1 State of the art and preliminary work

The comparison of nonparametric curves is a classic topic in econometrics and statistics. Depending on the specific application, the curves of interest are densities, distribution functions, time trends or regression curves. The problem of testing for equality of densities has been studied in Mammen (1992), Anderson et al. (1994) and Li et al. (2009) among others. Tests for equality of distribution functions can be found for example in Kiefer (1959), Anderson (1962) and Finner and Gontscharuk (2018). Tests for equality of trend or regression curves have been developed in Härdle and Marron (1990), Hall and Hart (1990), Delgado (1993), Degras et al. (2012), Zhang et al. (2012) and Hidalgo and Lee (2014) among many others. In the proposed project, we focus on the comparison of nonparametric trend curves.

The statistical problem of comparing trends has a wide range of applications in economics, finance and other fields such as climatology and biology. In economics, one may wish is to compare trends in real gross domestic product (GDP) across different countries (cp. Grier and Tullock, 1989). Another example concerns the dynamics of long-term interest rates. To better understand these dynamics, researchers aim to compare the yields of US Treasury bills at different maturities over time (cp. Park et al., 2009). In finance, it is of interest to compare the volatility trends of different stocks (cp. Nyblom and Harvey, 2000). Finally, in climatology, researchers are interested in comparing the trending behaviour of temperature time series across different spatial locations (cp. Karoly and Wu, 2005).

Classically, time trends are modelled stochastically in econometrics; see e.g. Stock and Watson (1988). Recently, however, there has been a growing interest in econometric models with deterministic time trends; see Cai (2007), Atak et al. (2011), Robinson (2012) and Chen et al. (2012) among others. Non- and semiparametric trend modelling has attracted particular interest in a panel data context. Li et al. (2010), Atak et al. (2011), Robinson (2012) and Chen et al. (2012) considered panel models where the observed time series have a common time trend. In many applications, however, the assumption of a common time trend is quite harsh. In particular when the number of observed time series is large, it is quite natural to suppose that the time trend may differ across time series. More flexible panel settings with heterogeneous trends have been studied, for example, in Zhang et al. (2012) and Hidalgo and Lee (2014).

In what follows, we consider a general panel framework with heterogeneous trends which is useful for a number of economic and financial applications: Suppose we observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \le t \le T\}$ for $1 \le i \le n$, where Y_{it} are real-valued random variables and $\mathbf{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ are d-dimensional random vectors. Each time series $\mathcal{Z}_i$ is modelled by the equation

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + \alpha_i + \varepsilon_{it} \tag{1}$$

for $1 \le t \le T$, where $m_i : [0,1] \to \mathbb{R}$ is a nonparametric (deterministic) trend function, $\mathbf{X}_{it}$ is a vector of regressors or controls and $\boldsymbol{\beta}_i$ is the corresponding parameter vector. Moreover, α_i are so-called fixed effect error terms and ε_{it} are standard regression errors with $\mathbb{E}[\varepsilon_{it}|\mathbf{X}_{it}] = 0$ for all t. Model (1) nests a number of panel models which have recently been considered in the literature. Special cases of model (1) with a nonparametric trend specification are for example considered in Atak et al. (2011), Zhang et al.

(2012) and Hidalgo and Lee (2014). Versions of model (1) with a parametric trend are studied in Vogelsang and Franses (2005), Sun (2011) and Xu (2012) among others.

Within the general framework of model (1), we can formulate a number of interesting statistical questions concerning the set of trend functions $\{m_i : 1 \le i \le n\}$.

(a) Testing for equality of nonparametric trend curves

In many application contexts, an important question is whether the time trends m_i in model (1) are all the same. Put differently, the question is whether the observed time series have a common trend. This question can formally be addressed by a statistical test of the null hypothesis

$$H_0 : \text{There exists a function } m : [0,1] \to \mathbb{R} \text{ such that } m_i = m \text{ for all } 1 \le i \le n.$$

A closely related question is whether all time trends have the same parametric form. To formulate the corresponding null hypothesis, let $m(\theta, \cdot) : [0,1] \to \mathbb{R}$ be a function which is known up to the finite-dimensional parameter $\theta \in \Theta$, where Θ denotes the parameter space. The null hypothesis of interest now reads as follows:

$$H_{0,\text{para}} : \text{ There exists } \theta \in \Theta \text{ such that } m_i(\cdot) = m(\theta, \cdot) \text{ for all } 1 \le i \le n.$$

If $m(\theta, w) = a + bw$ with $\theta = (a, b)$, for example, then H_0 is the hypothesis that all trends m_i are linear with the same intercept a and slope b. A somewhat simpler but yet important hypothesis is given by

$$H_{0,\text{const}} : m_i \equiv 0 \text{ for all } 1 \le i \le n.$$

Under this hypothesis, there is no time trend at all in the observed time series. Put differently, all the time trends m_i are constant. (Note that under the normalization constraint $\int_0^1 m_i(w)dw = 0$, m_i must be equal to zero if it is a constant function.) A major goal of our project is to develop new tests for the hypotheses H_0, $H_{0,\text{para}}$ and $H_{0,\text{const}}$ in model (1). In order to keep the exposition as clear as possible, we focus attention on the hypothesis H_0 in what follows. Tests of $H_{0,\text{para}}$, $H_{0,\text{const}}$ and related hypotheses have for example been studied in Lyubchich and Gel (2016) and Chen and Wu (2018).

In recent years, a number of different approaches have been developed to test the hypothesis H_0. Degras et al. (2012) consider the problem of testing H_0 within the model framework

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \qquad (1 \le t \le T,\, 1 \le i \le n), \tag{2}$$

where $\mathbb{E}[\varepsilon_{it}] = 0$ for all i and t and the terms α_i are assumed to be deterministic. Obviously, (2) is a special case of (1) which does not include additional regressors. Degras et al. (2012) construct an L_2-type statistic to test H_0. The statistic is based on the difference between estimators of the trend with and without imposing H_0. Let $\hat{m}_{i,h}$ be the estimator of m_i and $\hat{m}_h$ the estimator of the common trend m under H_0, where h denotes the bandwidth parameter. With these estimators, the authors define the statistic

$$\Delta_{n,T} = \sum_{i=1}^{n} \int_0^1 \big(\hat{m}_{i,h}(u) - \hat{m}_h(u)\big)^2 du, \tag{3}$$

which measures the L_2-distance between $\hat{m}_{i,h}$ and $\hat{m}_h$. In the theoretical part of their paper, they derive the limit distribution of $\Delta_{n,T}$. Chen and Wu (2018) develop theory for test statistics closely related to those from Degras et al. (2012), but under more general conditions on the error terms.

Zhang et al. (2012) investigate the problem of testing the hypothesis H_0 in a slightly restricted version of model (1), where $\beta_i = \beta$ for all i. The regression coefficients β_i are thus assumed to be homogeneous in their setting. They construct a residual-based test statistic as follows: First, they obtain profile least squares estimators $\hat{\beta}$ and $\hat{m}_h(t/T)$ of the parameter vector β and the common trend m under H_0, where h denotes the bandwidth. With these estimators, they compute the residuals $\hat{u}_{it} = Y_{it} - \hat{\beta}^T X_{it} - \hat{m}_h(t/T)$. These residuals are shown to have the form $\hat{u}_{it} = \Delta_i(t/T) + \eta_{it}$, where Δ_i is a deterministic function with the property that $\Delta_i \equiv 0$ under H_0 and η_{it} denotes the error term. Testing H_0 is thus equivalent to testing the hypothesis $H_0' : \Delta_i \equiv 0$ for all $1 \le i \le n$. The authors construct a test statistic for the hypothesis H_0' on the basis of nonparametric kernel estimators of the functions Δ_i and derive its limit distribution.

The tests of Zhang et al. (2012), Degras et al. (2012) and Chen and Wu (2018) are based on nonparametric estimators of the trend functions m_i that depend on one or several bandwidth parameters. Unfortunately, it is far from clear how to choose these bandwidths in an appropriate way. This is a general problem concerning essentially all tests based on nonparametric curve estimators. There are of course many theoretical results on optimal bandwidth choice for estimation purposes. However, the optimal bandwidth for curve estimation is usually not optimal for testing. Optimal bandwidth choice for tests is indeed an open problem, and only little theory for simple cases is available (cp. Gao and Gijbels, 2008). Since tests based on nonparametric curve estimators are commonly quite sensitive to the choice of bandwidth and theory for optimal bandwidth selection is not available, it appears preferable to work with bandwidth-free tests.

