

Multiscale Comparison of Nonparametric Trending Coefficients

Marina Khismatullina* Bernhard van der Sluis†

November 18, 2025

Abstract

This paper proposes a novel framework to test for slope heterogeneity between time-varying coefficients in panel data models. Our test not only allows us to detect whether the coefficient functions are the same across all units or not, but also determines which of them are different and where these differences are located. We establish the asymptotic validity of our multiscale test. As an extension of the proposed procedure, we show how to use the results to uncover latent group structures in the model. We apply our methods to test for heterogeneity in the effect of U.S. monetary shocks on 49 foreign economies and itself. We find evidence that such heterogeneity indeed exists and we discuss the clustering results for two groups.

Keywords: panel data, nonparametric regression, multiscale statistics, strong approximation.

JEL Classification: C12, C14, C23, C38

*E-mail: khismatullina@ese.eur.nl. Erasmus School of Economics, Erasmus University Rotterdam

†E-mail: vandersluis@ese.eur.nl. Erasmus School of Economics, Erasmus University Rotterdam

1 Introduction

The use of panel data is a fundamental tool in empirical research across various disciplines, including economics, finance, and climatology. The double-index structure of panel data allows researchers to capture both cross-sectional heterogeneity and time dynamics, making it indispensable for analyzing complex systems. However, the majority of panel data models assume time-invariant and/or common coefficients, assumptions that are often unrealistic in practical applications. Structural changes, policy interventions, and macroeconomic shocks frequently induce time variation in coefficients, and failure to capture such variation can lead to misleading conclusions. Similarly, cross-sectional units may respond differently to covariates. For instance, in finance, the sensitivity of asset returns to market return is often assumed to fluctuate over time (Chen, 1982; Jagannathan and Wang, 1996; Lewellen and Nagel, 2006). Moreover, investors are interested in the different risk exposures across assets for the same predictors (Fama and MacBeth, 1973; Adrian and Franzoni, 2009). In environmental research, the impact of climatic factors varies across regions and time (Lobell et al., 2014; Keane and Neal, 2020). These examples highlight the importance of developing statistical tools to analyze heterogeneous time-varying coefficients in panel data.

In this paper, we propose a new multiscale test that allows us to assess heterogeneity in time-varying regression coefficients in panel data models. Unlike traditional global hypothesis tests, which provide no insight into localized differences, our method offers a more detailed analysis by detecting both the presence of coefficient heterogeneity and its specific locations and time intervals. This level of detail makes the test particularly useful in practical applications, where understanding the specific structure of heterogeneity is crucial for policy making, risk management, and forecasting.

While several methods have been developed to allow for time-varying coefficients in panel data, they typically focus on global tests that do not identify where or when differences arise. Currently there exists limited theoretical guidance on (i) testing whether time-varying coefficients are all the same across units, and (ii) identifying specific time periods and units where coefficients differ. Existing methods, such as those proposed by Atak et al. (2025), allow researchers to test for global slope homogeneity, but do not provide localized information where these differences occur. However, simply detecting heterogeneity is often insufficient: practitioners need to understand which time periods and which groups of units drive these differences. Our proposed approach addresses this gap directly.

The main theoretical contribution of this paper is establishing that under mild regularity conditions, the proposed multiscale test accounts for the multiple testing problem inherent in this setting by controlling the family-wise error rate. Building on the strong approximation results for random vectors developed in Karmakar and Wu (2020), this result ensures the validity of simultaneous confidence statements about localized differ-

ences in trending coefficients. This extends previous work on the comparison of non-parametric time trends by Khismatullina and Vogt (2022, 2023) to the time-varying coefficient models.

This paper contributes to several strands of literature. First, the assumption of homogeneity of slope coefficients, whether they are assumed to be time-varying or not, is crucial for valid inference. If the coefficients are truly homogeneous and the individual heterogeneity is captured by fixed effects, cross-sectional data can be pooled to estimate coefficients more efficiently. However, if heterogeneity exists in the coefficients, cross-sectional averaging results in misleading inference (Hsiao and Tahmiscioglu, 1997; Baltagi et al., 2008; Wang et al., 2019). Several methods have been developed to address this issue, both for testing for homogeneity in panel data models with constant slope (Pesaran and Yamagata, 2008; Gao et al., 2020; Su and Chen, 2013; Ando and Bai, 2015) and with time-varying coefficients (Atak et al., 2025). However, as mentioned before, these approaches were developed for testing a global null hypothesis that states that the coefficients are the same for *all* units at *all* times, whereas our method provides additional information in case of slope heterogeneity. Specifically, by considering a multitude of localized null hypotheses instead of one global null, we are able to say when and where this heterogeneity exactly occurs.

Second, we extend prior work on multiscale testing to a panel data setting with time-trending coefficients. Multiscale tests were developed for various models and null hypotheses. For example, Khismatullina and Vogt (2020) propose a new multiscale approach to test for monotonicity of deterministic trend functions in a single nonparametric regression. This framework was later extended to multiple time series by Khismatullina and Vogt (2022) and Khismatullina and Vogt (2023), where the authors devise procedures to test for homogeneity of trend functions under different assumptions. In this paper, we build upon the theoretical work in the above studies to consider time-varying coefficient functions. Similarly as before, the objects of interest are functions of time that are estimated nonparametrically and compared on a wide range of scales, i.e., using various bandwidths. This is the essence of a multiscale approach, and this approach allows us to provide information on top of simply rejecting the null hypothesis of a global slope homogeneity.

Furthermore, in our paper, we propose a clustering algorithm that identifies groups of units exhibiting similar time trends. Specifically, in case of detected heterogeneity among the coefficient functions, it is still possible to allow for grouped patterns of heterogeneity that can be a result of similar underlying characteristics across units. Several approaches exist to identify the unobserved group structure. In the seminal paper of Bonhomme and Manresa (2015), a k -means algorithm is used to cluster the individual observations in a linear panel data model with group-specific fixed effects. This framework has been extended to time-varying slope coefficients, but the extensions are mostly limited to structural breaks (Okui and Wang, 2021; Lumsdaine et al.,

2023). Alternatively, Su et al. (2016) propose a classifier-Lasso simultaneous estimator of heterogeneous coefficients and group membership in linear and non-linear panel data models and Su et al. (2019) use the classifier-Lasso in a panel data model with group-specific time-varying coefficients. Their approach requires a sieve approximation of the time-varying coefficients together with a penalized least-squares objective function. Recently, this work has been extended by Haimerl et al. (2025) who instead of the classifier-Lasso use a pairwise adaptive group fused-Lasso to uncover a hidden group structure. In contrast, our estimator of the time-varying coefficients has a closed-form solution, the multiscale test statistics are easy to compute, and the clustering procedure is a straightforward application of an easy-to-track hierarchical agglomerative clustering algorithm. The most similar approach to ours was taken in Chen (2019) where the latent group structure in heterogeneous time-varying-coefficient panels is recovered via kernel estimation, hierarchical agglomerative clustering, and a generalized information criterion for the number of groups. In contrast to imposing group structure ex ante and estimating it directly as in Chen (2019), our framework treats grouping as a downstream step: test-implied dissimilarities feed into HAC clustering, yielding both a global decision on homogeneity and time-resolved evidence of heterogeneity.

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 presents our multiscale testing procedure, and Section 4 establishes its theoretical properties. In Section 5, we discuss how our method can be extended to perform clustering and uncover unknown group structure in the data. In Section 6, we apply our methods to an empirical dataset revisiting the relationship between U.S. monetary shock and 49 foreign economies and U.S. GDP from Iacoviello and Navarro (2019). Section 7 concludes. Additional proofs and supplementary results are provided in Appendix.

Finally, some words on notation. For a vector $\mathbf{v} = (v_1, \dots, v_D) \in \mathbb{R}^D$, we write $\|\mathbf{v}\| = \left(\sum_{i=1}^D v_i^2\right)^{1/2}$ and $\|\mathbf{v}\|_\infty = \max\{|v_1|, |v_2|, \dots, |v_D|\}$ to denote its Euclidean and maximum (ℓ^∞) norm, respectively. Similarly, for a random vector $\mathbf{V}$, we denote its $\mathcal{L}^q$ -norm for $q > 0$ as $\|\mathbf{V}\|_{\mathcal{L}^q} = (\mathbb{E}\|\mathbf{V}\|^q)^{1/q}$ (i.e., we always use the Euclidean base-norm). For a $D \times D$ square matrix $\mathbf{A} = (a_{d_1, d_2})_{1 \leq d_1, d_2 \leq D}$ we denote induced norms $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_1$ as $\|\mathbf{A}\| = \sup \left\{ \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} : \mathbf{x} \in \mathbb{R}^D, \mathbf{x} \neq \mathbf{0} \right\}$ and $\|\mathbf{A}\|_1 = \max_{d_2=1, \dots, D} \sum_{d_1=1}^D |a_{d_1, d_2}|$, respectively. If $\mathbf{A}$ is positive semi-definite with eigenvalues $\lambda_1 \geq \dots \geq \lambda_D$, we write $\rho_*(\mathbf{A}) = \lambda_D$. For a set $\mathcal{S}$, we denote its cardinality as $|\mathcal{S}|$. For a scalar x , we denote the greatest integer that is less or equal to x by $\lfloor x \rfloor$. Throughout the paper, C denotes a universal constant that may take different values on different lines.

Furthermore, we introduce a general class of potentially nonlinear time series that can be defined as follows. Let η_t for $t \in \mathbb{Z}$ be i.i.d. random variables. Define $\mathcal{F}_t = (\dots, \eta_{t-1}, \eta_t)$ and let $g : \mathbb{R}^\infty \rightarrow \mathbb{R}^D$ be a measurable function such that the random vector $g(\mathcal{F}_t)$ is well-defined. Following Wu (2005) and later Karmakar and Wu (2020), we define the

physical dependence measure of the process g as

$$\delta_q(g, t) = \|g(\mathcal{F}_t) - g(\mathcal{F}'_t)\|_{\mathcal{L}^q}, \quad (1)$$

where $\mathcal{F}'_t = (\dots, \eta_{-1}, \eta'_0, \eta_1, \dots, \eta_{t-1}, \eta_t)$ with η'_0 being an i.i.d. copy of η_0 . The quantity $\delta_q(g, t)$ measures the t -th lag dependence of the process g in terms of q -th moment. For a detailed explanation of the physical dependence measure and its properties, we refer the reader to Wu (2005).

2 The model framework

In this section, we specify the model we consider in our paper and describe the preliminary step that are needed before constructing the test statistics.

2.1 The model

Suppose that we observe a panel of N time series $\mathcal{T}_i = \{(Y_{it}, \mathbf{X}'_t) : t = 1, \dots, T\}$, of length T where the responses Y_{it} are real-valued random variables and the predictors $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,D})' \in \mathcal{X} \subset \mathbb{R}^D$ are random vectors of covariates common for all time series (which holds, for instance, in factor models). For the moment, we suppose that the set $\mathcal{X}$ is compact and $T \rightarrow \infty$, while N and D remain fixed.

We assume that each of the time series $\mathcal{T}_i$ follows a heterogeneous time-varying coefficients model:

$$Y_{it} = \mathbf{X}'_t \boldsymbol{\beta}_i(t/T) + \varepsilon_{it}, \quad t = 1, \dots, T, \quad (2)$$

where ε_{it} are idiosyncratic error terms which can potentially be serially correlated across t but are independent across i and $\boldsymbol{\beta}_i = (\beta_{i,1}, \dots, \beta_{i,D})' : [0, 1] \rightarrow \mathbb{R}^D$ are unknown vector-valued coefficient functions that may be different among the units. Technical assumptions on the dependency structure in model (2) are discussed later in Section 4. The main goal of this paper is to develop a rigorous statistical method that allows us to compare the unknown functions $\boldsymbol{\beta}_i$ across units, while providing information on where and when these differences occur. In what follows, we suppose that the functions $\boldsymbol{\beta}_i$ take the rescaled time t/T as an argument which is common in the literature for nonparametric estimation of time-varying parameters (e.g., Cai, 2007; Li et al., 2011). Such rescaling means that, as the time dimension increases, we collect more and more data locally which ensures consistency of the nonparametric estimator that is used for our testing procedure.

2.2 Preliminary steps

We assume that β_i are sufficiently smooth coefficient functions satisfying Lipschitz continuity conditions. Formally, it can be formulated as follows:

- (A1) For each $i = 1, \dots, N$, the trend coefficient functions β_i are Lipschitz continuous on $[0, 1]$ coordinate-wise, i.e., for all $d = 1, \dots, D$, $|\beta_{i,d}(v) - \beta_{i,d}(w)| \leq C|v - w|$ for all $v, w \in [0, 1]$.

We estimate β_i using a local constant regression approach. Specifically, for each fixed time point $u \in [0, 1]$, we minimize the locally weighted least squares criterion:

$$\sum_{t=1}^T (Y_{it} - \mathbf{X}_t' \beta_i(u))^2 K\left(\frac{t/T - u}{h}\right). \quad (3)$$

Here, K denotes a kernel function and h is a bandwidth that controls the smoothness of the estimator. We assume that $h \in [h_T^{\min}, h_T^{\max}]$ with formal requirements on $h_T^{\min}$ and $h_T^{\max}$ stated further. Minimizing (3) with respect to β_i results in the well-known Nadaraya-Watson estimator that has the following form:

$$\hat{\beta}_i(u, h) = \left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' K\left(\frac{t/T - u}{h}\right) \right]^{-1} \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t Y_{it} K\left(\frac{t/T - u}{h}\right). \quad (4)$$

Alternatively, we can estimate the unknown slope coefficients by a local linear approximation (e.g., Li et al., 2011; Chen and Hong, 2012) or a sieve approximation (e.g., Zhang and Zhou, 2021; Atak et al., 2025). However, the local constant approach allows us to use strong approximation theory for the errors ε_{it} which guarantees valid critical values for our testing procedure.

