{\log T}\sqrt{\log\log T}}{\log\log\log T}$. Let $\mathcal{M}_1$ be the set of triplets (i,j,k) for which (1) holds true. Then under $\mathcal{C}1-\mathcal{C}6$, it holds that

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_1 : |\hat{\psi}_{ijk}| > c_{\text{Gauss}}(\alpha, h_k)\Big) = 1-o(1)$$

Strategy of the proof

- Replace the statistic $\widehat{\Psi}_T$ under $H_0 : m = 0$ by a statistic $\widetilde{\Phi}_T$ with the same distribution and the property that

$$\left|\widetilde{\Phi}_T - \Phi_T\right| = o_p(\delta_T),$$

  where $\delta_T = o(1)$. To do so, we make use of strong approximation theory for dependent processes as derived in Berkes et al. (2014)

- Using the anti-concentration results for Gaussian random vectors (Chernozhukov et al. 2015), prove that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$, i.e.

$$\sup_{x \in \mathbb{R}} \mathrm{P}\Big(|\Phi_T - x| \le \delta_T\Big) = o(1).$$

- Show that

$$\sup_{x \in \mathbb{R}} \Big|\mathrm{P}(\widetilde{\Phi}_T \le x) - \mathrm{P}(\Phi_T \le x)\Big| = o(1).$$

Illustration

Graphical representation

How to represent the results of the test?

Plot the results of pairwise comparison $\mathcal{F}_{\text{reject}}(i,j)$:

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : \mathcal{I}_k \notin \mathcal{F}_{\text{reject}}(i,j)\Big) \geq 1 - \alpha + o(1)$$

Minimal intervals

An interval $\mathcal{I}_k \in \mathcal{F}_{\text{reject}}(i,j)$ is called **minimal** if there is no other interval $\mathcal{I}_{k'} \in \mathcal{F}_{\text{reject}}(i,j)$ with $\mathcal{I}_{k'} \subset \mathcal{I}_k$.

The set of minimal intervals is denoted $\mathcal{F}^{\text{min}}_{\text{reject}}(i,j)$.

We can make similar confidence statements about minimal intervals:

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : \mathcal{I}_k \notin \mathcal{F}^{\text{min}}_{\text{reject}}(i,j)\Big) \geq 1 - \alpha + o(1)$$

Discussion

We can claim, with confidence of about 95%, that the null hypothesis is violated for all intervals (and all pairs of time series) for which our test rejects the null.

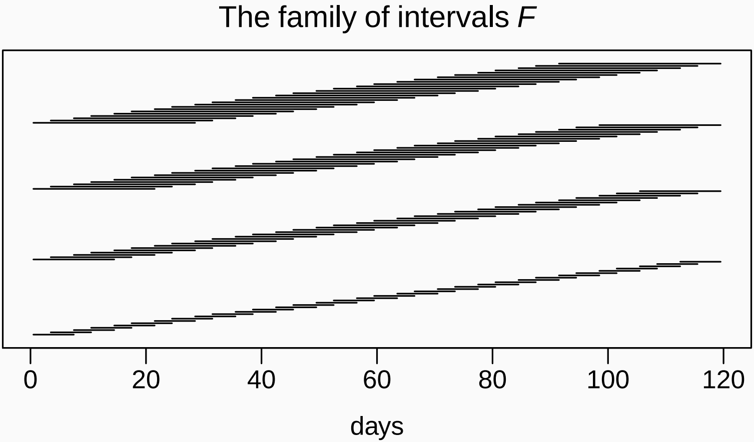
However, we can not say anything about the causes of such differences. This question requires further (probably not purely statistical) analysis.

Further possible extensions:

- introduce scaling factor in the trend function, that will allow to adjust for the size of the country (population, density, testing regimes, etc.);
- include the dependence between covariates and error terms;
- cluster the time series based on the trends they exhibit.

Thank you!