A classical way to obtain a bandwidth-free test of the hypothesis H_0 is to use CUSUM-type statistics which are based on partial sum processes. This approach is taken in Hidalgo and Lee (2014). A more modern approach to obtain a bandwidth-free test is to employ multiscale methods. These methods avoid the need to choose a bandwidth by considering a large collection of bandwidths simultaneously. More specifically, the basic idea is as follows: Let S_h be a test statistic for the null hypothesis of interest, which depends on the bandwidth h. Rather than considering only a single statistic S_h for a specific bandwidth h, a multiscale approach simultaneously considers a whole family of statistics $\{S_h : h \in \mathcal{H}\}$, where $\mathcal{H}$ is a set of bandwidth values. The multiscale test then proceeds as follows: For each bandwidth or scale h, one checks whether $S_h > q_h(\alpha)$, where $q_h(\alpha)$ is a bandwidth-dependent critical value (for given significance level α). The multiscale test rejects if $S_h > q_h(\alpha)$ for at least one scale h. The main theoretical difficulty in this approach is of course to derive appropriate critical values $q_h(\alpha)$. Specifically, the critical values $q_h(\alpha)$ need to be determined such that the multiscale test has the correct (asymptotic) level, that is, such that $\mathbb{P}(S_h > q_h(\alpha)$ for some $h \in \mathcal{H}) = (1 - \alpha) + o(1)$.

Multiscale methods have been developed for a variety of different test problems in recent years. Chaudhuri and Marron (1999, 2000) introduced the so-called SiZer method which has been extended in various directions; see for example Hannig and

Marron (2006) and Rondonotti et al. (2007). Horowitz and Spokoiny (2001) proposed a multiscale test for the parametric form of a regression function. Dümbgen and Spokoiny (2001) constructed a multiscale approach which works with additively corrected supremum statistics. This general approach has been very influential in recent years and has been further developed in numerous ways; see for example Dümbgen (2002), Rohde (2008) and Proksch et al. (2018) for multiscale methods in the regression context and Dümbgen and Walther (2008), Rufibach and Walther (2010), Schmidt-Hieber et al. (2013) and Eckle et al. (2017) for methods in the context of density estimation. Importantly, all of these studies are restricted to the case of independent data. It turns out that it is highly non-trivial to extend the multiscale approach of Dümbgen and Spokoiny (2001) to the case of dependent data. A first step into this direction has recently been made in Khismatullina and Vogt (2020). They developed multiscale methods to test for local increases/decreases of the nonparametric trend function m in the univariate time series model $Y_t = m(t/T) + \varepsilon_t$.

To the best of our knowledge, multiscale tests of the hypotheses H_0, $H_{0,\text{para}}$ and $H_{0,\text{const}}$ in model (1) are not available in the literature. The only exception is Park et al. (2009) who developed SiZer methods for the comparison of nonparametric trend curves in a strongly simplified version of model (1). Their analysis, however, is mainly methodological and not fully backed up by theory. Indeed, theory has only been derived for the special case $n = 2$, that is, for the case that only two time series are observed.

(b) Clustering of nonparametric trend curves

Consider the situation that the null hypothesis $H_0 : m_1 = \ldots = m_n$ is violated in the general panel data model (1). Even though some of the trend functions m_i are different in this case, there may still be groups of time series with the same time trend. Formally, a group stucture can be defined as follows within the framework of model (1): There exist sets or groups of time series $G_1, \ldots, G_{K_0}$ with $\{1, \ldots, n\} = \dot{\bigcup}_{k=1}^{K_0} G_k$ such that for each $1 \le k \le K_0$,

$$m_i = m_j \quad \text{for all } i, j \in G_k. \tag{4}$$

According to (4), the time series of a given group G_k all have the same time trend. In many applications, it is very natural to suppose that there is such a group structure in the data. An interesting statistical problem which we aim to investigate in our project is how to estimate the unknown groups $G_1, \ldots, G_{K_0}$ and their unknown number K_0 from the data.

Several approaches to this problem have been proposed in the context of models closely related to (1). Degras et al. (2012) used a repeated testing procedure based on L_2-type test statistics of the form (3) in order to estimate the unknown group structure in model (2). Zhang (2013) developed a clustering method within the same model framework which makes use of an extended Bayesian information criterion. Vogt and Linton (2017) constructed a thresholding method to estimate the unknown group structure in the panel model $Y_{it} = m_i(X_{it}) + u_{it}$, where X_{it} are random regressors and u_{it} are general error terms that may include fixed effects. Their approach can also be adapted to the case of fixed regressors $X_{it} = t/T$. As an alternative to a group structure, factor-type structures have been imposed on the trend and regression functions in panel models. Such factor-type structures are studied in Kneip et al. (2012), Boneva et al. (2015) and Boneva et al. (2016) among others.

The problem of estimating the unknown groups $G_1, \dots, G_{K_0}$ and their unknown number K_0 in model (1) has close connections to functional data clustering. There, the aim is to cluster smooth random curves that are functions of (rescaled) time and that are observed with or without noise. A number of different clustering approaches have been proposed in the context of functional data models; see for example Abraham et al. (2003), Tarpey and Kinateder (2003) and Tarpey (2007) for procedures based on k-means clustering, James and Sugar (2003) and Chiou and Li (2007) for model-based clustering approaches and Jacques and Preda (2014) for a recent survey.

The problem of finding the unknown group structure in model (1) is also closely related to a developing literature in econometrics which aims to identify unknown group structures in parametric panel regression models. In its simplest form, the panel regression model under consideration is given by the equation $Y_{it} = \beta_i^\top X_{it} + u_{it}$ for $1 \le t \le T$ and $1 \le i \le n$, where the coefficient vectors β_i are allowed to vary across individuals i and the error terms u_{it} may include fixed effects. Similar to the trend functions in model (1), the coefficients β_i are assumed to belong to a number of groups: there are K_0 groups $G_1, \dots, G_{K_0}$ such that $\beta_i = \beta_j$ for all $i, j \in G_k$ and all $1 \le k \le K_0$. The problem of estimating the unknown groups and their unknown number has been studied in different versions of this modelling framework; cp. Su et al. (2016), Su and Ju (2018) and Wang et al. (2018) among others. Bonhomme and Manresa (2015) considered a related model where the group structure is not imposed on the regression coefficients but rather on some unobserved time-varying fixed effect components of the panel model.

Virtually all the proposed procedures to cluster nonparametric curves in panel and functional data models related to (1) depend on a number of bandwidth or smoothing parameters required to estimate the nonparametric functions m_i. In general, nonparametric curve estimators are strongly affected by the chosen bandwidth parameters. A clustering procedure which is based on such estimators can be expected to be strongly influenced by the choice of bandwidths as well. Moreover, as in the context of statistical testing, there is no theory available on how to pick the bandwidths optimally for the clustering problem. Hence, as in the context of testing, it is desirable to construct a clustering procedure which is free of bandwidth or smoothing parameters that need to be selected.

One way to obtain a clustering method which does not require to select any bandwidth parameter is to use multiscale methods. This approach has recently been taken in **?**. They develop a clustering approach in the context of the panel model $Y_{it} = m_i(X_{it}) + u_{it}$, where X_{it} are random regressors and u_{it} are general error terms that may include fixed effects. Imposing the same group structure as in (4) on their model, they construct estimators of the unknown groups and their unknown number as follows: In a first step, they develop bandwidth-free multiscale statistics $\hat{d}_{ij}$ which measure the distance between pairs of functions m_i and m_j. To construct them, they make use of the multiscale testing methods described in part (a) of this section. In a second step, the statistics $\hat{d}_{ij}$ are employed as dissimilarity measures in a hierarchical clustering algorithm.

2 The model

The model setting is as follows. We observe time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ of length T for $1 \le i \le n$. Each time series $\mathcal{Y}_i$ satisfies the model equation

$$Y_{it} = \beta' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{5}$$

for $1 \le t \le T$, where β is a $d \times 1$ vector of unknown parameters, X_{it} is a $d \times 1$ vector of individual covariates, m_i is an unknown nonparametric trend function defined on $[0,1]$, α_i is a (deterministic or random) intercept term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ is a zero-mean stationary error process. For identification, we normalize the functions m_i such that $\int_0^1 m_i(u)du = 0$ for all $1 \le i \le n$. The term α_i can also be regarded as an additional error component. In the econometrics literature, it is commonly called a fixed effect error term. It can be interpreted as capturing unobserved characteristics of the time series $\mathcal{Y}_i$ which remain constant over time. We allow the error terms α_i to be dependent across i in an arbitrary way. Hence, by including them in model equation (5), we allow the n time series $\mathcal{Y}_i$ in our sample to be correlated with each other. Whereas the terms α_i may be correlated, the error processes $\mathcal{E}_i$ are assumed to be independent across i. In addition, each process $\mathcal{E}_i$ is supposed to satisfy the conditions (C1)–(C3). Finally note that throughout the paper, we restrict attention to the case where the number of time series n in model (1) is fixed. Extending our theoretical results to the case where n slowly grows with the sample size T is a possible topic for further research.

The stationary error processes $\{\varepsilon_{it}\}_{t\in\mathbb{Z}}$ are assumed to have the following properties:

(C1) For each i the variables ε_{it} allow for the representation $\varepsilon_{it} = G_i(\ldots, \eta_{it-1}, \eta_{it}, \eta_{it+1}, \ldots)$, where η_{it} are i.i.d. random variables across t and $G_i : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. Denote $Eta_{it} = i(\ldots, \eta_{it-1}, \eta_{it})$.