In the rest of the paper, we employ the following shorthanded notation: $K_{t,u,h} := K((t/T - u)/h)$ for the value of the kernel function K at $(t/T - u)/h$ and $\mathbf{M}_{XKX}(u, h) := \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' K_{t,u,h}$ for the (localized) design matrix. Using this notation, we can rewrite the estimator in (4) as

$$\hat{\beta}_i(u, h) = \mathbf{M}_{XKX}^{-1}(u, h) \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t Y_{it} K_{t,u,h} \right).$$

Remark 2.1. *Introducing time-varying coefficient functions in (2) generalizes the model in Khismatullina and Vogt (2022), which has constant coefficients β_i and a deterministic trend function. The authors provide a multiscale test for (local) differences in the trend functions. We allow all coefficients to be time-varying, where $X_{t,1}$ does not necessarily need to be constant.*

3 Testing procedure

This section presents the multiscale testing procedure for detecting heterogeneity in regression coefficients. The primary objective is to assess whether the time-varying coefficient functions β_i in model (2) are the same across all time series. Unlike conventional tests that focus on global comparison, our approach captures localized variations by testing over multiple time scales simultaneously.

3.1 Hypothesis formulation

We formally define the null hypothesis as:

$$H_0 : \beta_1(u) = \beta_2(u) = \dots = \beta_N(u) \quad \forall u \in [0, 1], \quad (5)$$

which states that the coefficient functions β_i are everywhere identical across all units $i = 1, \dots, N$. Under H_0 , model (2) simplifies to a homogeneous linear panel data model. In this setting, pooling data across units yields an efficient estimator of the common slope function. However, when H_0 is violated, pooling can lead to biased estimates and incorrect inference due to underlying coefficient heterogeneity (see, e.g., Hsiao and Tahmiscioglu, 1997). Testing H_0 is therefore critical for ensuring valid estimation and inference in time-varying panel models.

Several approaches exist for testing H_0 . Traditional parametric methods, such as the Wald test and the likelihood ratio test, compare restricted and unrestricted estimators. More recently, Atak et al. (2025) proposed a nonparametric test based on residuals obtained from the estimation procedure under H_0 . However, while these tests are helpful in detecting slope heterogeneity in general, they mostly focus on global departures from homogeneity and do not identify where and when coefficient differences occur.

In this section, we describe a multiscale testing framework that provides richer insights than standard methods. Following the perspective of Khismatullina and Vogt (2020, 2022, 2023), our approach is designed to answer three key questions:

1. Is the null hypothesis H_0 violated? That is, does at least one coefficient function differ somewhere across units?
2. If the null hypothesis H_0 is not true, then which coefficient functions differ?
3. At which time intervals do these differences arise?

To achieve this, we introduce a series of local null hypotheses, where "local" refers to a certain pair of units and a specific time intervals rather than the entire time domain. We then construct test statistics that allow for separate evaluation of each local null hypothesis while accounting for multiple testing problem at the same time. The core idea is to aggregate the conclusions from these local tests into a unified confidence

statement, providing a statistically rigorous framework for detecting heterogeneity in time-varying coefficients.

We proceed as follows. For any interval $[u - h, u + h] \subseteq [0, 1]$ (where u denotes the location and h the bandwidth) and for any pair of units $1 \leq i < j \leq N$, consider the local null hypothesis

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

that states that the coefficient functions of units i and j are identical on the interval $[u - h, u + h]$. Reformulating the global null hypothesis in terms of these local hypotheses gives:

$$H_0 : \text{The hypotheses } H_0^{[i,j]}(u, h) \text{ hold true for all intervals } [u - h, u + h] \subseteq [0, 1] \\ \text{and for all } 1 \leq i < j \leq N.$$

Thus, instead of testing a single global null hypothesis H_0 , our method assesses a collection of local ones, enabling us to identify specific deviations in coefficient functions both across units and over time. Moreover, the multiscale perspective ensures that our test is sensitive to both long-term trends and short-term structural changes.

3.2 Construction of the test statistic

The main idea of our multiscale method is to test simultaneously hypotheses $H_0^{[i,j]}(u, h)$ for all intervals $[u - h, u + h] \subseteq [0, 1]$ and all pairs (i, j) with $1 \leq i < j \leq N$. In practice, we are unable to consider all possible intervals $[u - h, u + h] \subseteq [0, 1]$, and we need to work with a countable subset of these intervals. We consider the following set. We define a collection of location-bandwidth points $\mathcal{G}_T = \{(u, h) : [u - h, u + h] \subseteq [0, 1] \text{ with } u = t/T, h = s/T \text{ for some } t, s = 1, \dots, T \text{ and } h \in [h_T^{\min}, h_T^{\max}]\}$. Here, as before, $h_T^{\min}$ and $h_T^{\max}$ denote the minimal and maximal bandwidths, respectively, with specific requirements on their selection provided below in Assumption (A12). The set $\mathcal{G}_T$ is chosen to be sufficiently rich so that the resulting intervals collectively cover the entire unit interval. We allow $\mathcal{G}_T$ to be quite large compared to the length of the time series T . Formal assumptions on $\mathcal{G}_T$, $h_T^{\min}$ and $h_T^{\max}$ are given in Section 4.1.

Now we are ready to construct the localized pairwise test statistics $\hat{S}_{ij}(u, h)$. For a given pair $(u, h) \in \mathcal{G}_T$ and all $i = 1, \dots, N$, we estimate $\beta_i(u)$ using the nonparametric kernel estimator $\hat{\beta}_i(u, h)$ as defined in (4). Based on these estimates, for each pair (i, j) , $1 \leq i < j \leq N$ we consider the following quantity:

$$\hat{S}_{ij}(u, h) = \left\| \hat{\Sigma}_{ij}^{-1/2} \mathbf{M}_{XX}(u, h) \left(\hat{\beta}_i(u, h) - \hat{\beta}_j(u, h) \right) \right\|_{\infty},$$

where $\hat{\Sigma}_{ij} = (\hat{\Sigma}_i + \hat{\Sigma}_j)/2$ with $\hat{\Sigma}_i$ and $\hat{\Sigma}_j$ represent appropriate estimators of the

long-run covariance matrices Σ_i and Σ_j of the processes $\{\mathbf{X}_{t\varepsilon_{it}}\}_{t=1}^T$ and $\{\mathbf{X}_{t\varepsilon_{jt}}\}_{t=1}^T$, respectively. Our procedure requires an estimator $\hat{\Sigma}_i$ to have the rate of convergence $\rho_T = o(\log T)$ which is easily achieved using most standard estimators, for example, HAC estimator as proposed in Andrews (1991). Discussion of the estimation procedure for Σ_i that we follows in the simulation study and in the empirical analysis is provided in Section 3.5.

Heuristically, the statistic $\hat{S}_{ij}(u, h)$ approximates the distance between β_i and β_j over the interval $[u-h, u+h]$. To ensure scale invariance, we normalize the test statistic using the (localized) variance of $\mathbf{X}_t$ through $\mathbf{M}_{XX}(u, h)$ and the estimated variances of the processes $\{\mathbf{X}_{t\varepsilon_{it}}\}_{t=1}^T$ and $\{\mathbf{X}_{t\varepsilon_{jt}}\}_{t=1}^T$. By construction, $\hat{S}_{ij}(u, h)$ is always nonnegative, with smaller values indicating that the difference between β_i and β_j over the given interval is quite small. In contrast, large positive values of $\hat{S}_{ij}(u, h)$ suggest substantial differences, making the rejection of the local null hypothesis $H_0^{[i,j]}(u, h)$ more likely.

To test the global null hypothesis H_0 , we aggregate the local statistics $\hat{S}_{ij}(u, h)$ over all location-bandwidth points $(u, h) \in \mathcal{G}_T$ and all pairs $(i, j), 1 \leq i < j \leq N$, using the aggregation scheme proposed by Dümbgen and Spokoiny (2001). The resulting multiscale test statistic is given by

$$\hat{\Psi}_T = \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} \{\hat{S}_{ij}(u, h) - \lambda(h)\}, \quad (6)$$

where $\lambda(h)$ are (appropriately chosen) additive correction terms. In brief, $\lambda(h)$ represents a term designed to account for the multiple testing problem inherent in our framework. The functional form of $\lambda(h)$ does not affect the asymptotic convergence results for our test statistic but may affect the finite sample performance. In our work, $\lambda(h)$ is determined by the extreme value theory and depends on the bandwidth h : drawing on the results from Dümbgen and Spokoiny (2001), the correction term for our setting is given by $\lambda(h) = \sqrt{2 \log(1/2h)}$.

3.3 Computation of the critical values

Suppose for a moment that the exact $(1 - \alpha)$ -quantile $q_T^*(\alpha)$ of the multiscale test statistic $\hat{\Psi}_T$ under H_0 were known. The multiscale test would then follow the rejection rule:

(T*) Reject H_0 if $\hat{\Psi}_T > q_T^*(\alpha)$.

By construction, this decision rule ensures a rigorous level- α -test, meaning that under H_0 we have $\mathbb{P}(\hat{\Psi}_T > q_T^*(\alpha)) = \alpha$.

In practice, however, $q_T^*(\alpha)$ is unknown and can not be computed analytically due to the complicated dependence structure of the individual test statistics across various location-bandwidth points. Hence, it cannot be used to set up the test. In what

follows, we show how to construct an asymptotic approximation $q_T(\alpha)$ that is both computationally feasible and statistically valid. In particular, we require that under H_0 , $q_T(\alpha)$ satisfies the asymptotic property

$$\mathbb{P}(\widehat{\Psi}_T > q_T(\alpha)) = \alpha + o(1). \quad (7)$$

We use strong approximation theory to compute the critical values $q_T(\alpha)$, and in Section 4, we show that this approach ensures an asymptotically valid testing procedure. For calculation of the critical values, we do the following. Under the null H_0 , the individual test statistics

$$\begin{aligned} \hat{S}_{ij}(u, h) &= \left\| \widehat{\Sigma}_{ij}^{-1/2} \mathbf{M}_{XKX}(u, h) \left(\hat{\beta}_i(u, h) - \hat{\beta}_j(u, h) \right) \right\|_{\infty} \\ &= \left\| \widehat{\Sigma}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t (Y_{it} - Y_{jt}) K_{t,u,h} \right) \right\|_{\infty} \\ &= \left\| \widehat{\Sigma}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \left(\mathbf{X}_t' \beta_i(t/T) + \varepsilon_{it} - \mathbf{X}_t' \beta_j(t/T) - \varepsilon_{jt} \right) K_{t,u,h} \right) \right\|_{\infty}, \end{aligned}$$

is exactly equal to

$$\hat{S}_{ij}^0(u, h) = \left\| \widehat{\Sigma}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t (\varepsilon_{it} - \varepsilon_{jt}) K_{t,u,h} \right) \right\|_{\infty}. \quad (8)$$

Under some regularity conditions and a certain rate of consistency of the estimators $\widehat{\Sigma}_i$ and $\widehat{\Sigma}_j$ of Σ_i and Σ_j , respectively, we can show that $\hat{S}_{ij}^0(u, h)$ can be approximated by

$$S_{ij}(u, h) = \left\| \Sigma_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T (\Sigma_i^{1/2} \mathbf{Z}_{it} - \Sigma_j^{1/2} \mathbf{Z}_{jt}) K_{t,u,h} \right) \right\|_{\infty},$$

where $\mathbf{Z}_{it}$ are i.i.d. across i and t multivariate standard normal random vectors that are independent of $\mathbf{X}_s$ and ε_{js} . Essentially, when constructing $S_{ij}(u, h)$, we have replaced the terms $\mathbf{X}_t \varepsilon_{it}$ by the scaled multivariate Gaussian random vectors $\Sigma_i^{1/2} \mathbf{Z}_{it}$. Under an additional simplifying assumption of $\Sigma_i = \Sigma_j$, the Gaussian version $S_{ij}(u, h)$ of the test statistic can be further reduced to

$$S_{ij}(u, h) = \left\| \frac{1}{\sqrt{Th}} \sum_{t=1}^T (\mathbf{Z}_{it} - \mathbf{Z}_{jt}) K_{t,u,h} \right\|_{\infty}. \quad (9)$$

Note that in this case, $S_{ij}(u, h)$ does not depend on any unknown quantities and its quantiles can be (approximately) calculated via Monte Carlo simulations.

In order to construct a global test statistic, we aggregate $S_{ij}(u, h)$ in a similar way as

$\hat{S}_{ij}(u, h)$:

$$\Phi_T = \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} \{S_{ij}(u, h) - \lambda(h)\}. \quad (10)$$

As we will rigorously prove in Section 4, the $1 - \alpha$ quantile $q^*(\alpha)$ of the aggregated data-dependent test statistic $\hat{\Psi}_T$ can be approximated by the $1 - \alpha$ quantile $q^*(\alpha)$ of the similarly aggregated Gaussian test statistic Φ_T . Equation (9) shows that the distribution of $S_{ij}(u, h)$ for each pair $(i, j), 1 \leq i < j \leq N$, and each $(u, h) \in \mathcal{G}_T$, and thus of Φ_T depends only on the Gaussian random vectors $\mathbf{Z}_{it}$. Therefore, accurate critical values of Φ_T can be computed by simulating draws of $\mathbf{Z}_{it}$ for a large number of times B . In our simulation study and application analysis, we choose $B = 5000$, but we have repeated the analysis with $B = 1000$ and the results are roughly the same. Here we formally describe our testing approach. For a given significance level $\alpha \in (0, 1)$, we can compute the critical values of Φ_T according to the following steps:

Step 1. Simulate B independent draws of a matrix

$$\mathcal{Z}^b = (\mathbf{Z}_{it}^b)_{i=1, \dots, N; t=1, \dots, T}, \quad b = 1, 2, \dots, B,$$

where each entry $\mathbf{Z}_{it}^b$ is an independent multivariate standard normal random variable of dimension D .

Step 2. Compute the simulated individual test statistic

$$S_{ij}^b(u, h) = \left\| \frac{1}{\sqrt{Th}} \sum_{t=1}^T (\mathbf{Z}_{it}^b - \mathbf{Z}_{jt}^b) K_{t,u,h} \right\|_{\infty}.$$

Step 3. Aggregate the simulated statistics over all pairs $(i, j), 1 \leq i < j \leq N$, and all location-bandwidth points $(u, h) \in \mathcal{G}_T$ as

$$\Phi_T^b = \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} \{S_{ij}^b(u, h) - \lambda(h)\}.$$

Step 4. Calculate the critical value $q_T(\alpha)$ as the empirical $(1 - \alpha)$ -quantile of the simulated values $\{\Phi_T^b\}_{b=1}^B$.