Family of time intervals



Simulation results for the size of the test

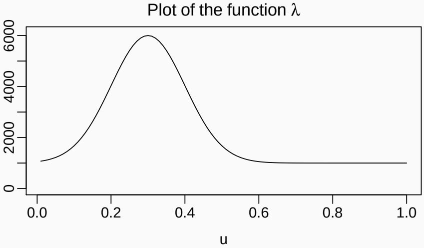
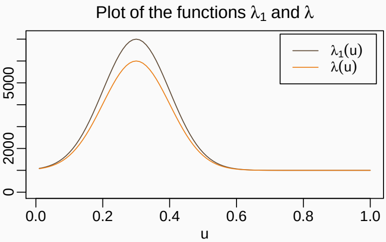


Table 1: Size of the multiscale test

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.011	0.047	0.093	0.010	0.044	0.087	0.008	0.037	0.075
$T = 250$	0.009	0.047	0.091	0.009	0.046	0.087	0.008	0.035	0.069
$T = 500$	0.010	0.044	0.083	0.008	0.048	0.093	0.007	0.035	0.077

Simulation results for the power of the test



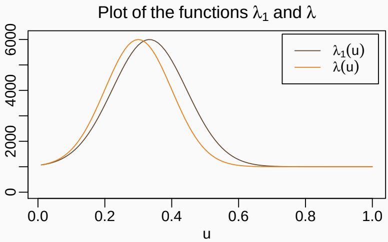


Table [illegible]: Power of the multiscale test for scenario [illegible]

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
$T = 250$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
$T = 500$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Estimator of σ^2

We estimate the overdispersion paramter σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_i^2 \text{ and } \hat{\sigma}_i^2 = \frac{\sum_{t=2}^{T}(X_{it} - X_{it-1})^2}{2\sum_{t=1}^{T} X_{it}}$$

We assume that λ_i is Lipschitz continuous. Then

$$X_{it} - X_{it-1} = \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}(\eta_{it} - \eta_{it-1}) + r_{it},$$

where $|r_{it}| \leq C(1 + |\eta_{it-1}|)/T$ with a sufficiently large C. Hence,

$$\frac{1}{T}\sum_{t=2}^{T}(X_{it} - X_{it-1})^2 = 2\sigma^2\Big\{\frac{1}{T}\sum_{t=2}^{T}\lambda_i(t/T)\Big\} + o_p(1)$$

Together with

$$\frac{1}{T}\sum_{t=1}^{T}X_{it} = \frac{1}{T}\sum_{t=1}^{T}\lambda_i(t/T) + o_p(1),$$

we get that $\hat{\sigma}_i^2 = \sigma^2 + o_p(1)$ for any i and thus $\hat{\sigma}^2 = \sigma^2 + o_p(1)$. Go back

Idea behind a_k and b_k

Dümbgen and Spokoiny (2001): the critical values $c_{ijk}(\alpha)$ depend on the scale of the testing problem, i.e. the length h_k of the time interval.

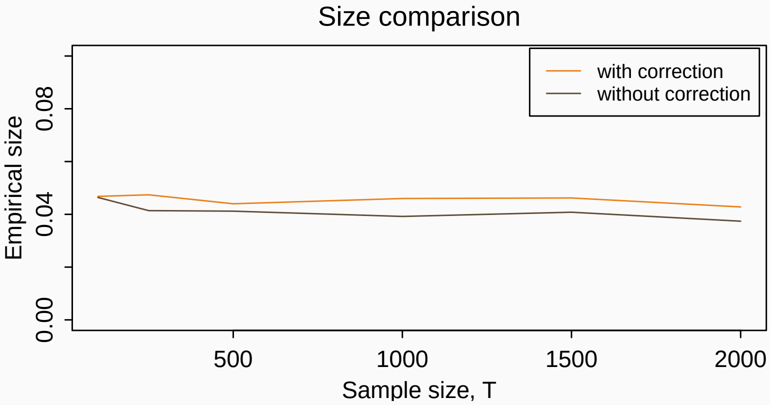
Specifically,

$$c_{ijk}(\alpha) = c(\alpha, h_k) := b_k + q(\alpha)/a_k,$$

where $a_k = \{\log(e/h_k)\}^{1/2} / \log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants and $q(\alpha)$ is chosen such that we control FWER.

Idea behind a_k and b_k, part 2

This choice of scale-dependent constants helps us balance the significance of hypotheses between the time intervals of different lengths h_k:



Go back

Idea behind the additive correction

Consider the uncorrected Gaussian statistic

$$\Phi^{\text{uncor}} = \max_{(i,j,k)} |\phi_{ijk}|$$

and let the family of intervals be

$$\mathcal{F} = \big\{[(m-1)h_l, mh_l] \text{ for } 1 \leq m \leq 1/h_l, 1 \leq l \leq L\big\}$$