(C2) It holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$, where $\|\varepsilon_{it}\|_q = (\mathbb{E}|\varepsilon_t|^q)^{1/q}$.

Following Wu (2005), we impose conditions on the dependence structure of the error process $\{\varepsilon_{it}\}_{t\in\mathbb{Z}}$ in terms of the physical dependence measure $d_{i,t,q} = \|\varepsilon_{it} - \varepsilon'_{it}\|_q$, where $\varepsilon'_{it} = G(\ldots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \ldots, \eta_{it-1}, \eta_{it}, \eta_{it+1}, \ldots)$ with $\{\eta'_{it}\}$ being an i.i.d. copy of $\{\eta_{it}\}$. In particular, we assume the following:

(C3) Define $\Theta_{i,t,q} = \sum_{|s|\ge t} d_{i,s,q}$ for $t \ge 0$. It holds that $\Theta_{i,t,q} = O(t^{-\tau_q}(\log t)^{-A})$, where $A > \frac{2}{3}(1/q + 1 + \tau_q)$ and $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$.

The conditions (C1)–(C3) are fulfilled by a wide range of stationary processes $\{\varepsilon_{it}\}_{t\in\mathbb{Z}}$.

3 Testing for equality of time trends

In this section, we adapt the multiscale method developed in Section **??** to test the hypothesis that the trend functions in model (1) are all the same. More formally, we test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ in model (5). As we will see, the proposed multiscale method does not only allow to test whether the null hypothesis is violated. It also provides information on where violations occur. More specifically, it allows to identify, with a pre-specified confidence, (i) trend functions which are different from each other and (ii) time intervals where these trend functions differ.

3.1 Construction of the test statistic in the presence of exogenous regressors

We now extend the model (1) to include the exogenous regressors:

$$Y_{it} = \beta_i' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{6}$$

It is obvious that if β_i is known, the problem of testing for the common time trend would be reduced to the one discussed before. That is, we would test $H_0 : m_1 = m_2 = \ldots = m_n$ in the model

$$\begin{aligned} Y_{it} - \beta_i' X_{it} &=: V_{it} \\ &= m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \end{aligned}$$

replacing Y_{it} by V_{it} in the construction of the test statistic. However, β_i is not known so we need to estimate it first. Given an estimator $\widehat{\beta}_i$, we then consider

$$\widehat{V}_{it} := Y_{it} - \widehat{\beta}_i' X_{it} = (\widehat{\beta}_i - \beta_i)' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it}$$

and construct the kernel averages $\widehat{\psi}_{ij,T}(u,h)$ based on $\widehat{V}_{it}$ instead of $\widehat{Y}_{it}$. Specifically, for any pair of time series i and j we define the kernel averages

$$\widehat{\psi}'_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)(\widehat{V}_{it} - \widehat{V}_{jt})$$

with the kernel weights defined in ??. Similar as in Section **??**, we aggregate the kernel averages $\widehat{\psi}'_{ij,T}(u,h)$ for all $(u,h) \in \mathcal{G}_T$ by the multiscale test statistic

$$\widehat{\Psi}'_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}'_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h) \Big\}.$$

We now focus on finding an appropriate estimator $\widehat{\beta}$ of β. For that purpose, we consider the time series $\{\Delta Y_{it}\}$ of the differences $\Delta Y_{it} = Y_{it} - Y_{it-1}$ for each i. We then have

$$\Delta Y_{it} = Y_{it} - Y_{it-1} = \beta_i' \Delta X_{it} + \bigg(m_i\Big(\frac{t}{T}\Big) - m_i\Big(\frac{t-1}{T}\Big) \bigg) + \Delta\varepsilon_{it},$$

where $\Delta X_{it} = X_{it} - X_{it-1}$ and $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Since $m_i(\cdot)$ is Lipschitz, we can use the fact that $\big|m_i\big(\frac{t}{T}\big) - m_i\big(\frac{t-1}{T}\big)\big| = O\big(\frac{1}{T}\big)$ and rewrite

$$\Delta Y_{it} = \beta_i' \Delta X_{it} + \Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big).$$

For a class of stochastic processes $\{\mathbf{L}(v, \mathcal{F}_t)\}_{t\in\mathbb{Z}}$, we say that the process is $\mathcal{L}^q$ *stochastic Lipschitz-continuous* over $[0,1]$ if

$$\sup_{0\le v_1 < v_2 \le 1} \frac{||\mathbf{L}(v_2, \mathcal{F}_0) - \mathbf{L}(v_1, \mathcal{F}_0)||_q}{|v_2 - v_1|} < \infty.$$

We denote the collection of $\mathcal{L}^q$ *stochastic Lipschitz-continuous* over $[0,1]$ classes by Lip_q.

3.2 Nonstationary regressors

We need the following assumptions on the independent variables X_{it} for each i:

(C4) The covariates X_{it} allow for the representation $X_{it} = \mathbf{H}_i(t/T, \mathcal{U}_{it})$, where $\mathcal{U}_{it} = (\ldots, u_{it-1}, u_{it})$ with u_{it} being i.i.d. random variables and $\mathbf{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^T : [0,1] \times \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^d$ is a measurable function such that $\mathbf{H}_i(v, \mathcal{U}_{it})$ is well defined for each $v \in [0,1]$.

(C5) Let $N_i(v)$ be the $d \times d$ matrix with kl-th entry $n_{i,kl}(v) = \mathbb{E}[H_{ik}(v, \mathcal{U}_{i0}), H_{il}(v, \mathcal{U}_{i0})]$. We assume that the smallest eigenvalue of $N_i(v)$ is bounded away from 0 on $v \in [0,1]$.

(C6) Let $\mathbf{H}_i(v, \mathcal{U}_{it}) \in Lip_2$ and $\sup_{0 \le v \le 1} ||\mathbf{H}_i(v, \mathcal{U}_{it})||_4 < \infty$.

We define the first-differenced regressors as follows.

$$\Delta \mathbf{X}_{it} = \mathbf{H}_i(v_t, \mathcal{U}_{it}) - \mathbf{H}_i(v_{t-1}, \mathcal{U}_{it-1}) := \Delta \mathbf{H}_i(v_t, \mathcal{U}_{it})$$

where $v_t = t/T$. Similarly,

$$\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1} = G_i(\ldots, \eta_{it-1}, \eta_{it}) - G_i(\ldots, \eta_{it-2}, \eta_{it-1}) = \Delta G_i(\ldots, \eta_{it-1}, \eta_{it}).$$

With these assumptions we can prove the following proposition.

Proposition 3.1. *Under Assumptions (C4) - (C6),*

$$\Delta \mathbf{H}_i(v_t, \mathcal{U}_{it}) \in Lip_2 \text{ and } \sup_{0 \le v \le 1} ||\Delta \mathbf{H}_i(v, \mathcal{U}_{it})||_4 < \infty$$

Proof of Proposition 3.1. Note the following

$$\sup_{0 \le v_t < v_s \le 1} \frac{||\Delta \mathbf{H}_i(v_s, \mathcal{U}_{i0}) - \Delta \mathbf{H}_i(v_t, \mathcal{U}_{i0})||_2}{|v_s - v_t|} \le$$

$$\le \sup_{0 \le v_t < v_s \le 1} \frac{||\mathbf{H}_i(v_s, \mathcal{U}_{i0}) - \mathbf{H}_i(v_{s-1}, \mathcal{U}_{i0}) - |\mathbf{H}_i(v_t, \mathcal{U}_{i0}) - \mathbf{H}_i(v_{t-1}, \mathcal{U}_{i0})|||_2}{|v_s - v_t|} \le$$

$$\le \sup_{0 \le v_t < v_s \le 1} \frac{||\mathbf{H}_i(v_s, \mathcal{U}_{i0}) - \mathbf{H}_i(v_t, \mathcal{U}_{i0})||_2}{|v_s - v_t|} + \sup_{0 \le v_{t-1} < v_{s-1} \le 1} \frac{||\mathbf{H}_i(v_{s-1}, \mathcal{U}_{i0}) - \mathbf{H}_i(v_{t-1}, \mathcal{U}_{i0})||_2}{|v_{s-1} - v_{t-1}|},$$

where $v_s - v_t = v_{s-1} - v_{t-1}$ by definition. By Assumption (C6),

$$\sup_{0 \le v_t < v_s \le 1} \frac{||\mathbf{H}_i(v_s, \mathcal{U}_{i0}) - \mathbf{H}_i(v_t, \mathcal{U}_{i0})||_2}{|v_s - v_t|} < \infty$$

and

$$\sup_{0 \le v_{t-1} < v_{s-1} \le 1} \frac{||\mathbf{H}_i(v_{s-1}, \mathcal{U}_{i0}) - \mathbf{H}_i(v_{t-1}, \mathcal{U}_{i0})||_2}{|v_{s-1} - v_{t-1}|} < \infty.$$

Thus,

$$\sup_{0 \le v_t < v_s \le 1} \frac{||\Delta \mathbf{H}_i(v_s, \mathcal{U}_{i0}) - \Delta \mathbf{H}_i(v_t, \mathcal{U}_{i0})||_2}{|v_s - v_t|} < \infty$$

and $\Delta \mathbf{H}_i(v_t, \mathcal{U}_{it}) \in Lip_2$. Moreover, by Assumption (C6),

$$\sup_{0 \le v_t \le 1} ||\Delta \mathbf{H}_i(v_t, \mathcal{U}_{it})||_4 \le \sup_{0 \le v_t \le 1} ||\mathbf{H}_i(v, \mathcal{U}_{it})||_4 + \sup_{0 \le v_{t-1} \le 1} ||\mathbf{H}_i(v_{t-1}, \mathcal{U}_{it-1})||_4 < \infty,$$

which completes the proof. □

To be able to prove the next proposition, we need an additional assumption on the relationship between the covariates and the error process.