This resampling approach ensures that the estimated critical values correctly approximate the unknown quantile $q_T^*(\alpha)$, allowing for valid inference. The computational complexity of this procedure is manageable even for large samples, as the Monte Carlo step can be computed in parallel and does not require re-estimation of the coefficient functions.

3.4 The testing procedure

For a given significance level $\alpha \in (0, 1)$, we conduct the multiscale test to assess the global null hypothesis H_0 using the following decision rule:

(T) Reject H_0 if $\widehat{\Psi}_T > q_T(\alpha)$.

By adopting this criterion, the proposed method serves as a test for the overall hypothesis H_0 . Alternatively, it can be interpreted as a simultaneous test for the family of local null hypotheses $H_0^{[i,j]}(u, h)$ across all location-bandwidth points $(u, h) \in \mathcal{G}_T$ and all unit pairs (i, j) , $1 \leq i < j \leq N$. This interpretation allows us to formulate a local decision rule:

(T_{mult}) For each interval $[u - h, u + h]$ with $(u, h) \in \mathcal{G}_T$ and for each pair (i, j) with $1 \leq i < j \leq N$, reject $H_0^{[i,j]}(u, h)$ if $\hat{S}_{ij}(u, h) - \lambda(h) > q_T(\alpha)$.

This local rule ensures that hypothesis testing is performed across multiple time intervals while our approach for constructing of the critical values simultaneously controls for resulting multiple comparisons.

A crucial property of the multiscale test is that it enables rigorous confidence statements about where and when significant deviations occur. Specifically, in Section 4, we prove the following result under suitable regularity conditions:

With asymptotic probability at least $1 - \alpha$, the hypothesis $H_0^{[i,j]}(u, h)$ is violated for all pairs (i, j) and for all intervals $[u - h, u + h]$ satisfying $\hat{S}_{ij}(u, h) - \lambda(h) > q_T(\alpha)$.

This result allows us to make the following simultaneous confidence statement:

With at least $(1 - \alpha)$ (asymptotic) confidence, we can claim that the hypothesis $H_0^{[i,j]}(u, h)$ is violated for all pairs of time series (i, j) and for all intervals $[u - h, u + h]$ for which our test rejects.

This provides a statistically robust way to detect heterogeneous time-varying coefficient functions across unit pairs. By accumulating local test results, we construct a comprehensive confidence region for coefficient differences over time and for all pair-wise comparisons, adding interpretability to the statistical inference.

3.5 Long-run covariance estimation

Define $\mathbf{v}_{it} = \mathbf{X}_t \varepsilon_{it}$. As was mentioned before, our multiscale test relies on a consistent estimator of the long-run variance $\Sigma_i = \sum_{\ell=-\infty}^{\infty} \text{Cov}(\mathbf{v}_{i0}, \mathbf{v}_{i\ell})$. In this section, we describe one of the possible estimators.

We can rewrite the long-run variance Σ_i as follows:

$$\Sigma_i = \lim_{T \rightarrow \infty} \sum_{\ell=-T+1}^{T-1} \Gamma_{i,T}(\ell), \text{ where}$$

$$\Gamma_{i,T}(\ell) = \begin{cases} \frac{1}{T} \sum_{t=\ell+1}^T \mathbb{E}(\mathbf{v}_{it} \mathbf{v}'_{i(t-\ell)}), & \ell \geq 0, \\ \frac{1}{T} \sum_{t=-\ell+1}^T \mathbb{E}(\mathbf{v}_{i(t+\ell)} \mathbf{v}'_{it}), & \ell < 0. \end{cases}$$

A popular choice in the time series literature is the kernel heteroskedasticity and autocorrelation consistent (HAC) covariance estimator. The idea of the kernel HAC estimator is to estimate autocovariances $\Gamma_{i,T}(\ell)$ by sample autocovariances $\hat{\Gamma}_{i,T}(\ell)$ and to attach lower weights to autocovariances with large $|\ell|$ using a kernel function. The kernel weights smooth the sample autocovariances, ensuring a positive semi-definite covariance estimator.

Specifically, we proceed as follows. Let κ denote a real-valued covariance kernel function with some fixed bandwidth χ . The class of fixed-bandwidth kernel HAC estimators is given by

$$\hat{\Sigma}_i(\chi) = \frac{T}{T-D} \sum_{\ell=-T+1}^{T-1} \kappa\left(\frac{\ell}{\chi}\right) \hat{\Gamma}_{i,T}(\ell) \quad (11)$$

$$\hat{\Gamma}_{i,T}(\ell) = \begin{cases} \frac{1}{T} \sum_{t=\ell+1}^T \hat{\mathbf{v}}_{it} \hat{\mathbf{v}}'_{i(t-\ell)}, & \ell \geq 0, \\ \frac{1}{T} \sum_{t=-\ell+1}^T \hat{\mathbf{v}}_{i(t+\ell)} \hat{\mathbf{v}}'_{it}, & \ell < 0, \end{cases} \quad (12)$$

where $\hat{\mathbf{v}}_{it} = \mathbf{X}_t \hat{\varepsilon}_{it}$ with $\hat{\varepsilon}_{it}$ being the residuals after plugging in some estimator $\hat{\beta}_i(t/T)$ of $\beta_i(t/T)$. Importantly, the covariance kernel function κ should give a smaller weight to $\hat{\Gamma}_{i,T}(\ell)$ as $|\ell|$ increases. The following assumptions guarantee that the kernel HAC covariance estimator $\hat{\Sigma}_i$ converges to a true long-run variance matrix Σ_i . The assumptions are a simplified version of Assumptions A-C in Andrews (1991) and Assumptions 5.1-5.3 in Belotti et al. (2023) where we used the fact that we consider a special case of a linear model compared to a general one discussed in these papers.

Let $m_i^{(abcd)}(t, t+r, t+s, t+o)$ denote the time- t fourth-order cumulant of $(\mathbf{v}_{it}^{(a)}, \mathbf{v}_{i(t+r)}^{(b)}, \mathbf{v}_{i(t+s)}^{(c)}, \mathbf{v}_{i(t+o)}^{(d)})$ where $\mathbf{v}_{it}^{(a)}$ is the a th element of a vector $\mathbf{v}_{it}$. That is,

$$m_i^{(abcd)}(t, t+r, t+s, t+o) = \mathbb{E} \left(\mathbf{v}_{it}^{(a)} \mathbf{v}_{i(t+r)}^{(b)} \mathbf{v}_{i(t+s)}^{(c)} \mathbf{v}_{i(t+o)}^{(d)} \right) - \mathbb{E} \left(\tilde{\mathbf{v}}_{it}^{(a)} \tilde{\mathbf{v}}_{i(t+r)}^{(b)} \tilde{\mathbf{v}}_{i(t+s)}^{(c)} \tilde{\mathbf{v}}_{i(t+o)}^{(d)} \right),$$

where $\{\tilde{\mathbf{v}}_{it}\}_{t=1}^T$ is a Gaussian sequence with the same mean and covariance structure as $\{\mathbf{v}_{it}\}_{t=1}^T$. Now, we can formulate the requirements on the dependence structure of the process $\{\mathbf{v}_{it}\}_{t=1}^T$ as follows.

(A2) For all $i = 1, \dots, N$, $T^\vartheta(\hat{\beta}_i - \beta_i) = O_p(1)$ for some $\vartheta \in (0, 1/2)$, $\sup_{t \geq 1} \mathbb{E}(\|\mathbf{v}_{it}\|^2) < \infty$, and $\sup_{t \geq 1} \mathbb{E}(\|\mathbf{X}_t \mathbf{X}'_t\|^2) < \infty$. Moreover, $\int |\kappa(y)| dy < \infty$.

(A3) For all $i = 1, \dots, N$, the process $\{\mathbf{v}_{it}\}_{t=1}^T$ is a mean-zero, fourth-order stationary sequence with $\sum_{\ell=-\infty}^{\infty} \|\Gamma_{i,T}(\ell)\| < \infty$ and

$$\sum_{k=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{o=-\infty}^{\infty} |m_i^{(abcd)}(0, k, s, o)| < \infty$$

for all $a, b, c, d = 1, \dots, D$.

(A4) For all $i = 1, \dots, N$, Assumption (A3) holds with $\mathbf{v}_{it}$ replaced by $(\mathbf{v}'_{it}, (\mathbf{X}_t \mathbf{X}'_t - \mathbb{E}(\mathbf{X}_t \mathbf{X}'_t)))'$.

Belotti et al. (2023) establish that under Assumptions (A2) - (A4), using nonparametric estimator $\hat{\beta}_i$ of β_i with data-driven bandwidths leads to the rate of convergence of a kernel HAC estimator to be polynomial in T . In the appendix, we provide a similar result for the rate of convergence for a fixed bandwidth parameter. This holds true for various kernel functions, and since the necessary condition for our multiscale procedure is to have a rate of convergence $\rho_T = o(\log T)$, our main results remain valid for a number of HAC specifications.

In the estimation procedure described above, we need to choose the bandwidth value which determines the truncation of the sample autocovariances. A common practice which we follow in the simulation study and in the empirical analysis is to set $\chi \approx T^{1/3}$. Alternatively, one can opt for the selection procedure proposed by Newey and West (1994), or the parametric selection methods by Andrews (1991) and Andrews and Monahan (1992).

4 Theory

In this section, we present the theoretical results that are necessary for validity of our multiscale testing procedure.

4.1 Assumptions

To establish the theoretical properties of our test, we impose a set of assumptions about the behavior of the regressors, errors, and underlying dependence structure in the model.

(A5) Processes $\boldsymbol{\varepsilon}_i = \{\varepsilon_{it}\}_{t=1}^T$ are independent across i .

(A6) For each $i = 1, \dots, N$, the process $\boldsymbol{\varepsilon}_i$ satisfies $\mathbb{E}[\varepsilon_{it} | \mathbf{X}_1, \dots, \mathbf{X}_T] = 0$, $\mathbb{E}[\varepsilon_{it}^2 | \mathbf{X}_1, \dots, \mathbf{X}_T] = \sigma_{\varepsilon}^2$ and $\mathbb{E}[\varepsilon_{it} \varepsilon_{is} | \mathbf{X}_1, \dots, \mathbf{X}_T] = \sigma_{\varepsilon, \varepsilon}$ for all $t, s = 1, \dots, T$.

(A7) For each $i = 1, \dots, N$, the variables $\mathbf{v}_{it} = \mathbf{X}_t \varepsilon_{it} \in \mathbb{R}^D$ allow for the representation $\mathbf{v}_{it} = g_i(\mathcal{F}_{it})$ where $\mathcal{F}_{it} = (\dots, \eta_{i(t-1)}, \eta_{it})$ and η_{is} are i.i.d. across s .

(A8) For each $i = 1, \dots, N$, it holds that $\mathbb{E}(\|\mathbf{v}_{it}\|^q) \leq C < \infty$ for some $q > 4$

(A9) For each i , $\sum_{s \geq t} \delta_q(g_i, s) = O(t^{-\xi}(\log t)^{-A})$, where

$$\begin{aligned} A &> \frac{(2q + q^2)\xi + q^2 + 3q + 2 + \gamma^{1/2}}{q(1 + q + 2\gamma)}, \\ \gamma &= q^2(q^2 + 4q - 12)\xi^2 + 2q(q^3 + q^2 - 4q - 4)\xi + (q^2 - q - 2)^2, \\ \xi &> \xi_0 \frac{q^2 - 4 + (q - 2)\sqrt{q^2 + 20q + 4}}{8q}, \end{aligned}$$

and q is defined in (A8).

(A10) There exists $\lambda_* > 0$ and $l_* \in \mathbb{N}$ such that for all i , $t \geq 1$ and $l \geq l_*$, the smallest eigenvalue of $\text{Var}(\sum_{s=t+1}^{t+l} \mathbf{v}_{is})$ is sufficiently bounded away from zero, i.e.,

$$\rho_* \left(\text{Var} \left(\sum_{s=t+1}^{t+l} \mathbf{v}_{is} \right) \right) \geq \lambda_* l.$$

(A11) For all (u, h) , it holds that $\mathbf{M}_{KXX}(u, h) = \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' K_{t,u,h}$ is positive definite with probability going to 1.

(A12) $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrary large but fixed constant $\theta > 0$. Furthermore, we assume that $h_T^{\min} \gg T^{2/(3q)-1/3}$ where q is defined in (A7) and $h_T^{\max} \leq 1/4$.

(A13) The kernel K is non-negative, symmetric about zero, integrates to one and there exists $\delta_K > 0$ and $c_K > 0$ such that $K(z) \geq c_K$ for all $|z| \leq \delta_K$. Moreover, it has compact support $[-1, 1]$, $\frac{1}{2} < \int_{-1}^1 K^2(y) dy < \infty$ and the function is Lipschitz continuous, that is, $|K(v) - K(w)| \leq C|v - w|$ for any $v, w \in \mathbb{R}$ and some constant $C > 0$.

Assumption (A5) rules out cross-sectional dependence in the idiosyncratic errors assuming that the dependence structure across units is fully explained by the common covariates. This condition allows us to construct the multiscale test statistics by combining unit-wise Gaussian approximations without having to control for additional cross-sectional correlation. Assumption (A6) imposes a homogeneous second-order structure for the idiosyncratic errors conditional on the common regressors $(\mathbf{X}_1, \dots, \mathbf{X}_T)$. Together with Assumption (A5) and the representation in Assumption (A7), this leads to a weakly dependent, second-order stationary structure for the processes $\{v_{it}\}_{t=1}^T$.

Assumption (A7) places the product process $\mathbf{v}_{it} = \mathbf{X}_t \varepsilon_{it}$ in the physical-dependence framework of Wu (2005). This representation allows us to quantify temporal dependence via the physical dependence measures $\delta_q(g_i, t)$ and to invoke the strong approximation results of Karmakar and Wu (2020). Assumption (A8) provides sufficient control of the tails of $\mathbf{v}_{it}$.