Then we can rewrite the uncorrected test statistic as

$$\Phi^{\text{uncor}} = \max_{i,j} \max_{\substack{1 \leq l \leq L, \\ 1 \leq m \leq 1/h_l}} \left| \frac{1}{\sqrt{2Th_l}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in [(m-1)h_l, mh_l]\Big)(Z_{it} - Z_{jt}) \right|$$

$\Rightarrow \quad \max_m \ldots = \sqrt{2\log(1/h_l)} + o_P(1) \to \infty$ as $h \to 0$ and the stochastic behavior of Φ^{uncor} is dominated by the elements with small bandwidths h_l. Go back

Dependence measure

Following Wu (2005), we define the *physical dependence measure* for the process $\boldsymbol{L}(\mathcal{F}_t)$ as the following:

$$\delta_q(\boldsymbol{L}, t) = ||\boldsymbol{L}(\mathcal{F}_t) - \boldsymbol{L}(\mathcal{F}_t')||_q,$$

where $\mathcal{F}_t = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ and $\mathcal{F}_t' = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ is a coupled process of $\mathcal{F}_t$ with ϵ_0' being an i.i.d. copy of ϵ_0.

Intuitively, $\delta_q(\boldsymbol{L}, t)$ measures the dependency of $\boldsymbol{L}(\mathcal{F}_t)$ on ϵ_0, i.e., how replacing ϵ_0 by an i.i.d. copy while keeping all other innovations in place affects the output $\boldsymbol{L}(\mathcal{F}_t)$.

Technical assumptions

- $\mathcal{C}1'$ The variables ε_{it} allow for the representation $\varepsilon_{it} = G_i(\ldots, \eta_{it-1}, \eta_{it})$, where η_{it} are i.i.d. random variables across t and $G_i : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. Denote $\mathcal{J}_{it} = (\ldots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$.
- $\mathcal{C}1'''$ Define $\Theta_{i,t,q} = \sum_{s \geq t} \delta_q(G_i, s)$ for $t \geq 0$. For each i it holds that $\Theta_{i,t,q} = O(t^{-\tau_q}(\log t)^{-A})$, where $A > \frac{2}{3}(1/q + 1 + \tau_q)$ and $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$. Go back

Technical assumptions, part 2

- $\mathcal{C}3'$ $\boldsymbol{X}_{it}$ allow for the representation $\boldsymbol{X}_{it} = \boldsymbol{H}_i(\ldots, u_{it-1}, u_{it})$ with u_{it} being i.i.d. random variables and $\boldsymbol{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^\top : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^d$ being a measurable function such that $\boldsymbol{H}_i(\mathcal{U}_{it})$ is well defined. Denote $\mathcal{U}_{it} = (\ldots, u_{it-1}, u_{it})$.
- $\mathcal{C}3''$ Let $\boldsymbol{N}_i$ be the $d \times d$ matrix with $n_{i,kl} = \mathbb{E}[H_{ik}(\mathcal{U}_{i0})H_{il}(\mathcal{U}_{i0})]$ being kl-th entry. We assume that the smallest eigenvalue of $\boldsymbol{N}_i$ is strictly bigger than 0.
- $\mathcal{C}3'''$ Let $\mathbb{E}[\mathrm{H}_i(\mathcal{U}_{i0})] = 0$ and $||\mathrm{H}_i(\mathcal{U}_{it})||_{q'} < \infty$ for some $q' > \max\{2\theta, 4\}$, where θ will be introduced further.
- $\mathcal{C}4'$ $\sum_{s=0}^{\infty} \delta_{q'}(\mathrm{H}_i, s) < \infty$ for q' from Assumption $\mathcal{C}3'''$.
- $\mathcal{C}4''$ For each i it holds that $\sum_{s=t}^{\infty} \delta_{q'}(\mathrm{H}_i, s) = O(t^{-\alpha})$ for q' from Assumption $\mathcal{C}3'''$ and for some $\alpha > 1/2 - 1/q'$. Go back

Technical assumptions, part 3

$\mathcal{C}6$ Let $\zeta_{i,t} = (u_{it}, \eta_{it})^\top$. Define $\mathcal{I}_{it} = (\ldots, \zeta_{i,t-1}, \zeta_{i,t})$ and $\mathsf{U}_i(\mathcal{I}_{it}) = \mathsf{H}_i(\mathcal{U}_{it}) G_i(\mathcal{J}_{it})$. With this notation at hand, we assume that $\sum_{s=0}^{\infty} \delta_2(\mathsf{U}_i, s) < \infty$. Go back