(C7) For each i $\{\eta_{it}\}_{t\in\mathbb{Z}}$ from Assumption (C1) and $\{u_{it}\}_{t\in\mathbb{Z}}$ from Assumption (C4) are independent of each other.

Proposition 3.2. *Under Assumptions (C1) - (C7),*

$$\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\Big| = O_P(1)$$

Proof of Proposition 3.2. Let $\zeta_{i,t} = (u_{i,t}, \eta_{i,t})^T$. We need the following notation:

$$\begin{aligned}
&\mathcal{I}_{i,t} && := (\ldots, \zeta_{i,t-2}, \zeta_{i,t-1}, \zeta_{i,t}),\\
&\Delta\mathbf{U}_i(v, \mathcal{I}_{i,t}) && := \Delta\mathbf{H}_i(v_t, \mathcal{U}_{it})\Delta G_i(\ldots, \eta_{it-1}, \eta_{it}),\\
&\mathcal{P}_{i,t}(\cdot) && := \mathbb{E}[\cdot|\mathcal{I}_{i,t}] - \mathbb{E}[\cdot|\mathcal{I}_{i,t-1}],\\
&\kappa_i && := \frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\varepsilon_{it},\\
&\kappa_{i,s}^{\mathcal{P}} && := \frac{1}{T}\sum_{t=1}^{T}\mathcal{P}_{i,t-s}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\big).
\end{aligned}$$

Then,

$$\begin{aligned}
||\kappa_{i,s}^{\mathcal{P}}||^2 &= \Big|\Big|\frac{1}{T}\sum_{t=1}^{T}\mathcal{P}_{i,t-s}\big(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}\big)\Big|\Big|^2 \leq\\
&\leq \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s-1})\Big|\Big|^2 =\\
&= \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta\mathbf{X}_{it,s}^{*}\Delta\varepsilon_{it,s}^{*}|\mathcal{I}_{i,t-s})\Big|\Big|^2,
\end{aligned}$$

where $\Delta\mathbf{X}_{it,s}^{*}\Delta\varepsilon_{it,s}^{*}$ denotes $\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}$ with $\zeta_{i,t-s}$ replaced by its i.i.d. copy $\zeta_{i,t-s}^{*}$. In this case $\mathbb{E}(\Delta\mathbf{X}_{it,s}^{*}\Delta\varepsilon_{it,s}^{*}|\mathcal{I}_{i,t-s-1}) = \mathbb{E}(\Delta\mathbf{X}_{it,s}^{*}\Delta\varepsilon_{it,s}^{*}|\mathcal{I}_{i,t-s})$. Furthermore, by linearity of the expectation and Jensen's inequality we have

$$\begin{aligned}
||\kappa_{i,s}^{\mathcal{P}}||^2 &\leq \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\mathbb{E}(\Delta\mathbf{X}_{it}\Delta\varepsilon_{it}|\mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta\mathbf{X}_{it,s}^{*}\Delta\varepsilon_{it,s}^{*}|\mathcal{I}_{i,t-s})\Big|\Big|^2 \leq\\
&\leq \frac{1}{T^2}\sum_{t=1}^{T}\Big|\Big|\Delta\mathbf{X}_{it}\Delta\varepsilon_{it} - \Delta\mathbf{X}_{it,s}^{*}\Delta\varepsilon_{it,s}^{*}\Big|\Big|^2
\end{aligned}$$

□

3.3 Analysis of UK weather station data

To illustrate our test method from Section 3, we examine a dataset of monthly mean temperatures from 34 different UK weather stations. The data are publicly available on the webpage of the UK Met Office. We use a subset of 25 stations for which data are available over the time span from 1986 to 2017. We thus observe a time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ of length $T = 386$ for each station $i \in \{1, \dots, 25\}$. The time series $\mathcal{Y}_i$ is assumed to follow the model

$$Y_{it} = \alpha_i(t) + m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}, \tag{7}$$

where m_i is an unknown nonparametric time trend and $\alpha_i(t)$ is a month-specific intercept which captures the seasonality pattern in the data. We suppose that $\alpha_i(t) = \alpha_i(t + 12\ell)$ for any integer ℓ, that is, we have a different intercept $\alpha_i(k)$ for each month $k = 1, \dots, 12$. The test method and the underlying theory from Section 3 can be easily adapted to model (7), which is a slight extension of model (**??**). The details are provided below. As in Section **??**, the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ is assumed to have the AR(1) structure $\varepsilon_{it} = a_i \varepsilon_{i,t-1} + \eta_{it}$ for each i, where η_{it} are i.i.d. innovations with mean zero.

In this supplement, we provide the proofs that are omitted in the paper. Specifically, we prove Proposition **??** and derive the technical results from Sections **??** and 3. We employ the same notation as summarized at the beginning of the Appendix in the paper.

Proof of Proposition ??

We only need to prove part (b). The arguments closely follow those for the proof of Proposition **??**. To start with, we introduce the notation $\widehat{\psi}'_T(u,h) = \widehat{\psi}'^A_T(u,h) + \widehat{\psi}'^B_T(u,h)$, where $\widehat{\psi}'^A_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h)\varepsilon_t$ and $\widehat{\psi}'^B_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h) m_T(\frac{t}{T})$. We further write $m_T(\frac{t}{T}) = m_T(u) + m'_T(\xi_{u,t,T})(\frac{t}{T} - u)$, where $\xi_{u,t,T}$ is an intermediate point between u and t/T. The local linear weights $w'_{t,T}(u,h)$ are constructed such that $\sum_{t=1}^T w'_{t,T}(u,h) = 0$, which implies that

$$\widehat{\psi}'^B_T(u,h) = \sum_{t=1}^T w'_{t,T}(u,h)\Big(\frac{\frac{t}{T} - u}{h}\Big) h m'_T(\xi_{u,t,T}). \tag{S.1}$$

By assumption, there exists $(u_0, h_0) \in \mathcal{G}_T$ with $[u_0 - h_0, u_0 + h_0] \subseteq [0,1]$ such that $m'_T(w) \ge c_T \sqrt{\log T/(T h_0^3)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$. (The case that $-m'_T(w) \ge c_T\sqrt{\log T/(T h_0^3)}$ for all w can be treated analogously.) Since the kernel K is symmetric and $u_0 = t/T$ for some t, it holds that $S_{T,1}(u_0, h_0) = 0$, which in turn implies that

$$\begin{aligned} & w'_{t,T}(u_0,h_0)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) \\ & = K\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)^2 \Big/ \Big\{\sum_{t=1}^T K^2\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big)^2\Big\}^{1/2} \ge 0. \end{aligned}$$

From this and the assumption that $m_T'(w) \ge c_T \sqrt{\log T/(Th_0^3)}$ for all $w \in [u_0 - h_0, u_0 + h_0]$, we get that

$$\widehat{\psi}_T'^B(u_0, h_0) \ge c_T \sqrt{\frac{\log T}{Th_0}} \sum_{t=1}^T w_{t,T}'(u_0, h_0) \Big(\frac{\frac{t}{T} - u_0}{h_0} \Big). \tag{S.2}$$

Analogous to (**??**), we can show that for any $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$,

$$\Big| \sum_{t=1}^T w_{t,T}'(u, h) \Big(\frac{\frac{t}{T} - u}{h} \Big) - \kappa' \sqrt{Th} \Big| \le \frac{C}{\sqrt{Th}}, \tag{S.3}$$

where $\kappa' = (\int K(\varphi)\varphi^2 d\varphi)/(\int K^2(\varphi)\varphi^2 d\varphi)^{1/2}$ and the constant C does not depend on u, h and T. (S.3) implies that $\sum_{t=1}^T w_{t,T}'(u, h)(\frac{t}{T} - u)/h \ge \kappa' \sqrt{Th}/2$ for sufficiently large T and any $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$. From this and (S.2), we can infer that

$$\widehat{\psi}_T'^B(u_0, h_0) \ge \frac{\kappa' c_T \sqrt{\log T}}{2} \tag{S.4}$$

for sufficiently large T. Furthermore, by arguments very similar to those for the proof of Proposition **??**, it follows that

$$\max_{(u,h) \in \mathcal{G}_T} |\widehat{\psi}_T'^A(u, h)| = O_p(\sqrt{\log T}). \tag{S.5}$$