Assumptions (A9) and (A10) correspond to the assumptions in Karmakar and Wu (2020, Theorem 2.1-2.2) that justify the strong approximation result. Assumption (A9) specifies how fast the physical dependence coefficients $\delta_q(g_i, t)$ must decay. It implies that the process $\{\mathbf{v}_{it}\}_{t=1}^T$ is short-memory and yields a strong approximation error of order $T^{1/q}$. Note that one could allow for stronger serial dependence by choosing a smaller ξ at the price of a slower approximation rate. In that case, our main theoretical results would still remain valid. Assumption (A10) is a common assumption in time series literature that ensures that the eigenvalues of the covariance matrices of the increment processes are substantially large and away from zero, implying well-conditioned and positive definite covariance matrices.

Assumption (A11) concerns the localized design matrix $\mathbf{M}_{KX}(u, h)$ and requires this matrix to be positive definite with probability tending to one. This excludes local multicollinearity of the regressors on each interval $[u - h, u + h]$ in the grid G_T and ensures that the kernel estimator $\hat{\beta}_i(u, h)$ is well defined for all $(u, h) \in G_T$ and that the normalization by $\mathbf{M}_{KX}(u, h)$ in the test statistic is stable.

Assumption (A12) governs the practical choice of the grid $\mathcal{G}_T$ and is a weak restriction commonly used in multiscale testing (Khismatullina and Vogt, 2020, 2022). The condition $|G_T| = O(T^\theta)$ for some fixed $\theta > 0$ allows the number of location–bandwidth points to grow polynomially in T , which is large enough in practice to cover a rich collection of intervals. The lower bound $h_T^{\min} \gg T^{2/(3q)-1/3}$ prevents us from using bandwidths that are too small, that is, from forming intervals with too few observations for the limit theory to apply. The upper bound $h_T^{\max} \leq 1/4$ rules out overly large bandwidths that would average over almost the entire sample and blur local features.

Assumption (A13) is a standard assumption in the nonparametric time series literature that is satisfied by many standard kernels such as the Epanechnikov kernel that we use in our simulation and application studies. The only non-standard restriction that we impose on the kernel function is the requirement $\frac{1}{2} < \int_{-1}^1 K^2(y)dy$ which is necessary for the anti-concentration bounds used in the Gaussian approximation of the multiscale statistic. This restriction is also satisfied by the most common kernel functions such as Epanechnikov (which we use in our empirical analysis), biweight or triweight kernels.

4.2 Theoretical results

In this section, we establish the theoretical properties of the multiscale test introduced in Section 3. We first show that the distribution of the test statistic $\hat{\Psi}_T$ can be approximated by the distribution of a suitable Gaussian max-type statistic Φ_T . This strong approximation justifies the use of the simulated critical values $q_T(\alpha)$ defined in Section 3.3. We then derive the asymptotic size of our test for the global null hypothesis H_0 , study local power properties under a class of shrinking alternatives, and finally show that the procedure controls the family-wise error rate when simultaneously testing the

full number of the local null hypotheses $H_0^{[i,j]}(u, h)$.

Theorem 4.1. *Let Assumptions (A1) - (A13) be fulfilled. Moreover, assume that the estimators $\widehat{\Sigma}_i$ of the long-run covariance matrices Σ_i satisfy $\|\widehat{\Sigma}_i - \Sigma_i\|_1 = o_p(\rho_T)$ for some $\rho_T = o(\log T)$, and that the long-run covariance matrices are homogeneous across individuals, $\Sigma_i = \Sigma$ for all i . Then, under the global null H_0 ,*

$$\sup_{y \in \mathbb{R}} \left| \mathbb{P}(\widehat{\Psi}_T \leq y) - \mathbb{P}(\Phi_T \leq y) \right| = o(1).$$

Theorem 4.1 shows that, under H_0 , the Kolmogorov distance between the distribution of the multiscale statistic $\widehat{\Psi}_T$ and that of the Gaussian test statistic Φ_T converges to zero as the sample size increases. The additional assumption on $\widehat{\Sigma}_i$ requires the long-run covariance matrices Σ_i to be estimated with sufficient accuracy. As discussed in Section 3.5, this rate can be achieved by standard HAC-type estimators. The homogeneity condition $\Sigma_i = \Sigma$ for all i simplifies the covariance structure across units and allows us to work with a common Gaussian approximation. Intuitively, Theorem 4.1 ensures that the high-dimensional dependence across location–bandwidth points $(u, h) \in \mathcal{G}_T$ and pairs of units (i, j) is adequately captured by the Gaussian coupling embodied in Φ_T . Proof of Theorem 4.1 is given in Appendix.

The result in Theorem 4.1 is necessary for deriving the theoretical properties of the proposed multiscale test. Recall that $q_T(\alpha)$ denotes the $(1 - \alpha)$ -quantile of Φ_T . First, we show that the test based on the critical value $q_T(\alpha)$ has asymptotically the correct size: the rejection probability under H_0 converges to the nominal level α .

Proposition 4.2 (Asymptotic size). *Suppose that the assumptions of Theorem 4.1 are fulfilled. Then, under H_0 ,*

$$\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) = (1 - \alpha) + o(1).$$

Note that the size control is obtained despite the fact that the dimension of $\widehat{\Psi}_T$, that is, the number of location–bandwidth points $(u, h) \in \mathcal{G}_T$ and ordered pairs (i, j) , may grow with T .

Second, we study the behavior of the test under a certain class of local alternatives and we show that under this certain class, our test has sufficient power. Let $\beta_i = \beta_{i,T}$ be a sequence of coefficient functions that depend on the time length T . The following class of alternatives ensures that, for some pair of units, sequences $\beta_{i,T}$ and $\beta_{j,T}$ are sufficiently separated on at least one interval in the multiscale grid.

Proposition 4.3 (Local power). *Let the assumptions of Theorem 4.1 be fulfilled. Moreover, assume that the following holds: for some pair (i, j) , there exists $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$ and some coordinate $d \in \{1, \dots, D\}$ such that*

$|\beta_{i,T}(w) - \beta_{j,T}(w)| \geq c_T \sqrt{\log T / (Th)}$ for all $w \in [u - h, u + h]$, where $\{c_T\}$ is any sequence of positive numbers with $c_T \rightarrow \infty$ as $T \rightarrow \infty$. Then it holds that

$$\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) = o(1),$$

and hence, the test rejects H_0 with probability tending to 1.

So far we have focused on the global null hypothesis H_0 of complete homogeneity across units. The construction of $\widehat{\Psi}_T$ also enables multiple testing of the collection of local null hypotheses $H_0^{[i,j]}(u, h)$ defined in Section 3.1. We now formalize the multiple testing problem and show that the procedure controls the family-wise error rate (FWER).

Let $\mathcal{M} = \{(u, h, i, j) : (u, h) \in \mathcal{G}_T \text{ and } 1 \leq i < j \leq N\}$ denote the collection of all location-scale points $(u, h) \in \mathcal{G}_T$ and all pairs of units (i, j) . Let $\mathcal{M}_0 \subseteq \mathcal{M}$ denote the collection of tuples (u, h, i, j) for which the local null hypothesis $H_0^{[i,j]}(u, h)$ is true. Our property of interest, FWER, is defined as the probability of rejecting $H_0^{[i,j]}(u, h)$ for at least one tuple $(u, h, i, j) \in \mathcal{M}_0$. That is,

$$\text{FWER}(\alpha) = \mathbb{P}\left(\exists (u, h, i, j) \in \mathcal{M}_0 : \hat{S}_{ij}(u, h) - \lambda(h) > q_T(\alpha)\right)$$

for some given significance level $\alpha \in (0, 1)$. We would like to bound the FWER, controlling the rate of false discoveries. The following proposition shows that our multiscale test asymptotically controls the FWER at level α .

Proposition 4.4. *Suppose that the assumptions of Theorem 4.1 are fulfilled. Then, for any given $\alpha \in (0, 1)$,*

$$\text{FWER}(\alpha) \leq \alpha + o(1).$$

Equivalently, Proposition 4.4 implies that

$$\mathbb{P}\left(\forall (u, h, i, j) \in \mathcal{M}_0 : \hat{S}_{ij}(u, h) - \lambda(h) \leq q_T(\alpha)\right) \geq 1 - \alpha + o(1),$$

that is, with probability at least $1 - \alpha$ (asymptotically), none of the true local null hypotheses is rejected by our procedure. This result provides the theoretical justification for the simultaneous confidence statement (R) in Section 3.4: the set of accepted local hypotheses delivers a joint confidence statement about all coefficient functions across all units and all intervals, with an asymptotic coverage probability at least $1 - \alpha$.

5 Clustering

The global null hypothesis H_0 says that the coefficient functions are fully homogeneous across individuals. A rejection of H_0 implies the presence of heterogeneity in the coeffi-

cient functions. Rather than full heterogeneity across units, it is plausible that groups of individuals share the same coefficient functions - a structure commonly referred to as grouped heterogeneity (Bonhomme and Manresa, 2015; Su et al., 2016; Chen, 2019). In this section, we discuss how our multiscale test can be used to uncover such group structure in the data.

We assume that the group pattern follows the form

$$\beta_i = \sum_{k=1}^{K_0} \gamma_k \mathbb{1}\{i \in G_k\},$$

where K_0 is the number of groups, G_k denotes the k -th group, with the partition $\bigcup_{k=1}^{K_0} G_k = \{1, 2, \dots, N\}$ and $G_k \cap G_{k'} = \emptyset$ for $k \neq k'$. Each group shares a common coefficient function γ_k , and these functions differ across groups, i.e., $\gamma_k \neq \gamma_{k'}$ for $k \neq k'$. Since the group memberships are typically unknown, they need to be estimated from the data.

To estimate the group structure, we use partially aggregated individual multiscale test statistics

$$\max_{(u,h) \in \mathcal{G}_T} \{\hat{S}_{ij}(u, h) - \lambda(h)\}$$

as a dissimilarity measure between series i and j . Following Chen (2019) and Khismatullina and Vogt (2022), we use these distances as inputs for a hierarchical agglomerative clustering (HAC) algorithm. Given a specified number of groups K_0 , the algorithm returns the estimated group memberships $\{\hat{G}_1, \dots, \hat{G}_{K_0}\}$. We refer the reader to Chen (2019), Khismatullina and Vogt (2022), and Hastie et al. (2009) for more details about the HAC in this context.

If the number of groups is unknown, we estimate it by selecting the smallest number of groups $\hat{K}$ such that the maximum within-group distance does not exceed the critical value $q_T(\alpha)$. We show that $\hat{K}$ is a consistent estimator of K_0 in the current setting.¹ Below we state our assumptions on the group-specific coefficient functions.

- (A13) For each group $k = 1, \dots, K_0$, the trend coefficient functions $\gamma_k = (\gamma_{k,1}, \dots, \gamma_{k,D})'$ are Lipschitz continuous on $[0, 1]$ coordinate-wise with normalization $\int_0^1 \gamma_{k,d}(u) du = 0$ for each $d = 1, \dots, D$. Moreover, for any $k \neq k'$, the coefficient functions are different in the following way: There exists $(u, h) \in \mathcal{G}_T$ and a coordinate $d \in \{1, \dots, D\}$ such that $|\gamma_k(w) - \gamma_{k'}(w)|_d \geq c_T \sqrt{\log T / (Th)}$ for all $w \in [u - h, u + h]$, where $0 < c_T \rightarrow \infty$.

The first part of Assumption (A13) repeats Assumption (A1), whereas the second part is a necessary condition for the consistency of $\hat{K}$ and is common in the literature.

¹Alternative selection procedures exist, such that the information criterion by Chen (2019), and may be validated in the current setting.

The following proposition states that the hierarchical clustering algorithm produces consistent group-membership estimates and that the estimator $\widehat{K}$ of the true number of groups K_0 is consistent.

Proposition 5.1. *Let the assumptions of Theorem 4.1 and (A13) be fulfilled. Then we have that*

$$\mathbb{P}(\{\widehat{G}_1, \dots, \widehat{G}_{\widehat{K}}\} = \{G_1, \dots, G_{K_0}\}) \geq (1 - \alpha) + o(1),$$

and

$$\mathbb{P}(\widehat{K} = K_0) \geq (1 - \alpha) + o(1).$$

Importantly, our multiscale approach allows us to identify the locations where two group-specific coefficients γ_k and $\gamma_{k'}$ are different from each other. For each pair of groups (k, k') , let $\mathcal{A}_T^{[k, k']}$ denote the set of $(u, h) \in \mathcal{G}_T$ for which there is at least one local difference between individuals across groups, that is,

$$\mathcal{A}_T^{[k, k']}(\alpha) = \{(u, h) \in \mathcal{G}_T : \hat{S}_{ij}(u, h) > q_T(\alpha) \text{ for some } i \in \widehat{G}_k, j \in \widehat{G}_{k'}\}.$$

We define the set $\mathcal{S}_T^{[k, k']}(\alpha)$ as the set of intervals $[u - h, u + h] \subseteq [0, 1]$ for which $(u, h) \in \mathcal{A}_T^{[k, k']}(\alpha)$, i.e.,

$$\mathcal{S}_T^{[k, k']}(\alpha) = \{[u - h, u + h] : (u, h) \in \mathcal{A}_T^{[k, k']}(\alpha)\}.$$

Moreover, let

$$E_T^{[k, k']}(\alpha) = \{\forall [u - h, u + h] \in \mathcal{S}_T^{[k, k']}(\alpha) : \gamma_k(v) \neq \gamma_{k'}(v) \text{ for some } v \in [u - h, u + h]\}$$

be the event that the group-specific coefficients differ on all intervals in $\mathcal{S}_T^{[k, k']}(\alpha)$. The following proposition says the multiscale test correctly identifies the time intervals where the group-specific coefficients γ_k and $\gamma_{k'}$ are different from each other with asymptotic confidence $1 - \alpha$.

Proposition 5.2. *Let the assumptions of Proposition 5.1 be fulfilled. Then we have that the event*

$$E_T(\alpha) = \{\cap_{1 \leq k < k' \leq \widehat{K}} E_T^{[k, k']}(\alpha)\} \cap \{\widehat{K} = K_0 \text{ and } \widehat{G}_k = G_k \text{ for all } k\}$$

asymptotically occurs with probability at least $1 - \alpha$, that is,

$$\mathbb{P}(E_T(\alpha)) \geq (1 - \alpha) + o(1).$$

Remark 5.3. *In this work we assume that the group structure is the same across all D coordinates. Recent works by Cheng et al. (2019) and Leng et al. (2023) discuss models with multi-dimensional group structure, where different parameters may correspond to different group structures. We leave this extension for further research.*

6 Empirical application

In our empirical study, we revisit the relationship between U.S. interest rates and foreign output. Iacoviello and Navarro (2019) find that U.S. monetary tightening results in larger drops in GDP for emerging economies than for advanced economies, and they document substantial heterogeneity in the effect of U.S. monetary shocks on foreign economies. We employ our multiscale test to test for heterogeneity across economies and across time. While doing so, we allow for a time-varying relationship between U.S. monetary shocks and foreign GDP, as opposed to the static relationship in Iacoviello and Navarro (2019).

The quarterly interest rate and foreign GDP data spans the period between 1965-Q1 and 2016-Q2, and covers 49 foreign countries. We include the U.S. GDP in our GDP data, bringing the number of time series to $N = 50$. The dataset is taken from Iacoviello and Navarro (2019) and is publicly available at the authors' website.² For the detailed description of the dataset we refer the reader to the original paper.

For our analysis, we construct the U.S. monetary shock variable (denoted further by X_t) as a residual of a regression on several control variables as described by Iacoviello and Navarro (2019). The GDP data (denoted by y_{it} for a country i) is de-seasonalized by extracting the residual from an country-specific regression on the fourth-order lag and a quadratic trend (corresponding to the controls in Iacoviello and Navarro (2019)). Afterwards, we demean each of the obtained time series separately in order to account for the individual fixed-effects. As a result, we estimate the time-varying GDP response to an U.S. monetary shock from the following model:

$$y_{i(t+h)} = \beta_i(t/T)X_t + \epsilon_{it},$$

for some lag h . In what follows, we fix $h = 1$ in the current draft.

The test is applied exactly as described in Section 3.3. The nonparametric estimators are computed as described in Section 2.2, where we employ the Epanechnikov kernel. We estimate the long-run variances Σ_i , $i = 1, \dots, N$, by the Bartlett-kernel HAC estimator with data-driven bandwidth (see Section 3.5). We construct the grid $\mathcal{G}_T$

²<https://www.matteoiacoviello.com/research.htm>

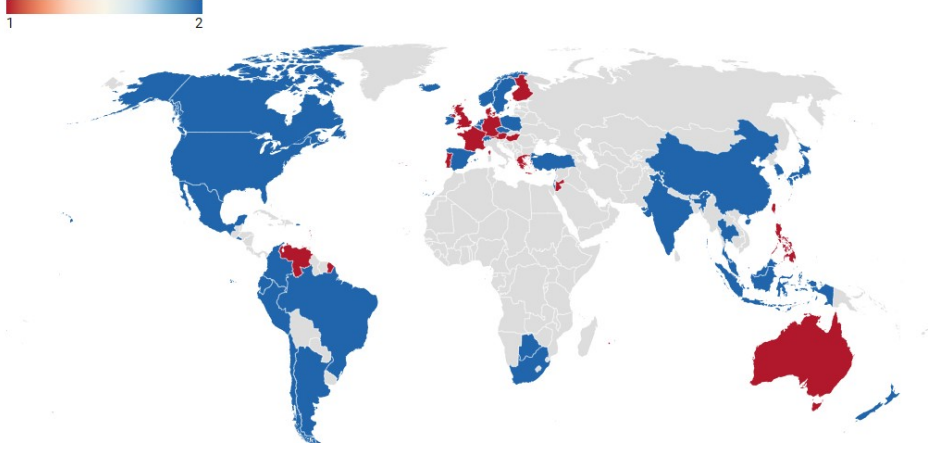


Figure 1: Estimated cluster memberships using hierarchical clustering. The number of clusters is set to $K = 2$.

following Khismatullina and Vogt (2022) as $\mathcal{G}_T = U_T \times H_T$ with

$$U_T = \{u \in [0, 1] : u = \frac{5t}{T} \text{ for some } t \in \mathbb{N}\},$$

$$H_T = \left\{ h \in \left[\frac{1}{T^{1/3}}, \frac{1}{4} \right] : h = \frac{5t-3}{T} \text{ for some } t \in \mathbb{N} \right\}.$$

Furthermore, we compute the critical value $q_T(\alpha)$ of our test by Monte Carlo simulation based on $B = 1000$ Gaussian samples. The results are as follows. First, we reject the null hypothesis of global homogeneity at 5% level. This indicates that there is a heterogeneous effect of U.S. monetary shocks on foreign GDP. To investigate the heterogeneity further, we partition the countries into clusters. Given the natural partition into advanced vs. emerging countries and to be able to directly compare with Iacoviello and Navarro (2019), we use $K = 2$ clusters.³ We visualize the clusters for $K = 2$ in Figure 1. The figure illustrates that the second cluster is mostly made up from European countries, while the first clusters covers South America and Asia. The clusters correspond mostly with the partition into advanced and emerging economies made by Iacoviello and Navarro (2019), although some differences arise. We find that the time-varying effect of the U.S. monetary shock are similar for the U.S. and Canadian economies, however, we find significant differences between these economies and most European countries which together with Australia mostly constitute a separate cluster. Our results indicate that allowing for time-varying exposure to U.S. monetary shocks, the heterogeneity in exposure may be more sophisticated than the classical advanced vs. emerging economies partition that is currently used. To sum up, our test serves as a statistically rigorous starting point for further analysis of the various economic relationships.

³Our selection procedure indicates $K = 18$ clusters but for interpretability reasons we chose to report the results for $K = 2$ clusters.

7 Conclusion

In this paper, we introduce a multiscale test for heterogeneous time-varying coefficients in panels and prove that it asymptotically controls the family-wise error rate. We also show that under local alternatives our test has asymptotic power of one. The proposed multiscale procedure allows us to make simultaneous confidence statements about which of the coefficient functions are different and where these differences occur. Furthermore, the results of our test can be used as a dissimilarity measure for a subsequent HAC clustering algorithm. For the real-life application, we study the relationship between U.S interest rates and foreign output of 49 economies and the economy of U.S. itself. The results show that foreign GDP responses to U.S. monetary shocks exhibit heterogeneity and can be partitioned into two clusters with crisis-specific dynamics. Future research will relax the common-regressor assumption and extend strong approximation theory to other forms of dependencies.

8 Acknowledgement

The authors thank Chen Zhou and Wendun Wang for their insightful suggestions during the preparation of the paper.

References

- ADRIAN, T. and FRANZONI, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance*, **16** 537–556.
- ANDO, T. and BAI, J. (2015). A simple new test for slope homogeneity in panel data models with interactive effects. *Economics Letters*, **136** 112–117.
- ANDREWS, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59** 817–858.
- ANDREWS, D. W. K. and MONAHAN, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, **60** 953–966.
- ATAK, A., YANG, T. T., ZHANG, Y. and ZHOU, Q. (2025). Specification tests for time-varying coefficient panel data models. *Econometric Theory*, **41** 123–170.
- BALTAGI, B. H., BRESSON, G. and PIROTTE, A. (2008). To pool or not to pool? In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Springer, 517–546.
- BELOTTI, F., CASINI, A., CATANIA, L., GRASSI, S. and PERRON, P. (2023). Simultaneous bandwidths determination for DK-HAC estimators and long-run variance estimation in nonparametric settings. *Econometric Reviews*, **42** 281–306.
- BONHOMME, S. and MANRESA, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, **83** 1147–1184.
- CAI, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, **136** 163–188.
- CHEN, B. and HONG, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, **80** 1157–1183.

- CHEN, J. (2019). Estimating latent group structure in time-varying coefficient panel data models. *The Econometrics Journal*, **22** 223–240.
- CHEN, S.-N. (1982). An examination of risk-return relationship in bull and bear markets using time-varying betas. *Journal of Financial and Quantitative Analysis*, **17** 265–286.
- CHENG, X., SCHORFHEIDE, F. and SHAO, P. (2019). Clustering for multi-dimensional heterogeneity. Tech. rep., Working paper.
- DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Annals of Statistics*, **29** 124–152.
- FAMA, E. F. and MACBETH, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, **81** 607–636.
- GAO, J., XIA, K. and ZHU, H. (2020). Heterogeneous panel data models with cross-sectional dependence. *Journal of Econometrics*, **219** 329–353.
- HAIMERL, P., SMEEKES, S. and WILMS, I. (2025). Estimation of latent group structures in time-varying panel data models. *arXiv preprint arXiv:2503.23165*.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics, Springer.
- HSIAO, C. and TAHMISCIOGLU, A. K. (1997). A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association*, **92** 455–465.
- IACOVIELLO, M. and NAVARRO, G. (2019). Foreign effects of higher us interest rates. *Journal of International Money and Finance*, **95** 232–250.
- JAGANNATHAN, R. and WANG, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance*, **51** 3–53.
- KARMAKAR, S. and WU, W. B. (2020). Optimal Gaussian approximation for multiple time series. *Statistica Sinica*, **30** 1399–1417.

- KEANE, M. and NEAL, T. (2020). Climate change and US agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics*, **11** 1391–1429.
- KHISMATULLINA, M. and VOGT, M. (2020). Multiscale inference and long-run variance estimation in non-parametric regression with time series errors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **82** 5–37.
- KHISMATULLINA, M. and VOGT, M. (2022). Multiscale comparison of nonparametric trend curves. *arXiv preprint arXiv:2209.10841*.
- KHISMATULLINA, M. and VOGT, M. (2023). Nonparametric comparison of epidemic time trends: The case of COVID-19. *Journal of Econometrics*, **232** 87–108.
- LENG, X., CHEN, H. and WANG, W. (2023). Multi-dimensional latent group structures with heterogeneous distributions. *Journal of Econometrics*, **233** 1–21.
- LEWELLEN, J. and NAGEL, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, **82** 289–314.
- LI, D., CHEN, J. and GAO, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, **14** 387–408.
- LOBELL, D. B., ROBERTS, M. J., SCHLENKER, W., BRAUN, N., LITTLE, B. B., REJESUS, R. M. and HAMMER, G. L. (2014). Greater sensitivity to drought accompanies maize yield increase in the US Midwest. *Science*, **344** 516–519.
- LUMSDAINE, R. L., OKUI, R. and WANG, W. (2023). Estimation of panel group structure models with structural breaks in group memberships and coefficients. *Journal of Econometrics*, **233** 45–65.
- NEWAY, W. K. and WEST, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, **61** 631–653.
- OKUI, R. and WANG, W. (2021). Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, **220** 447–473.

- PESARAN, M. H. and YAMAGATA, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, **142** 50–93.
- SU, L. and CHEN, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, **29** 1079–1135.
- SU, L., SHI, Z. and PHILLIPS, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, **84** 2215–2264.
- SU, L., WANG, X. and JIN, S. (2019). Sieve estimation of time-varying panel data models with latent structures. *Journal of Business & Economic Statistics*, **37** 334–349.
- WANG, W., ZHANG, X. and PAAP, R. (2019). To pool or not to pool: What is a good strategy for parameter estimation and forecasting in panel regressions? *Journal of Applied Econometrics*, **34** 724–745.
- WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, **102** 14150–14154.
- ZHANG, Y. and ZHOU, Q. (2021). Partially linear functional-coefficient dynamic panel data models: Sieve estimation and specification testing. *Econometric Reviews*, **40** 983–1006.

A Proofs

Proof of Theorem 4.1. With the help of this theorem, we justify that inference using the critical values from Section 3 is valid. We first summarize the main proof strategy, which splits up into three steps, and then fill in the details. We will also make use of an auxiliary test statistics

$$\widehat{\Psi}_T^0 = \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \{\hat{S}_{ij}^0(u,h) - \lambda(h)\},$$

where

$$\hat{S}_{ij}^0(u,h) = \left\| \widehat{\Sigma}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t ((\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j)) K_{t,u,h} \right) \right\|_{\infty},$$

as specified in (8). $\hat{S}_{ij}^0(u,h)$ is different from $\hat{S}_{ij}(u,h)$ only in the fact that it does not depend on the coefficient functions β_i and β_j . Under the null H_0 , when we assume that $\beta_i \equiv \beta_j$ for all $1 \leq i < j \leq N$, $\widehat{\Psi}_T$ is exactly equal to $\widehat{\Psi}_T^0$.

Step 1

We first have a closer look at the statistic $\widehat{\Psi}_T^0$. In particular, we show that there exists an identically distributed version $\widetilde{\Psi}_T$ of $\widehat{\Psi}_T^0$ which is close to the Gaussian statistic Φ_T from (10). More formally, we prove the following result.

Proposition A.1. *There exist statistics $\{\widetilde{\Psi}_T : T = 1, 2, \dots\}$ with the following two properties:*

(i) $\widetilde{\Psi}_T$ has the same distribution as $\widehat{\Psi}_T^0$ for any T , and (ii)

$$|\widetilde{\Psi}_T - \Phi_T| = o_p(\delta_T),$$

where $\delta_T = T^{1/q} / \sqrt{T(h_T^{\min})^3} + \rho_T \sqrt{\log T}$ and Φ_T is a Gaussian statistic as defined in (10).

The proof makes heavy use of strong approximation theory for dependent processes. As it is quite technical, we defer the full proof to the next subsection.

Step 2

In this step, we establish some properties of the Gaussian statistic Φ_T . Specifically, we prove the following result.

Proposition A.2. *It holds that*

$$\sup_{y \in \mathbb{R}} \mathbb{P}(|\Phi_T - y| \leq \delta_T) = o(1),$$

$$\text{where } \delta_T = \frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}} + \rho_T \sqrt{\log T}.$$

Roughly speaking, this proposition says that the random variable Φ_T does not concentrate too strongly in small regions of the form $[y - \delta_T, y + \delta_T]$ with δ_T converging to 0. The main technical tool for deriving it are anti-concentration bounds for Gaussian random vectors. The proof of this proposition with $\delta_T = T^{1/q} / \sqrt{T(h_T^{\min})^3} + \rho_T \sqrt{\log T}$ can be found later in this section and closely follows the proof in Khismatullina and Vogt (2022) (Proposition A.8).