With the help of (S.4), (S.5) and the fact that $\lambda(h) \le \lambda(h_{\min}) \le C\sqrt{\log T}$, we can finally conclude that

$$\begin{aligned} \widehat{\Psi}_T' &\ge \max_{(u,h) \in \mathcal{G}_T} \frac{|\widehat{\psi}_T'^B(u, h)|}{\widehat{\sigma}} - \max_{(u,h) \in \mathcal{G}_T} \Big\{ \frac{|\widehat{\psi}_T'^A(u, h)|}{\widehat{\sigma}} + \lambda(h) \Big\} \\ &= \max_{(u,h) \in \mathcal{G}_T} \frac{|\widehat{\psi}_T'^B(u, h)|}{\widehat{\sigma}} + O_p(\sqrt{\log T}) \\ &\ge \frac{\kappa' c_T \sqrt{\log T}}{2\widehat{\sigma}} + O_p(\sqrt{\log T}). \end{aligned} \tag{S.6}$$

Since $q_T'(\alpha) = O(\sqrt{\log T})$ for any fixed $\alpha \in (0, 1)$, (S.6) immediately yields that $\mathbb{P}(\widehat{\Psi}_T' \le q_T'(\alpha)) = o(1)$.

Proof of Proposition ??

In what follows, we show that

$$\mathbb{P}(E_T^+) \ge (1 - \alpha) + o(1). \tag{S.7}$$

The other statement of Proposition **??** can be verified by analogous arguments. (S.7) is a consequence of the following two observations:

(i) For all $(u,h) \in \mathcal{G}_T$ with

$$\Big|\frac{\widehat{\psi}'_T(u,h) - \mathbb{E}\widehat{\psi}'_T(u,h)}{\widehat{\sigma}}\Big| - \lambda(h) \le q'_T(\alpha) \quad \text{and} \quad \frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} - \lambda(h) > q'_T(\alpha),$$

it holds that $\mathbb{E}[\widehat{\psi}'_T(u,h)] > 0$.

(ii) For all $(u,h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0,1]$, $\mathbb{E}[\widehat{\psi}'_T(u,h)] > 0$ implies that $m'(v) > 0$ for some $v \in [u-h, u+h]$.

Observation (i) is trivial, (ii) can be seen as follows: Let (u,h) be any point with $(u,h) \in \mathcal{G}_T$ and $[u-h, u+h] \subseteq [0,1]$. It holds that $\mathbb{E}[\widehat{\psi}'_T(u,h)] = \widehat{\psi}'^B_T(u,h)$, where $\widehat{\psi}'^B_T(u,h)$ has been defined in the proof of Proposition **??**. There, we have already seen that

$$\widehat{\psi}'^B_T(u,h) = \sum_{t=1}^{T} w'_{t,T}(u,h)\Big(\frac{\frac{t}{T}-u}{h}\Big)\, hm'(\xi_{u,t,T}),$$

where $\xi_{u,t,T}$ is some intermediate point between u and t/T. Moreover, $S_{T,1}(u,h) = 0$, which implies that $w'_{t,T}(u,h)(\frac{t}{T}-u)/h \ge 0$ for any t. Hence, $\mathbb{E}[\widehat{\psi}'_T(u,h)] = \widehat{\psi}'^B_T(u,h)$ can only take a positive value if $m'(v) > 0$ for some $v \in [u-h, u+h]$.

We now proceed as in the proof of Proposition **??**. From observations (i) and (ii), we can infer the following: On the event

$$\big\{\widehat{\Phi}'_T \le q'_T(\alpha)\big\} = \Big\{\max_{(u,h)\in\mathcal{G}_T}\Big(\Big|\frac{\widehat{\psi}'_T(u,h) - \mathbb{E}\widehat{\psi}'_T(u,h)}{\widehat{\sigma}}\Big| - \lambda(h)\Big) \le q'_T(\alpha)\Big\},$$

it holds that for all $(u,h) \in \mathcal{A}^+_T$, $m'(v) > 0$ for some $v \in I_{u,h} = [u-h, u+h]$. We thus obtain that $\{\widehat{\Phi}'_T \le q'_T(\alpha)\} \subseteq E^+_T$. This in turn implies that

$$\mathbb{P}(E^+_T) \ge \mathbb{P}\big(\widehat{\Phi}'_T \le q'_T(\alpha)\big) = (1-\alpha) + o(1),$$

where the last equality holds by Theorem **??**.

Proof of Theorem ??

By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), there exist a standard Brownian motion $\mathbb{B}_i$ and a sequence $\{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ for each i such that the following holds: (i) $\mathbb{B}_i$ and $\{\widetilde{\varepsilon}_{it} : t \in \mathbb{N}\}$ are independent across i, (ii) $[\widetilde{\varepsilon}_{i1}, \dots, \widetilde{\varepsilon}_{iT}] \overset{\mathcal{D}}{=} [\varepsilon_{i1}, \dots, \varepsilon_{iT}]$ for each i and T, and (iii)

$$\max_{1\le t\le T}\Big|\sum_{s=1}^{t}\widetilde{\varepsilon}_{is} - \sigma_i\mathbb{B}_i(t)\Big| = o\big(T^{1/q}\big) \quad \text{a.s.}$$

for each i, where $\sigma_i^2 = \sum_{k\in\mathbb{Z}}\mathrm{Cov}(\varepsilon_{i0}, \varepsilon_{ik})$ denotes the long-run error variance. We define

$$\widetilde{\Phi}_{n,T} = \max_{1\le i<j\le N}\widetilde{\Phi}_{ij,T} \quad \text{with} \quad \widetilde{\Phi}_{ij,T} = \max_{(u,h)\in\mathcal{G}_T}\Big\{\Big|\frac{\widetilde{\phi}_{ij,T}(u,h)}{(\widetilde{\sigma}_i^2 + \widetilde{\sigma}_j^2)^{1/2}}\Big| - \lambda(h)\Big\},$$

where $\widetilde{\phi}_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{(\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i) - (\widetilde{\varepsilon}_{jt} - \bar{\widetilde{\varepsilon}}_j)\}$ with $\bar{\widetilde{\varepsilon}}_i = \bar{\widetilde{\varepsilon}}_{i,T} = T^{-1}\sum_{t=1}^T \widetilde{\varepsilon}_{it}$. Moreover, $\widetilde{\sigma}_i^2$ is the same estimator as $\widehat{\sigma}_i^2$ with $\widehat{Y}_{it} = (m_i(\frac{t}{T}) - \bar{m}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$ replaced by $\widetilde{Y}_{it} = (m_i(\frac{t}{T}) - \bar{m}_i) + (\widetilde{\varepsilon}_{it} - \bar{\widetilde{\varepsilon}}_i)$, where we set $\bar{m}_i = \bar{m}_{i,T} = T^{-1}\sum_{t=1}^T m_i(\frac{t}{T})$. By construction, $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\Phi}_{n,T}$ for each $T \geq 1$. In addition, we let

$$\Phi^*_{n,T} = \max_{1\leq i<j\leq N} \Phi^*_{ij,T} \quad \text{with} \quad \Phi^*_{ij,T} = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi^*_{ij,T}(u,h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \Big| - \lambda(h) \Big\},$$

where $\phi^*_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\{\sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j)\}$ and $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$. Without loss of generality, we also set $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$ in the Gaussian statistic $\Phi_{n,T}$ which is defined in Section **??**.

We now follow the proof strategy for Theorem **??**. Slightly modifying the arguments for Proposition **??**, we can show that

$$\big|\widetilde{\Phi}_{n,T} - \Phi^*_{n,T}\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T}\Big). \tag{S.8}$$

Moreover, it holds that

$$\big|\Phi_{n,T} - \Phi^*_{n,T}\big| = o_p\big(\rho_T\sqrt{\log T}\big), \tag{S.9}$$

which is a consequence of the following facts: (i) the variables Z_{it} are i.i.d. standard normal, (ii) $|\mathcal{G}_T| = O(T^\theta)$ for some large but fixed constant θ, (iii) $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$, and (iv) $\max_{(u,h)\in\mathcal{G}_T} |\sum_{t=1}^T w_{t,T}(u,h)| \leq CTh_{\max}$, where the constant C is independent of T. Finally, by arguments very similar to those for Proposition **??**, we obtain that

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\big|\Phi^*_{n,T} - x\big| \leq \delta_T\Big) = o(1) \tag{S.10}$$

with $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$. Combining (S.8)–(S.10) with Lemma **??**, we can infer that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\widetilde{\Phi}_{n,T} \leq x\big) - \mathbb{P}\big(\Phi^*_{n,T} \leq x\big)\big| = o(1) \tag{S.11}$$

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\Phi_{n,T} \leq x\big) - \mathbb{P}\big(\Phi^*_{n,T} \leq x\big)\big| = o(1). \tag{S.12}$$

From (S.11) and (S.12), it immediately follows that

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(\widetilde{\Phi}_{n,T} \leq x\big) - \mathbb{P}\big(\Phi_{n,T} \leq x\big)\big| = o(1),$$

which in turn implies that $\mathbb{P}(\widetilde{\Phi}_{n,T} \leq q_{n,T}(\alpha)) = (1-\alpha) + o(1)$. Since $\widetilde{\Phi}_{n,T}$ has the same distribution as $\widehat{\Phi}_{n,T}$, this completes the proof of Theorem **??**.

Proof of Proposition ??

Consider the event

$$B_{n,T} = \Big\{ \max_{1\leq \ell\leq N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} \leq q_{n,T}(\alpha) \;\text{ and }\; \min_{1\leq \ell<\ell'\leq N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T} > q_{n,T}(\alpha) \Big\}.$$

The term $\max_{1\le \ell \le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T}$ is the largest multiscale distance between two time series i and j from the same group, whereas $\min_{1\le \ell<\ell' \le N} \min_{i\in G_\ell, j\in G_{\ell'}} \widehat{\Psi}_{ij,T}$ is the smallest multiscale distance between two time series from two different groups. On the event $B_{n,T}$, it obviously holds that

$$\max_{1\le \ell \le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} < \min_{1\le \ell<\ell' \le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T}. \tag{S.13}$$

Hence, any two time series from the same class have a smaller distance than any two time series from two different classes. With the help of Theorem **??**, it is easy to see that

$$\mathbb{P}\Big(\max_{1\le \ell \le N} \max_{i,j\in G_\ell} \widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha)\Big) \ge (1-\alpha) + o(1).$$

Moreover, the same arguments as for part (b) of Proposition **??** show that

$$\mathbb{P}\Big(\min_{1\le \ell<\ell' \le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T} \le q_{n,T}(\alpha)\Big) = o(1).$$

Taken together, these two statements imply that

$$\mathbb{P}\big(B_{n,T}\big) \ge (1-\alpha) + o(1). \tag{S.14}$$

In what follows, we show that on the event $B_{n,T}$, (i) $\{\widehat{G}_1^{[n-N]}, \dots, \widehat{G}_N^{[n-N]}\} = \{G_1, \dots$ $\dots, G_N\}$ and (ii) $\widehat{N} = N$. From (i), (ii) and (S.14), the statements of Proposition **??** easily follow.

Proof of (i). Suppose we are on the event $B_{n,T}$. The proof proceeds by induction on the iteration steps r of the HAC algorithm.

Base case ($r = 0$): In the first iteration step, the HAC algorithm merges two singleton clusters $\widehat{G}_i^{[0]} = \{i\}$ and $\widehat{G}_j^{[0]} = \{j\}$ with i and j belonging to the same group G_k. This is a direct consequence of (S.13). The algorithm thus produces a partition $\{\widehat{G}_1^{[1]}, \dots, \widehat{G}_{n-1}^{[1]}\}$ whose elements $\widehat{G}_\ell^{[1]}$ all have the following property: $\widehat{G}_\ell^{[1]} \subseteq G_k$ for some k, that is, each cluster $\widehat{G}_\ell^{[1]}$ contains elements from only one group.

Induction step ($r \curvearrowright r+1$): Now suppose we are in the r-th iteration step for some $r < n - N$. Assume that the partition $\{\widehat{G}_1^{[r]}, \dots, \widehat{G}_{n-r}^{[r]}\}$ is such that for any ℓ, $\widehat{G}_\ell^{[r]} \subseteq G_k$ for some k. Because of (S.13), the dissimilarity $\widehat{\Delta}(\widehat{G}_\ell^{[r]}, \widehat{G}_{\ell'}^{[r]})$ gets minimal for two clusters $\widehat{G}_\ell^{[r]}$ and $\widehat{G}_{\ell'}^{[r]}$ with the property that $\widehat{G}_\ell^{[r]} \cup \widehat{G}_{\ell'}^{[r]} \subseteq G_k$ for some k. Hence, the HAC algorithm produces a partition $\{\widehat{G}_1^{[r+1]}, \dots, \widehat{G}_{n-(r+1)}^{[r+1]}\}$ whose elements $\widehat{G}_\ell^{[r+1]}$ are all such that $\widehat{G}_\ell^{[r+1]} \subseteq G_k$ for some k.

The above induction argument shows the following: For any $r \le n-N$, the partition $\{\widehat{G}_1^{[r]}, \dots, \widehat{G}_{n-r}^{[r]}\}$ consists of clusters $\widehat{G}_\ell^{[r]}$ which all have the property that $\widehat{G}_\ell^{[r]} \subseteq G_k$ for some k. This in particular holds for the partition $\{\widehat{G}_1^{[n-N]}, \dots, \widehat{G}_N^{[n-N]}\}$, which implies that $\{\widehat{G}_1^{[n-N]}, \dots, \widehat{G}_N^{[n-N]}\} = \{G_1, \dots, G_N\}$. $\square$

Proof of (ii). To start with, consider any partition $\{\widehat{G}_1^{[n-r]},\dots,\widehat{G}_r^{[n-r]}\}$ with $r < N$ elements. Such a partition must contain at least one element $\widehat{G}_\ell^{[n-r]}$ with the following property: $\widehat{G}_\ell^{[n-r]} \cap G_k \neq \emptyset$ and $\widehat{G}_\ell^{[n-r]} \cap G_{k'} \neq \emptyset$ for some $k \neq k'$. On the event $B_{n,T}$, it obviously holds that $\widehat{\Delta}(S) > q_{n,T}(\alpha)$ for any S with the property that $S \cap G_k \neq \emptyset$ and $S \cap G_{k'} \neq \emptyset$ for some $k \neq k'$. Hence, we can infer that on the event $B_{n,T}$, $\max_{1\le \ell \le r} \widehat{\Delta}(\widehat{G}_\ell^{[n-r]}) > q_{n,T}(\alpha)$ for any $r < N$.

Next consider the partition $\{\widehat{G}_1^{[n-r]},\dots,\widehat{G}_r^{[n-r]}\}$ with $r = N$ and suppose we are on the event $B_{n,T}$. From (i), we already know that $\{\widehat{G}_1^{[n-N]},\dots,\widehat{G}_N^{[n-N]}\} = \{G_1,\dots,G_N\}$. Moreover, it is easy to see that $\widehat{\Delta}(G_\ell) \le q_{n,T}(\alpha)$ for any ℓ. Hence, we obtain that $\max_{1\le\ell\le N} \widehat{\Delta}(\widehat{G}_\ell^{[n-N]}) = \max_{1\le\ell\le N}\widehat{\Delta}(G_\ell) \le q_{n,T}(\alpha)$.

Putting everything together, we can conclude that on the event $B_{n,T}$,

$$\min\Big\{ r = 1,2,\dots \,\Big|\, \max_{1\le\ell\le r} \widehat{\Delta}\big(\widehat{G}_\ell^{[n-r]}\big) \le q_{n,T}(\alpha)\Big\} = N,$$

that is, $\widehat{N} = N$. □

Proof of Proposition ??

We consider the event

$$D_{n,T} = \Big\{ \widehat{\Phi}_{n,T} \le q_{n,T}(\alpha) \text{ and } \min_{1\le \ell<\ell'\le N} \min_{\substack{i\in G_\ell,\\ j\in G_{\ell'}}} \widehat{\Psi}_{ij,T} > q_{n,T}(\alpha)\Big\},$$

where we write the statistic $\widehat{\Phi}_{n,T}$ as

$$\widehat{\Phi}_{n,T} = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}_{ij,T}(u,h) - \mathbb{E}\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h)\Big\}.$$

The event $D_{n,T}$ can be analysed by the same arguments as those applied to the event $B_{n,T}$ in the proof of Proposition **??**. In particular, analogous to (S.14) and statements (i) and (ii) in this proof, we can show that

$$\mathbb{P}\big(D_{n,T}\big) \ge (1-\alpha) + o(1) \tag{S.15}$$

and

$$D_{n,T} \subseteq \big\{ \widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell \big\}. \tag{S.16}$$

Moreover, we have that

$$D_{n,T} \subseteq \bigcap_{1\le\ell<\ell'\le\widehat{N}} E_{n,T}(\ell,\ell'), \tag{S.17}$$

which is a consequence of the following observation: For all i, j and $(u,h) \in \mathcal{G}_T$ with

$$\Big| \frac{\widehat{\psi}_{ij,T}(u,h) - \mathbb{E}\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h) \le q_{n,T}(\alpha) \quad \text{and} \quad \Big| \frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \Big| - \lambda(h) > q_{n,T}(\alpha),$$

it holds that $\mathbb{E}[\widehat{\psi}_{ij,T}(u,h)] \neq 0$, which in turn implies that $m_i(v) - m_j(v) \neq 0$ for some $v \in I_{u,h}$. From (S.16) and (S.17), we obtain that

$$D_{n,T} \subseteq \Big\{ \bigcap_{1 \le \ell < \ell' \le \widehat{N}} E_{n,T}(\ell,\ell') \Big\} \cap \big\{ \widehat{N} = N \text{ and } \widehat{G}_\ell = G_\ell \text{ for all } \ell \big\} = E_{n,T}.$$

This together with (S.15) implies that $\mathbb{P}(E_{n,T}) \geq (1-\alpha) + o(1)$, thus completing the proof.