Step 3

We now use Steps 1 and 2 to prove that

$$\sup_{y \in \mathbb{R}} |\mathbb{P}(\hat{\Psi}_T^0 \leq y) - \mathbb{P}(\Phi_T \leq y)| = o_P(1).$$

The proof of this proposition can be found in Khismatullina and Vogt (2022), statement (A.5). $\square$

Proof of Proposition 5.1. The proof of this results follows from the steps in the proof of Proposition 5.1 in Khismatullina and Vogt (2022) and therefore omitted for brevity. $\square$

Proof of Proposition 5.2. We define the auxiliary statistic

$$\hat{\Psi}_T^0 = \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} \{\hat{S}_{ij}^0(u, h) - \lambda(h)\}, \quad (\text{A.1})$$

where $\hat{S}_{ij}^0(u, h)$ is defined in (8). Let us consider the event

$$B_T = \{\hat{\Psi}_T^0 \leq q_T(\alpha) \quad \text{and} \quad \min_{1 \leq k < k' \leq K_0} \min_{\substack{i \in G_k \\ j \in G_{k'}}} \hat{\Psi}_{ij, T} \geq q_T(\alpha)\},$$

where the statistic $\hat{\Psi}_{ij, T}$ is defined as

$$\hat{\Psi}_{ij, T} = \max_{(u, h) \in \mathcal{G}_T} \{\hat{S}_{ij}(u, h) - \lambda(h)\},$$

for each i and j . In other words, $\min_{1 \leq k < k' \leq K_0} \min_{\substack{i \in G_k \\ j \in G_{k'}}} \hat{\Psi}_{ij, T}$ corresponds to the smallest

between-group multiscale distance. In case event B_T holds, it holds that

$$\hat{\Psi}_T^0 < \min_{1 \leq k < k' \leq K_0} \min_{\substack{i \in G_k \\ j \in G_{k'}}} \hat{\Psi}_{ij,T}.$$

Since $\hat{\Psi}_T^0$ is identical to $\hat{\Psi}_T$ under the null H_0 , we can use Proposition 4.2 to conclude that

$$\mathbb{P} \left(\hat{\Psi}_T^0 \leq q_T(\alpha) \right) \geq (1 - \alpha) + o(1).$$

Furthermore, applying the same arguments used for Proposition 4.3 yields that

$$\mathbb{P} \left(\min_{1 \leq k < k' \leq K_0} \min_{\substack{i \in G_k \\ j \in G_{k'}}} \hat{\Psi}_{ij,T} \leq q_T(\alpha) \right).$$

Therefore, we obtain that

$$\mathbb{P}(B_T) \geq (1 - \alpha) + o(1). \quad (\text{A.2})$$

Analogous to the proof of Proposition 5.1 in Khismatullina and Vogt (2022), we can show that

$$B_T \subseteq \{\hat{K} = K_0 \text{ and } \hat{G}_k = G_k \text{ for all } k\}, \quad (\text{A.3})$$

as a result of the HAC algorithm to obtain clusters. Moreover, since for all i, j and $(u, h) \in \mathcal{G}_T$ for which it holds that

$$\hat{S}_{ij}^0(u, h) - \lambda(h) \geq q_T(\alpha) \quad \text{and} \quad \hat{S}_{ij}(u, h) - \lambda(h) \geq q_T(\alpha),$$

it must hold that

$$\sum_{t=1}^T \mathbf{X}_t \left(\mathbf{X}_t' \boldsymbol{\beta}_i(t/T) - \mathbf{X}_t' \boldsymbol{\beta}_j(t/T) \right) K_{t,u,h} \neq 0,$$

which implies that $\boldsymbol{\beta}_i(v) - \boldsymbol{\beta}_j(v) \neq 0$ for some $v \in [u - h, u + h]$. Therefore, we obtain that

$$B_T \subseteq \bigcap_{1 \leq k < k' \leq \hat{K}} E_T^{k,k'}(\alpha). \quad (\text{A.4})$$

Combining the results in (A.3) and (A.4), we obtain that

$$B_T \subseteq \left\{ \bigcap_{1 \leq k < k' \leq \hat{K}} E_T^{k,k'}(\alpha) \right\} \cap \{ \hat{K} = K_0 \text{ and } \hat{G}_k = G_k \text{ for all } k \} = E_T(\alpha).$$

Hence, result (A.2) implies that $\mathbb{P}(E_T(\alpha)) \geq (1 - \alpha) + o(1)$. $\square$

B Additional Proofs

Proof of Proposition A.1. Consider the stationary D -dimensional process $\mathcal{N}_i = \{\mathbf{v}_{it} = \mathbf{X}_t \varepsilon_{it} : 1 \leq t \leq T\}$ for some fixed $i \in \{1, \dots, N\}$. By Theorem 2.2 in Karmakar and Wu (2020), the following strong approximation result holds true: On a richer probability space, there exist a sequence of random vectors $\tilde{\mathcal{N}}_i = \{\tilde{\mathbf{v}}_{it} : t \in \mathbb{N}\}$ such that $[\tilde{\mathbf{v}}_{i1}, \dots, \tilde{\mathbf{v}}_{iT}] \stackrel{\mathcal{D}}{=} [\mathbf{v}_{i1}, \dots, \mathbf{v}_{iT}]$ for each T and

$$\max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{is} - \boldsymbol{\Sigma}_i^{1/2} IB(t) \right\| = o_p(T^{1/q}), \quad (\text{B.1})$$

where $IB(\cdot)$ is a centered Brownian motion in $\mathbb{R}^D$ and $\boldsymbol{\Sigma}_i = \sum_{k \in \mathbb{Z}} \text{Cov}(\mathbf{v}_{i0}, \mathbf{v}_{ik})$ is the long-run covariance matrix of $\{\mathbf{v}_{it} : 1 \leq t \leq T\}$. Let $\tilde{S}_{ij}(u, h)$ be defined similarly to $\hat{S}_{ij}^0(u, h)$ but with $\tilde{\mathbf{v}}_{it}$ instead of $\mathbf{v}_{it}$:

$$\tilde{S}_{ij}(u, h) = \left\| \tilde{\boldsymbol{\Sigma}}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T ((\tilde{\mathbf{v}}_{it} - \tilde{\mathbf{v}}_i) - (\tilde{\mathbf{v}}_{jt} - \tilde{\mathbf{v}}_j)) K_{t,u,h} \right) \right\|_{\infty},$$

where the estimator $\tilde{\boldsymbol{\Sigma}}_{ij}$ is constructed from the samples $\tilde{\mathcal{N}}_i$ and $\tilde{\mathcal{N}}_j$ in the same way as $\hat{\boldsymbol{\Sigma}}_{ij}$ is constructed from $\mathcal{N}_i$ and $\mathcal{N}_j$. Note that since $[\tilde{\mathbf{v}}_{i1}, \dots, \tilde{\mathbf{v}}_{iT}] \stackrel{\mathcal{D}}{=} [\mathbf{v}_{i1}, \dots, \mathbf{v}_{iT}]$ and $\|\hat{\boldsymbol{\Sigma}}_{ij}^{-1/2} - \boldsymbol{\Sigma}_{ij}^{-1/2}\|_1 = o_p(\rho_T)$, we have that $\|\tilde{\boldsymbol{\Sigma}}_{ij}^{-1/2} - \boldsymbol{\Sigma}_{ij}^{-1/2}\|_1 = o_p(\rho_T)$ (which we will use later in the proof).

Similarly to $\hat{\Psi}_T^0$, $\tilde{\Psi}_T$ is defined as

$$\tilde{\Psi}_T = \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \left\{ \tilde{S}_{ij}(u, h) - \lambda(h) \right\}.$$

Moreover, we introduce an auxiliary statistic

$$\Phi_T^\diamond = \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \left\{ S_{ij}^\diamond(u, h) - \lambda(h) \right\},$$

where

$$S_{ij}^\diamond(u, h) = \left\| \tilde{\Sigma}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \left(\Sigma_i^{1/2}(\mathbf{Z}_{it} - \bar{\mathbf{Z}}_i) - \Sigma_j^{1/2}(\mathbf{Z}_{jt} - \bar{\mathbf{Z}}_j) \right) K_{t,u,h} \right) \right\|_\infty,$$

which is different from $S_{ij}(u, h)$ only in using $\tilde{\Sigma}_{ij}^{-1/2}$ instead of $\Sigma_{ij}^{-1/2}$ as a normalization matrix. As a reminder, we also state here the formula for Φ_T :

$$\Phi_T = \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} \{S_{ij}(u, h) - \lambda(h)\},$$

with

$$S_{ij}(u, h) = \left\| \Sigma_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \left(\Sigma_i^{1/2}(\mathbf{Z}_{it} - \bar{\mathbf{Z}}_i) - \Sigma_j^{1/2}(\mathbf{Z}_{jt} - \bar{\mathbf{Z}}_j) \right) K_{t,u,h} \right) \right\|_\infty.$$

Using this notation, we trivially have

$$|\tilde{\Psi}_T - \Phi_T| \leq |\tilde{\Psi}_T - \Phi_T^\diamond| + |\Phi_T^\diamond - \Phi_T|.$$

In what follows, we prove that

$$|\tilde{\Psi}_T - \Phi_T^\diamond| = o_p \left(\frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}} \right), \quad (\text{B.2})$$

$$|\Phi_T^\diamond - \Phi_T| = o_p(\rho_T \sqrt{\log T}). \quad (\text{B.3})$$

For the sake of simplicity, we introduce the following notation:

$$\begin{aligned} \mathbf{a}_{ij,T}(u, h) &= \frac{1}{\sqrt{Th}} \sum_{t=1}^T \{(\tilde{\mathbf{v}}_{it} - \tilde{\mathbf{v}}_i) - (\tilde{\mathbf{v}}_{jt} - \tilde{\mathbf{v}}_j)\} K_{t,u,h} \\ \mathbf{b}_{ij,T}(u, h) &= \frac{1}{\sqrt{Th}} \sum_{t=1}^T \{\Sigma_i^{1/2}(\mathbf{Z}_{it} - \bar{\mathbf{Z}}_i) - \Sigma_j^{1/2}(\mathbf{Z}_{jt} - \bar{\mathbf{Z}}_j)\} K_{t,u,h}. \end{aligned}$$

We then have that

$$\begin{aligned} \tilde{S}_{ij,T}(u, h) &= \|\tilde{\Sigma}_{ij}^{-1/2} \mathbf{a}_{ij,T}(u, h)\|_\infty, \\ S_{ij,T}^\diamond(u, h) &= \|\tilde{\Sigma}_{ij}^{-1/2} \mathbf{b}_{ij,T}(u, h)\|_\infty, \\ S_{ij,T}(u, h) &= \|\Sigma_{ij}^{-1/2} \mathbf{b}_{ij,T}(u, h)\|_\infty. \end{aligned}$$

First, we consider $|\tilde{\Psi}_T - \Phi_T^\diamond|$. Due to the triangle inequality,

$$\begin{aligned} |\tilde{S}_{ij,T}(u, h) - S_{ij,T}^\diamond(u, h)| &= \left| \left\| \tilde{\Sigma}_{ij}^{-1/2} \mathbf{a}_{ij,T}(u, h) \right\|_\infty - \left\| \tilde{\Sigma}_{ij}^{-1/2} \mathbf{b}_{ij,T}(u, h) \right\|_\infty \right| \\ &\leq \left\| \tilde{\Sigma}_{ij}^{-1/2} (\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h)) \right\|_\infty. \end{aligned}$$

Now using the definition of matrix-vector multiplication, we get the following

$$\begin{aligned} &\left\| \tilde{\Sigma}_{ij}^{-1/2} (\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h)) \right\|_\infty \\ &= \max_{d=1, \dots, D} \left| \left[\tilde{\Sigma}_{ij}^{-1/2} (\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h)) \right]_d \right| \\ &= \max_{d=1, \dots, D} \left| \sum_{e=1}^D \left[\tilde{\Sigma}_{ij}^{-1/2} \right]_{e,d} \left[\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h) \right]_e \right| \\ &\leq \max_{d=1, \dots, D} \sum_{e=1}^D \left| \left[\tilde{\Sigma}_{ij}^{-1/2} \right]_{e,d} \left[\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h) \right]_e \right| \\ &\leq \max_{e=1, \dots, D} \left| \left[\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h) \right]_e \right| \cdot \max_{d=1, \dots, D} \sum_{e=1}^D \left| \left[\tilde{\Sigma}_{ij}^{-1/2} \right]_{e,d} \right| \\ &= \left\| \mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h) \right\|_\infty \cdot \left\| \tilde{\Sigma}_{ij}^{-1/2} \right\|_1. \end{aligned}$$

Then,

$$\begin{aligned} |\tilde{\Psi}_T - \Phi_T^\diamond| &\leq \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} |\tilde{S}_{ij,T}(u, h) - S_{ij,T}^\diamond(u, h)| \\ &\leq \max_{1 \leq i < j \leq N} \left\| \tilde{\Sigma}_{ij}^{-1/2} \right\|_1 \max_{1 \leq i < j \leq N} \max_{(u, h) \in \mathcal{G}_T} \left\| \mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h) \right\|_\infty, \end{aligned} \tag{B.4}$$