Proof of Proposition ??

The proof makes use of the following three lemmas, which correspond to Lemmas 5–7 in Chernozhukov et al. (2015).

Lemma 3.3. *Let $(W_1,\ldots,W_p)^\top$ be a (not necessarily centred) Gaussian random vector in $\mathbb{R}^p$ with $\mathrm{Var}(W_j) = 1$ for all $1 \le j \le p$. Suppose that $\mathrm{Corr}(W_j,W_k) < 1$ whenever $j \neq k$. Then the distribution of $\max_{1 \le j \le p} W_j$ is absolutely continuous with respect to Lebesgue measure and a version of the density is given by*

$$f(x) = f_0(x) \sum_{j=1}^{p} e^{\mathbb{E}[W_j]x - \mathbb{E}[W_j]^2/2}\, \mathbb{P}\big(W_k \le x \textit{ for all } k \neq j \,\big|\, W_j = x\big).$$

Lemma 3.4. *Let $(W_0,W_1,\ldots,W_p)^\top$ be a (not necessarily centred) Gaussian random vector with $\mathrm{Var}(W_j) = 1$ for all $0 \le j \le p$. Suppose that $\mathbb{E}[W_0] \geq 0$. Then the map*

$$x \mapsto e^{\mathbb{E}[W_0]x - \mathbb{E}[W_0]^2/2}\, \mathbb{P}\big(W_j \le x \textit{ for } 1 \le j \le p \,\big|\, W_0 = x\big)$$

is non-decreasing on $\mathbb{R}$.

Lemma 3.5. *Let $(X_1,\ldots,X_p)^\top$ be a centred Gaussian random vector in $\mathbb{R}^p$ with $\max_{1 \le j \le p} \mathbb{E}[X_j^2] \le \sigma^2$ for some $\sigma^2 > 0$. Then for any $r > 0$,*

$$\mathbb{P}\Big(\max_{1 \le j \le p} X_j \geq \mathbb{E}\Big[\max_{1 \le j \le p} X_j \Big] + r \Big) \le e^{-r^2/(2\sigma^2)}.$$

The proof of Lemmas 3.3 and 3.4 can be found in Chernozhukov et al. (2015). Lemma 3.5 is a standard result on Gaussian concentration whose proof is given e.g. in Ledoux (2001); see Theorem 7.1 therein. We now closely follow the arguments for the proof of Theorem 3 in Chernozhukov et al. (2015). The proof splits up into three steps.

Step 1. Pick any $x \geq 0$ and set

$$W_j = \frac{X_j - x}{\sigma_j} + \frac{\overline{\mu} + x}{\underline{\sigma}}.$$

By construction, $\mathbb{E}[W_j] \geq 0$ and $\mathrm{Var}(W_j) = 1$. Defining $Z = \max_{1\leq j\leq p} W_j$, it holds that

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\leq j\leq p} X_j - x\Big| \leq \delta\Big) &\leq \mathbb{P}\Big(\Big|\max_{1\leq j\leq p} \frac{X_j - x}{\sigma_j}\Big| \leq \frac{\delta}{\underline{\sigma}}\Big) \\ &\leq \sup_{y\in\mathbb{R}} \mathbb{P}\Big(\Big|\max_{1\leq j\leq p} \frac{X_j - x}{\sigma_j} + \frac{\overline{\mu} + x}{\underline{\sigma}} - y\Big| \leq \frac{\delta}{\underline{\sigma}}\Big) \\ &= \sup_{y\in\mathbb{R}} \mathbb{P}\Big(|Z - y| \leq \frac{\delta}{\underline{\sigma}}\Big).\end{aligned}$$

Step 2. We now bound the density of Z. Without loss of generality, we assume that $\mathrm{Corr}(W_j, W_k) < 1$ for $k \neq j$. The marginal distribution of W_j is $N(\nu_j, 1)$ with $\nu_j = \mathbb{E}[W_j] = (\mu_j/\sigma_j + \overline{\mu}/\underline{\sigma}) + (x/\underline{\sigma} - x/\sigma_j) \geq 0$. Hence, by Lemmas 3.3 and 3.4, the random variable Z has a density of the form

$$f_p(z) = f_0(z) G_p(z), \tag{S.18}$$

where the map $z \mapsto G_p(z)$ is non-decreasing. Define $\overline{Z} = \max_{1\leq j\leq p}(W_j - \mathbb{E}[W_j])$ and set $\overline{z} = 2\overline{\mu}/\underline{\sigma} + x(1/\underline{\sigma} - 1/\overline{\sigma})$ such that $\mathbb{E}[W_j] \leq \overline{z}$ for any $1 \leq j \leq p$. With these definitions at hand, we obtain that

$$\begin{aligned}\int_z^\infty f_0(u)du\, G_p(z) &\leq \int_z^\infty f_0(u)G_p(u)du = \mathbb{P}(Z > z) \\ &\leq P(\overline{Z} > z - \overline{z}) \leq \exp\Big(-\frac{(z - \overline{z} - \mathbb{E}[\overline{Z}])_+^2}{2}\Big),\end{aligned}$$

where the last inequality follows from Lemma 3.5. Since $W_j - \mathbb{E}[W_j] = (X_j - \mu_j)/\sigma_j$, it holds that

$$\mathbb{E}[\overline{Z}] = \mathbb{E}\Big[\max_{1\leq j\leq p}\Big\{\frac{X_j - \mu_j}{\sigma_j}\Big\}\Big] =: a_p.$$

Hence, for every $z \in \mathbb{R}$,

$$G_p(z) \leq \frac{1}{1 - F_0(z)} \exp\Big(-\frac{(z - \overline{z} - a_p)_+^2}{2}\Big). \tag{S.19}$$

Mill's inequality states that for $z > 0$,

$$z \leq \frac{f_0(z)}{1 - F_0(z)} \leq z\frac{1 + z^2}{z^2}.$$

Since $(1 + z^2)/z^2 \leq 2$ for $z \geq 1$ and $f_0(z)/\{1 - F_0(z)\} \leq 1.53 \leq 2$ for $z \in (-\infty, 1)$, we can infer that

$$\frac{f_0(z)}{1 - F_0(z)} \leq 2(z \vee 1) \quad \text{for any } z \in \mathbb{R}.$$

This together with (S.18) and (S.19) yields that

$$f_p(z) \leq 2(z \vee 1) \exp\Big(-\frac{(z - \overline{z} - a_p)_+^2}{2}\Big) \quad \text{for any } z \in \mathbb{R}.$$

Step 3. By Step 2, we get that for any $y \in \mathbb{R}$ and $u > 0$,

$$\mathbb{P}(|Z - y| \le u) = \int_{y-u}^{y+u} f_p(z)dz \le 2u \max_{z \in [y-u, y+u]} f_p(z) \le 4u(\overline{z} + a_p + 1),$$

where the last inequality follows from the fact that the map $z \mapsto ze^{-(z-a)^2/2}$ (with $a > 0$) is non-increasing on $[a+1, \infty)$. Combining this bound with Step 1, we further obtain that for any $x \ge 0$ and $\delta > 0$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le 4\delta\Big\{\frac{2\overline{\mu}}{\underline{\sigma}} + |x|\Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big) + a_p + 1\Big\}/\underline{\sigma}. \tag{S.20}$$

This inequality also holds for $x < 0$ by an analogous argument, and hence for all $x \in \mathbb{R}$.