Now, using summation by parts we get that

$$\begin{aligned} &\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h) \\ &= \frac{1}{\sqrt{Th}} \sum_{t=1}^T K_{t,u,h} \{ (\tilde{\mathbf{v}}_{it} - \tilde{\mathbf{v}}_i) - (\tilde{\mathbf{v}}_{jt} - \tilde{\mathbf{v}}_j) - \Sigma_i^{1/2}(\mathbf{Z}_{it} - \bar{\mathbf{Z}}_i) + \Sigma_j^{1/2}(\mathbf{Z}_{jt} - \bar{\mathbf{Z}}_j) \} \\ &= \frac{1}{\sqrt{Th}} \sum_{t=1}^{T-1} \mathbf{c}_{ij,t} (K_{t,u,h} - K_{t+1,u,h}) + \frac{1}{\sqrt{Th}} \mathbf{c}_{ij,T} K_{T,u,h}, \end{aligned}$$

where

$$\mathbf{c}_{ij,t} = \sum_{s=1}^t \{ (\tilde{\mathbf{v}}_{is} - \tilde{\mathbf{v}}_i) - (\tilde{\mathbf{v}}_{js} - \tilde{\mathbf{v}}_j) - \Sigma_i^{1/2}(\mathbf{Z}_{is} - \bar{\mathbf{Z}}_i) + \Sigma_j^{1/2}(\mathbf{Z}_{js} - \bar{\mathbf{Z}}_j) \}$$

and $\mathbf{c}_{ij,T} = \mathbf{0}$ for all pairs (i, j) by construction. From this, it follows that

$$\max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h)\|_\infty \leq \max_{(u,h) \in \mathcal{G}_T} W_T(u, h) \max_{1 \leq i < j \leq N} \max_{1 \leq t \leq T} \|\mathbf{c}_{ij,t}\|_\infty \quad (\text{B.5})$$

with

$$W_T(u, h) = \frac{1}{\sqrt{Th}} \sum_{t=1}^{T-1} |K_{t,u,h} - K_{t+1,u,h}|.$$

Due to the Lipschitz continuity of the kernel function $K(\cdot)$, we have that

$$\max_{(u,h) \in \mathcal{G}_T} W_T(u, h) = O\left(\frac{1}{\sqrt{T(h_T^{\min})^3}}\right). \quad (\text{B.6})$$

Using the triangular inequality yields that

$$\begin{aligned} \max_{1 \leq t \leq T} \|\mathbf{c}_{ij,t}\|_\infty &\leq \max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{is} - \Sigma_i^{1/2} \sum_{s=1}^t \mathbf{Z}_{is} \right\|_\infty + \max_{1 \leq t \leq T} \left\| t(\tilde{\mathbf{v}}_i - \Sigma_i^{1/2} \bar{\mathbf{Z}}_i) \right\|_\infty \\ &\quad + \max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{js} - \Sigma_j^{1/2} \sum_{s=1}^t \mathbf{Z}_{js} \right\|_\infty + \max_{1 \leq t \leq T} \left\| t(\tilde{\mathbf{v}}_j - \Sigma_j^{1/2} \bar{\mathbf{Z}}_j) \right\|_\infty \\ &\leq 2 \max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{is} - \Sigma_i^{1/2} \sum_{s=1}^t \mathbf{Z}_{is} \right\|_\infty + 2 \max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{js} - \Sigma_j^{1/2} \sum_{s=1}^t \mathbf{Z}_{js} \right\|_\infty \\ &= 2 \max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{is} - \Sigma_i^{1/2} IB(t) \right\|_\infty + 2 \max_{1 \leq t \leq T} \left\| \sum_{s=1}^t \tilde{\mathbf{v}}_{js} - \Sigma_j^{1/2} IB(t) \right\|_\infty \end{aligned}$$

Applying the strong approximation result (B.1) (which is formulated for the Euclidean norm, but the maximum norm is equivalent to it within the fixed number of dimensions D), we can infer that

$$\max_{1 \leq i < j \leq N} \max_{1 \leq t \leq T} \|\mathbf{c}_{ij,t}\|_\infty = o_p(T^{1/q}). \quad (\text{B.7})$$

Plugging (B.6) and (B.7) into (B.5), we obtain that

$$\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h)\|_\infty = o_p\left(\frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}}\right). \quad (\text{B.8})$$

Therefore, coming back to (B.4), we obtain that

$$\begin{aligned} |\tilde{\Psi}_T - \Phi_T^\diamond| &\leq \max_{1 \leq i < j \leq N} \|\tilde{\Sigma}_{ij}^{-1}\|_1 \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{a}_{ij,T}(u, h) - \mathbf{b}_{ij,T}(u, h)\|_\infty \\ &= O_p(1) \cdot o_p\left(\frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}}\right) = o_p\left(\frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}}\right), \end{aligned}$$

where we used the fact that $\max_{1 \leq i < j \leq N} \|\tilde{\Sigma}_{ij}^{-1/2}\|_1 = O_p(1)$, since Σ_i is positive definite with bounded from ∞ and from 0 eigenvalues for all i according to Assumption (A10).

Hence, we just proved (B.2), i.e., that

$$|\tilde{\Psi}_T - \Phi_T^\diamond| = o_p \left(\frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}} \right).$$

Next, we consider $|\Phi_T^\diamond - \Phi_T|$ in (B.3). Similarly to (B.4), we have that

$$\begin{aligned} |\Phi_T^\diamond - \Phi_T| &\leq \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} |S_{ij,T}^\diamond(u,h) - S_{ij,T}(u,h)| \\ &\leq \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \|(\tilde{\Sigma}_{ij}^{-1/2} - \Sigma_{ij}^{-1/2}) \mathbf{b}_{ij,T}(u,h)\|_\infty \\ &\leq \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{b}_{ij,T}(u,h)\|_\infty \max_{1 \leq i < j \leq N} \|\tilde{\Sigma}_{ij}^{-1/2} - \Sigma_{ij}^{-1/2}\|_1. \end{aligned}$$

First, we want to bound $\max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{b}_{ij,T}(u,h)\|_\infty$:

$$\begin{aligned} \left[\mathbf{b}_{ij,T}(u,h) \right]_d &= \left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T \Sigma_i^{1/2} (\mathbf{Z}_{it} - \bar{\mathbf{Z}}_i) K_{t,u,h} - \frac{1}{\sqrt{Th}} \sum_{t=1}^T \Sigma_j^{1/2} (\mathbf{Z}_{jt} - \bar{\mathbf{Z}}_j) K_{t,u,h} \right]_d \\ &= \left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T \left(K_{t,u,h} - \frac{1}{T} \sum_{s=1}^T K_{s,u,h} \right) \Sigma_i^{1/2} \mathbf{Z}_{it} \right]_d - \left[\frac{1}{\sqrt{Th}} \sum_{t=1}^T \left(K_{t,u,h} - \frac{1}{T} \sum_{s=1}^T K_{s,u,h} \right) \Sigma_j^{1/2} \mathbf{Z}_{jt} \right]_d \\ &\sim N \left(0, \frac{1}{Th} \sum_{t=1}^T \left(K_{t,u,h} - \frac{1}{T} \sum_{s=1}^T K_{s,u,h} \right)^2 [\Sigma_i + \Sigma_j]_{dd} \right). \end{aligned}$$

We want to use standard results on maxima of Gaussian random variables, and for that we need a uniform bound on the variance. Since $[\Sigma_i + \Sigma_j]_{dd} \leq C$ for fixed N and D , we want to show that $\frac{1}{Th} \sum_{t=1}^T \left(K_{t,u,h} - \frac{1}{T} \sum_{s=1}^T K_{s,u,h} \right)^2 = \frac{1}{Th} \sum_{t=1}^T K_{t,u,h}^2 - \frac{1}{T^2 h} \left(\sum_{t=1}^T K_{t,u,h} \right)^2 \leq C$

for all (u, h) . We can regard this value as a Riemann sum. Hence,

$$\begin{aligned}
& \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{1}{h} K^2 \left(\frac{t/T - u}{h} \right) - \lim_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T \frac{1}{\sqrt{h}} K \left(\frac{t/T - u}{h} \right) \right)^2 \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{1}{h} K^2 \left(\frac{t/T - u}{h} \right) - \left(\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{1}{h} K \left(\frac{t/T - u}{h} \right) \right)^2 \\
&= \int_0^1 \frac{1}{h} K^2 \left(\frac{z - u}{h} \right) dz - \left(\int_0^1 \frac{1}{\sqrt{h}} K \left(\frac{z - u}{h} \right) dz \right)^2 \\
&= \int_{-u/h}^{(1-u)/h} K^2(y) dy - h \left(\int_{-u/h}^{(1-u)/h} K(y) dy \right)^2 \\
&\leq \int_{-1}^1 K^2(y) dy < \infty,
\end{aligned}$$

where we used that K has a compact support $[-1, 1]$ and according to the Assumption (A13), $\int_{-1}^1 K^2(y) dy < \infty$, and for sufficiently large T , $\frac{1}{Th} \sum_{t=1}^T \left(K_{t,u,h} - \frac{1}{T} \sum_{s=1}^T K_{s,u,h} \right)^2 \leq 2 \int_{-1}^1 K^2(y) dy < C$ for all (u, h) . Therefore, all the coordinates of $\mathbf{b}_{ij,T}(u, h)$ are normally distributed random variables with bounded variance. By Assumption (A12) and standard results on maxima of Gaussian random variables, we have that

$$\max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{b}_{ij,T}(u, h)\|_\infty = O_p(\sqrt{\log T}). \quad (\text{B.9})$$

Hence, taking into account that $\|\tilde{\Sigma}_{ij}^{-1/2} - \Sigma_{ij}^{-1/2}\|_1 = o_p(\rho_T)$ for all i and j , we have that

$$\begin{aligned}
|\Phi_T^\diamond - \Phi_T| &\leq \max_{1 \leq i < j \leq N} \max_{(u,h) \in \mathcal{G}_T} \|\mathbf{b}_{ij,T}(u, h)\|_\infty \max_{1 \leq i < j \leq N} \|\tilde{\Sigma}_{ij}^{-1/2} - \Sigma_{ij}^{-1/2}\|_1 \\
&= O_p(\sqrt{\log T}) \cdot o_p(\rho_T) = o_p(\rho_T \sqrt{\log T}).
\end{aligned}$$

This completes the proof of (B.3). $\square$

Proof of Proposition A.2. The proof is an application of anti-concentration bounds for Gaussian random vectors. We in particular make use of the following anti-concentration inequality from Nazarov (2003), which can also be found as Lemma A.1 in Chernozhukov et al. (2017).

Lemma A.1. *Let $\mathbf{Z} = (Z_1, \dots, Z_p)^\top$ be a centered Gaussian random vector in $\mathbb{R}^p$ such that $\mathbb{E}[Z_j^2] \geq b$ for all $1 \leq j \leq p$ and some constant $b > 0$. Then for every $\mathbf{z} \in \mathbb{R}^p$ and $a > 0$,*

$$\mathbb{P}(\mathbf{Z} \leq \mathbf{z} + a) - \mathbb{P}(\mathbf{Z} \leq \mathbf{z}) \leq Ca\sqrt{\log p},$$

where the constant C only depends on b .

To apply this result, we introduce the following notation: We write $x = (u, h)$ and $\mathcal{G}_T = \{x_1, \dots, x_p\}$, where $p := |\mathcal{G}_T| = O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. For $k = 1, \dots, p$, $d = 1, \dots, D$ and $1 \leq i < j \leq N$, we further let

$$Z_{ij,2k-1,d} = [\Sigma_{ij}^{-1/2} \mathbf{b}_{ij,T}(x_{k1}, x_{k2})]_d \quad \text{and} \quad Z_{ij,2k,d} = -[\Sigma_{ij}^{-1/2} \mathbf{b}_{ij,T}(x_{k1}, x_{k2})]_d$$

along with $\lambda_{ij,2k-1,d} := \lambda_{ij,2k-1} = \lambda(x_{k2})$ and $\lambda_{ij,2k,d} := \lambda_{ij,2k} = \lambda(x_{k2})$, where $x_k = (x_{k1}, x_{k2})$. Under our assumptions, it holds that $\mathbb{E}[Z_{ij,l,d}] = 0$ and $\mathbb{E}[Z_{ij,l,d}^2] \geq b > 0$ for all i, j, l and d . First, as before,

$$\left[\Sigma_{ij}^{-1/2} \mathbf{b}_{ij,T}(x_{k1}, x_{k2}) \right]_d \sim N \left(0, \frac{2}{Tx_{k2}} \sum_{t=1}^T \left(K_{t,x_{k1},x_{k2}} - \frac{1}{T} \sum_{s=1}^T K_{s,x_{k1},x_{k2}} \right)^2 \right).$$

We want to show that $\frac{2}{Tx_{k2}} \sum_{t=1}^T \left(K_{t,x_{k1},x_{k2}} - \frac{1}{T} \sum_{s=1}^T K_{s,x_{k1},x_{k2}} \right)^2 = \frac{2}{Tx_{k2}} \sum_{t=1}^T K_{t,x_{k1},x_{k2}}^2 - \frac{2}{T^2 x_{k2}} \left(\sum_{t=1}^T K_{t,x_{k1},x_{k2}} \right)^2 \geq b$ for all k . As before, we regard this value as a Riemann sum:

$$\begin{aligned} & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{2}{x_{k2}} K^2 \left(\frac{t/T - x_{k1}}{x_{k2}} \right) - \lim_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T \frac{\sqrt{2}}{\sqrt{x_{k2}}} K \left(\frac{t/T - x_{k1}}{x_{k2}} \right) \right)^2 \\ &= 2 \int_{-x_{k1}/x_{k2}}^{(1-x_{k1})/x_{k2}} K^2(y) dy - 2x_{k2} \left(\int_{-x_{k1}/x_{k2}}^{(1-x_{k1})/x_{k2}} K(y) dy \right)^2 \\ &\geq 2 \int_{-4x_{k1}}^{4-4x_{k1}} K^2(y) dy - \frac{1}{2} \left(\int_{-\infty}^{\infty} K(y) dy \right)^2 \\ &\geq \int_{-1}^1 K^2(y) dy - \frac{1}{2} = 2b > 0, \end{aligned}$$

where we used the following facts: $x_{k2} \leq h_T^{\max} \leq 1/4$ as stated in Assumption (A12), K is a symmetric function with a compact support $[-1, 1]$, the interval $[-4x_{k1}, 4 - 4x_{k1}]$ fully covers either $[-1, 0]$ or $[0, 1]$ for all $x_{k1} \in [0, 1]$, and according to Assumption (A13), $\int_{-1}^1 K^2(y) dy - \frac{1}{2} > 0$. Therefore, for sufficiently large T , we have $\frac{2}{Tx_{k2}} \sum_{t=1}^T \left(K_{t,x_{k1},x_{k2}} - \frac{1}{T} \sum_{s=1}^T K_{s,x_{k1},x_{k2}} \right)^2 \geq \left(\int_{-1}^1 K^2(y) dy - \frac{1}{2} \right) / 2 = b > 0$.

We next construct the random vector $\mathbf{Z} = (Z_{ij,l,d} : 1 \leq i < j \leq N, 1 \leq l \leq 2p, 1 \leq d \leq D)$ by stacking the variables $Z_{ij,l,d}$ in a certain order (which can be chosen freely) and construct the vector $\boldsymbol{\lambda} = (\lambda_{ij,l,d} : 1 \leq i < j \leq N, 1 \leq l \leq 2p, 1 \leq d \leq D)$ in an analogous way. Since the variables $Z_{ij,l,d}$ are normally distributed, $\mathbf{Z}$ is a Gaussian random vector of length $(N-1)NpD$.

With this notation at hand, we can express the probability $\mathbb{P}(\Phi_T \leq q)$ as follows for each

$q \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}(\Phi_T \leq q) &= \mathbb{P}\left(\max_{1 \leq i < j \leq n} \max_{1 \leq l \leq 2p} \left\{ \max_{1 \leq d \leq D} Z_{ij,l,d} - \lambda_{ij,l} \right\} \leq q\right) \\
&= \mathbb{P}\left(\max_{1 \leq i < j \leq n} \max_{1 \leq l \leq 2p} \max_{1 \leq d \leq D} \{Z_{ij,l,d} - \lambda_{ij,l,d}\} \leq q\right) \\
&= \mathbb{P}(Z_{ij,l,d} \leq \lambda_{ij,l,d} + q \text{ for all } (i, j, l, d)) = \mathbb{P}(\mathbf{Z} \leq \boldsymbol{\lambda} + q).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{P}(|\Phi_T - x| \leq \delta_T) &= \mathbb{P}(x - \delta_T \leq \Phi_T \leq x + \delta_T) \\
&= \mathbb{P}(\Phi_T \leq x + \delta_T) - \mathbb{P}(\Phi_T \leq x) \\
&\quad + \mathbb{P}(\Phi_T \leq x) - \mathbb{P}(\Phi_T \leq x - \delta_T) \\
&= \mathbb{P}(\mathbf{Z} \leq \boldsymbol{\lambda} + x + \delta_T) - \mathbb{P}(\mathbf{Z} \leq \boldsymbol{\lambda} + x) \\
&\quad + \mathbb{P}(\mathbf{Z} \leq \boldsymbol{\lambda} + x) - \mathbb{P}(\mathbf{Z} \leq \boldsymbol{\lambda} + x - \delta_T) \\
&\leq 2C\delta_T \sqrt{\log((N-1)NpD)},
\end{aligned}$$

where the last line is by Lemma A.1. With $\delta_T = \frac{T^{1/q}}{\sqrt{T(h_T^{\min})^3}} + \rho_T \sqrt{\log T}$, this immediately implies Proposition A.2. $\square$

Proof of Proposition 4.2. To proof the result, we first show that

$$\mathbb{P}(\Phi_T \leq q_T(\alpha)) = 1 - \alpha. \tag{B.10}$$

We prove (B.10) by contradiction. Suppose (B.10) is false. By definition of $q_T(\alpha)$, it holds that $\mathbb{P}(\Phi_T \leq q_T(\alpha)) \geq 1 - \alpha$, and thus, there exists $\xi > 0$ such that $\mathbb{P}(\Phi_T \leq q_T(\alpha)) = 1 - \alpha + \xi$. Following the proof of Proposition A.2, we have that for any $\delta > 0$,

$$\mathbb{P}(\Phi_T \leq q_T(\alpha)) - \mathbb{P}(\Phi_T \leq q_T(\alpha) - \delta) \leq 2C\delta \sqrt{\log((N-1)NpD)}.$$

Combining this result with $\mathbb{P}(\Phi_T \leq q_T(\alpha)) = 1 - \alpha + \xi$ yields

$$\mathbb{P}(\Phi_T \leq q_T(\alpha) - \delta) \geq (1 - \alpha) + \xi - 2C\delta \sqrt{\log((N-1)NpD)} > (1 - \alpha),$$

for $\delta > 0$ small enough. This contradicts the definition of $q_T(\alpha)$, saying that $q_T(\alpha) := \inf_{q \in \mathbb{R}} \{\mathbb{P}(\Phi_T \leq q) \geq (1 - \alpha)\}$. Therefore, by contradiction, (B.10) holds. The result of

Proposition 4.2 follows directly now by applying Theorem 4.1.

$$\begin{aligned}
|\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) - (1 - \alpha)| &\stackrel{\text{(B.10)}}{=} |\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) - \mathbb{P}(\Phi_T \leq q_T(\alpha))| \\
&\leq \sup_{y \in \mathbb{R}} |\mathbb{P}(\widehat{\Psi}_T \leq y) - \mathbb{P}(\Phi_T \leq y)| \\
&\stackrel{\text{Thm. 4.1}}{=} o(1).
\end{aligned}$$

□

Proof of Proposition 4.3. By assumption, there exists a pair (i_0, j_0) and $(u_0, h_0) \in \mathcal{G}_T$ with $[u_0 - h_0, u_0 + h_0] \subset [0, 1]$ assume that there exists at least one coordinate $d_0 \in \{1, \dots, D\}$ such that

$$[\beta_{i_0, T}(w) - \beta_{j_0, T}(w)]_{d_0} \geq c_T \sqrt{\frac{\log T}{Th}} \quad \text{for all } w \in [u_0 - h_0, u_0 + h_0]. \quad (\text{B.11})$$

where $c_T \rightarrow \infty$ as $T \rightarrow \infty$. Since

$$\widehat{\Psi}_T \geq \hat{S}_{i_0 j_0}(u_0, h_0) - \lambda(h_0),$$

we obtain

$$\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) \leq \mathbb{P}(\hat{S}_{i_0 j_0}(u_0, h_0) - \lambda(h_0) \leq q_T(\alpha)).$$

Thus, it suffices to show that

$$\mathbb{P}(\hat{S}_{i_0 j_0}(u_0, h_0) - \lambda(h_0) \leq q_T(\alpha)) = o(1).$$

For notational simplicity, we write $(i, j, u, h) = (i_0, j_0, u_0, h_0)$ as follows.

Using the representation in Section 3.3,

$$\hat{S}_{ij}(u, h) = \left\| \hat{\Sigma}_{ij}^{-1/2} \left(\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \left(\mathbf{X}_t' \beta_{i, T}(t/T) + \varepsilon_{it} - \mathbf{X}_t' \beta_{j, T}(t/T) - \varepsilon_{jt} \right) K_{t, u, h} \right) \right\|_{\infty}.$$

Hence, we can write

$$\hat{S}_{ij}(u, h) = \left\| A_T(u, h) + B_T(u, h) \right\|_{\infty},$$

with

$$A_T(u, h) := \hat{\Sigma}_{ij}^{-1/2} \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' (\beta_{i, T}(t/T) - \beta_{j, T}(t/T)) K_{t, u, h},$$

$$B_T(u, h) := \hat{\Sigma}_{ij}^{-1/2} \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t (\varepsilon_{it} - \varepsilon_{jt}) K_{t, u, h}.$$

The term $A_T(u, h)$ represents the deterministic “signal” induced by $\beta_i(w) - \beta_j(w)$, and $B_T(u, h)$ is the stochastic “noise” due to the errors.

For fixed (i, j, u, h) , the random vector

$$\frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t (\varepsilon_{it} - \varepsilon_{jt}) K_{t,u,h}$$

is a kernel-weighted sum of mean-zero weakly dependent random vectors with finite q -th moments, by Assumptions (A5) - (A9). Standard moment bounds for such sums (as used in the proof of Theorem 4.1) yield that its Euclidean norm is $O_p(1)$, and hence its ℓ_∞ -norm is also $O_p(1)$.

The long-run covariance estimators satisfy

$$\|\widehat{\Sigma}_{ij}^{-1/2} - \Sigma_{ij}^{-1/2}\|_1 = o_p(1),$$

and Assumption (A10) implies that the eigenvalues of Σ_{ij} are bounded away from 0 and ∞ . Thus, multiplication by $\widehat{\Sigma}_{ij}^{-1/2}$ preserves stochastic order, and we obtain

$$\|B_T(u, h)\|_\infty = O_p(1). \quad (\text{B.12})$$

Now we look at $A_T(u, h)$. By Assumption (A13), there exists $c_K > 0$ such that for all w with $|w - u| \leq h\delta_K$,

$$K\left(\frac{w - u}{h}\right) \geq c_K. \quad (\text{B.13})$$

Consider the signal term before covariance normalization

$$C_T(u, h) := \frac{1}{\sqrt{Th}} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' (\beta_{i,T}(t/T) - \beta_{j,T}(t/T)) K_{t,u,h}.$$

Since $K_{t,u,h}$ vanishes outside $[u - h, u + h]$, at most $O(Th)$ summands are nonzero; the factor $1/\sqrt{Th}$ therefore yields an overall scale $\sqrt{Th}$.

By Assumption (A11) and standard arguments for kernel-weighted laws of large numbers,

$$\frac{1}{Th} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' K_{t,u,h} \rightarrow \Gamma(u, h),$$

where $\Gamma(u, h)$ is positive definite with eigenvalues bounded away from 0 and ∞ .

Using (B.11)-(B.13), the boundedness of $\beta_i - \beta_j$ (from the smoothness of β_i and β_j), and a

Riemann-sum approximation, we obtain

$$\left\| \frac{1}{Th} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' (\beta_{i,T}(t/T) - \beta_{j,T}(t/T)) K_{t,u,h} \right\|_{\infty} \geq c_1 c_T \sqrt{\frac{\log T}{Th}}$$

for some constant $c_1 > 0$ and all sufficiently large T , with probability tending to one. Hence

$$\begin{aligned} \|C_T(u, h)\|_{\infty} &= \sqrt{Th} \left\| \frac{1}{Th} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t' (\beta_{i,T}(t/T) - \beta_{j,T}(t/T)) K_{t,u,h} \right\|_{\infty} \\ &\geq c_1 \sqrt{Th} c_T \sqrt{\frac{\log T}{Th}} \\ &= c_1 c_T \sqrt{\log T} \end{aligned}$$

with probability tending to one.

As before, multiplication by $\widehat{\Sigma}_{ij}^{-1/2}$ preserves the stochastic order, hence, for some constant $c_2 > 0$,

$$\|\widehat{\Sigma}_{ij}^{-1/2} v\|_{\infty} \geq c_2 \|v\|_{\infty} \quad \text{for all } v \in \mathbb{R}^D,$$

with probability tending to one. Applying this to $v = C_T(u, h)$ and combining with the bound on $\|C_T(u, h)\|_{\infty}$ yields

$$\|A_T(u, h)\|_{\infty} = \|\widehat{\Sigma}_{ij}^{-1/2} C_T(u, h)\|_{\infty} \geq c_3 c_T \sqrt{\log T} \quad (\text{B.14})$$

with probability tending to 1 for some constant $c_3 > 0$.

By the triangle inequality and (B.12), (B.14),

$$\hat{S}_{ij}(u, h) = \|A_T(u, h) + B_T(u, h)\|_{\infty} \geq \|A_T(u, h)\|_{\infty} - \|B_T(u, h)\|_{\infty} \geq c_3 c_T \sqrt{\log T} - O_p(1).$$

Thus, for any $\varepsilon > 0$, there exists T_0 such that for all $T \geq T_0$,

$$\mathbb{P}\left(\hat{S}_{ij}(u, h) \geq \frac{1}{2} c_3 c_T \sqrt{\log T}\right) \geq 1 - \varepsilon. \quad (\text{B.15})$$

The multiscale correction term is $\lambda(h) = \sqrt{2 \log 1/(2h)}$. Note that for some sufficiently large constant C we have

$$\lambda(h) = \sqrt{2 \log \{1/(2h)\}} \leq \sqrt{2 \log \{1/(2h_{\min, T})\}} \leq C \sqrt{\log T}. \quad (\text{B.16})$$

The critical value $q_T(\alpha)$ is the $(1 - \alpha)$ -quantile of the Gaussian max-type statistic Φ_T defined

in (10). Each component of Φ_T is the ℓ_∞ -norm of a Gaussian vector with variance uniformly bounded and bounded away from zero, and the number of components is of order $|\mathcal{G}_T| \times N^2 \times D = O(T^\theta)$ for some $\theta > 0$ by Assumption (A12). Standard Gaussian tail bounds for maxima imply that

$$q_T(\alpha) = O(\sqrt{\log T}). \quad (\text{B.17})$$

Combining (B.16) and (B.17), there exists $C_4 > 0$ such that

$$q_T(\alpha) + \lambda(h) \leq C_4 \sqrt{\log T} \quad \text{for all sufficiently large } T. \quad (\text{B.18})$$

From (B.15) and (B.18), and since $c_T \rightarrow \infty$, we can choose T large enough such that

$$\frac{1}{2} c_3 c_T > C_4,$$

which implies

$$\mathbb{P}(\hat{S}_{ij}(u, h) - \lambda(h) > q_T(\alpha)) \rightarrow 1.$$

Equivalently,

$$\mathbb{P}(\hat{S}_{ij}(u, h) - \lambda(h) \leq q_T(\alpha)) = o(1),$$

which yields the necessary statement. □

Proof of Proposition 4.4. The proof of this result follows the steps in the proof of Proposition 4.5 in Khismatullina and Vogt (2022) and therefore omitted for brevity. □

References in Appendix

- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, **45** 2309–2352.
- KARMAKAR, S. and WU, W. B. (2020). Optimal Gaussian approximation for multiple time series. *Statistica Sinica*, **30** 1399–1417.
- KHISMATULLINA, M. and VOGT, M. (2022). Multiscale comparison of nonparametric trend curves. *arXiv preprint arXiv:2209.10841*.
- NAZAROV, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. Springer, 169–187.