Now let $0 < \delta \le \underline{\sigma}$ and define $b_p = \mathbb{E}\max_{1\le j\le p}\{X_j - \mu_j\}$. For any $|x| \le \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$, (S.20) yields that

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) &\le \frac{4\delta}{\underline{\sigma}}\Big\{\overline{\mu}\Big(\frac{3}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big) + a_p + \Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big)b_p \\ &\qquad + \Big(\frac{\overline{\sigma}}{\underline{\sigma}} - 1\Big)\sqrt{2\log\Big(\frac{\underline{\sigma}}{\delta}\Big)} + 2 - \frac{\underline{\sigma}}{\overline{\sigma}}\Big\} \\ &\le C\delta\Big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\Big\}\end{aligned} \tag{S.21}$$

with a sufficiently large constant $C > 0$ that depends only on $\underline{\sigma}$ and $\overline{\sigma}$. For $|x| \ge \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$, we obtain that

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \frac{\delta}{\underline{\sigma}}, \tag{S.22}$$

which can be seen as follows: If $x > \delta + \overline{\mu}$, then $|\max_j X_j - x| \le \delta$ implies that $|x| - \delta \le \max_j X_j \le \max_j\{X_j - \mu_j\} + \overline{\mu}$ and thus $\max_j\{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}$. Hence, it holds that

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}\Big). \tag{S.23}$$

If $x < -(\delta + \overline{\mu})$, then $|\max_j X_j - x| \le \delta$ implies that $\max_j\{X_j - \mu_j\} \le -|x| + \delta + \overline{\mu}$. Hence, in this case,

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) &\le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j - \mu_j\} \le -|x| + \delta + \overline{\mu}\Big) \\ &\le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}\Big),\end{aligned} \tag{S.24}$$

where the last inequality follows from the fact that for centred Gaussian random variables V_j and $v > 0$, $\mathbb{P}(\max_j V_j \le -v) \le \mathbb{P}(V_1 \le -v) = P(V_1 \ge v) \le \mathbb{P}(\max_j V_j \ge v)$. With (S.23) and (S.24), we obtain that for any $|x| \ge \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$,

$$\begin{aligned}&\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}\Big) \\ &\quad \le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j - \mu_j\} \ge \mathbb{E}\Big[\max_{1\le j\le p}\{X_j - \mu_j\}\Big] + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}\Big) \le \frac{\delta}{\underline{\sigma}},\end{aligned}$$

the last inequality following from Lemma 3.5. To sum up, we have established that for any $0 < \delta \leq \underline{\sigma}$ and any $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\Big|\max_{1\leq j\leq p} X_j - x\Big| \leq \delta\Big) \leq C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\} \tag{S.25}$$

with some constant $C > 0$ that does only depend on $\underline{\sigma}$ and $\overline{\sigma}$. For $\delta > \underline{\sigma}$, (S.25) trivially follows upon setting $C \geq 1/\underline{\sigma}$. This completes the proof.

References

Abraham, C., Cornillon, P. A., Matzner-Løber, E. and Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics*, **30** 581–595.

Anderson, N. H., Hall, P. and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis*, **50** 41–54.

Anderson, T. W. (1962). On the distribution of the two-sample Cramér-von Mises criterion. *Annals of Mathematical Statistics*, **33** 1148–1159.

Atak, A., Linton, O. and Xiao, Z. (2011). A semiparametric panel model for unbalanced data with application to climate change in the United Kingdom. *Journal of Econometrics*, **164** 92–115.

Berkes, I., Liu, W. and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Annals of Probability*, **42** 794–817.

Boneva, L., Linton, O. and Vogt, M. (2015). A semiparametric model for heterogeneous panel data with fixed effects. *Journal of Econometrics*, **188** 327–345.

Boneva, L., Linton, O. and Vogt, M. (2016). The effect of fragmentation in trading on market quality in the UK equity market. *Journal of Applied Econometrics*, **31** 192–213.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, **83** 1147–1184.

Cai, Z. (2007). Trending time-varying coefficients time series models with serially correlated errors. *Journal of Econometrics*, **136** 163–188.

Chaudhuri, P. and Marron, J. S. (1999). SiZer for the exploration of structures in curves. *Journal of the American Statistical Association*, **94** 807–823.

Chaudhuri, P. and Marron, J. S. (2000). Scale space view of curve estimation. *Annals of Statistics*, **28** 408–428.

Chen, J., Gao, J. and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, **171** 71–85.

Chen, L. and Wu, W. B. (2018). Testing for trends in high-dimensional time series. *Forthcoming in Journal of the American Statistical Association.*

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields*, **162** 47–70.

Chiou, J.-M. and Li, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B*, **69** 679–699.

Degras, D., Xu, Z., Zhang, T. and Wu, W. B. (2012). Testing for parallelism among trends in multiple time series. *IEEE Transactions on Signal Processing*, **60** 1087–1097.

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statistics & Probability Letters*, **17** 199–204.

Dümbgen, L. (2002). Application of local rank tests to nonparametric regression. *Journal of Nonparametric Statistics*, **14** 511–537.

Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Annals of Statistics*, **29** 124–152.

Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Annals of Statistics*, **36** 1758–1785.

Eckle, K., Bissantz, N. and Dette, H. (2017). Multiscale inference for multivariate deconvolution. *Electronic Journal of Statistics*, **11** 4179–4219.

Finner, H. and Gontscharuk, V. (2018). Two-sample Kolmogorov-Smirnov-type tests revisited: old and new tests in terms of local levels. *Annals of Statistics*, **46** 3014–3037.

Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, **103** 1584–1594.

Grier, K. B. and Tullock, G. (1989). An empirical analysis of cross-national economic growth, 1951–1980. *Journal of Monetary Economics*, **24** 259–276.

Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, **85** 1039–1049.

Hannig, J. and Marron, J. S. (2006). Advanced distribution theory for SiZer. *Journal of the American Statistical Association*, **101** 484–499.

Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Annals of Statistics*, **18** 63–89.

Hidalgo, J. and Lee, J. (2014). A CUSUM test for common trends in large heterogeneous panels. In *Essays in Honor of Peter C. B. Phillips*. Emerald Group Publishing Limited, 303–345.

Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69** 599–631.

Jacques, J. and Preda, C. (2014). Functional data clustering: a survey. *Advances in Data Analysis and Classification*, **8** 231–255.

James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, **98** 397–408.

Karoly, D. J. and Wu, Q. (2005). Detection of regional surface temperature trends. *Journal of Climate*, **18** 4337–4343.

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.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramér-v. Mises tests. *Annals of Mathematical Statistics*, **30** 420–447.

Kneip, A., Sickles, R. C. and Song, W. (2012). A new panel data treatment for heterogeneity in time trends. *Econometric Theory*, **28** 590–628.

Ledoux, M. (2001). *Concentration of Measure Phenomenon.* American Mathematical Society.

Li, D., Chen, J. and Gao, J. (2010). Nonparametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, **14** 387–408.

Li, Q., Maasoumi, E. and Racine, J. S. (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data. *Journal of Econometrics*, **148** 186–200.

Lyubchich, V. and Gel, Y. R. (2016). A local factor nonparametric test for trend synchronism in multiple time series. *Journal of Multivariate Analysis*, **150** 91–104.

Mammen, E. (1992). *When does bootstrap work? Asymptotic results and simulations.* New York, Springer.

Nyblom, J. and Harvey, A. (2000). Tests of common stochastic trends. *Econometric Theory*, **16** 176–199.

Park, C., Vaughan, A., Hannig, J. and Kang, K.-H. (2009). SiZer analysis for the comparison of time series. *Journal of Statistical Planning and Inference*, **139** 3974–3988.

Proksch, K., Werner, F. and Munk, A. (2018). Multiscale scanning in inverse problems. *Forthcoming in Annals of Statistics.*

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics*, **169** 4–14.

Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Annals of Statistics*, **36** 1346–1374.

Rondonotti, V., Marron, J. S. and Park, C. (2007). SiZer for time series: a new approach to the analysis of trends. *Electronic Journal of Statistics*, **1** 268–289.

Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *Journal of Computational and Graphical Statistics*, **19** 175–190.

Schmidt-Hieber, J., Munk, A. and Dümbgen, L. (2013). Multiscale methods for shape constraints in deconvolution: confidence statements for qualitative features. *Annals of Statistics*, **41** 1299–1328.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, **83** 1097–1107.

Su, L. and Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, **206** 554–573.

Su, L., Shi, Z. and Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, **84** 2215–2264.

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, **164** 345–366.

Tarpey, T. (2007). Linear transformations and the k-means clustering algorithm. *The American Statistician*, **61** 34–40.

Tarpey, T. and Kinateder, K. K. (2003). Clustering functional data. *Journal of Classification*, **20** 093–114.

Vogelsang, T. J. and Franses, P. H. (2005). Testing for common deterministic trend slopes. *Journal of Econometrics*, **126** 1–24.

Vogt, M. and Linton, O. (2017). Classification of non-parametric regression functions in longitudinal data models. *Journal of the Royal Statistical Society: Series B*, **79** 5–27.

Wang, W., Phillips, P. C. and Su, L. (2018). Homogeneity pursuit in panel data models: theory and application. *Journal of Applied Econometrics*, **33** 797–815.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natn. Acad. Sci. USA*, **102** 14150–14154.

Xu, K.-L. (2012). Robustifying multivariate trend tests to nonstationary volatility. *Journal of Econometrics*, **169** 147–154.

Zhang, T. (2013). Clustering high-dimensional time series based on parallelism. *Journal of the American Statistical Association*, **108** 577–588.

Zhang, Y., Su, L. and Phillips, P. C. (2012). Testing for common trends in semi-parametric panel data models with fixed effects. *The Econometrics Journal*, **15** 56–100.