

# Multiscale Testing for Equality of Nonparametric Trend Curves

Marina Khismatullina<sup>1</sup>  
University of Bonn

Michael Vogt<sup>2</sup>  
University of Bonn

We develop multiscale methods to test qualitative hypotheses about nonparametric time trends. In many applications, practitioners are interested in whether the observed time series has a time trend at all, that is, whether the trend function is non-constant. Moreover, they would like to get further information about the shape of the trend function. Among other things, they would like to know in which time regions there is an upward/downward movement in the trend. When multiple time series are observed, another important question is whether the observed time series all have the same time trend. We design multiscale tests to formally approach these questions. We derive asymptotic theory for the proposed tests and investigate their finite sample performance by means of simulations. In addition, we illustrate the methods by two applications to temperature data.

**Key words:** Multiscale statistics; nonparametric regression; time series errors; shape constraints; strong approximations; anti-concentration bounds.

**AMS 2010 subject classifications:** 62E20; 62G10; 62G20; 62M10.

## 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 to compare trends in real gross domestic product (GDP) across different countries (cp. Grier and Tullock, 1989). Another example concerns the dynamics of

---

<sup>1</sup>Address: Bonn Graduate School of Economics, University of Bonn, 53113 Bonn, Germany. Email: [marina.k@uni-bonn.de](mailto:marina.k@uni-bonn.de).

<sup>2</sup>Corresponding author. Address: Department of Economics and Hausdorff Center for Mathematics, University of Bonn, 53113 Bonn, Germany. Email: [michael.vogt@uni-bonn.de](mailto:michael.vogt@uni-bonn.de).

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  $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \leq t \leq T\}$  for  $1 \leq i \leq n$ , where  $Y_{it}$  are real-valued random variables and  $\mathbf{X}_{it} = (X_{it,1}, \dots, X_{it,d})^\top$  are  $d$ -dimensional random vectors. Each time series  $\mathcal{Z}_i$  is modelled by the equation

$$Y_{it} = m_i\left(\frac{t}{T}\right) + \beta_i^\top \mathbf{X}_{it} + \alpha_i + \varepsilon_{it} \quad (1.1)$$

for  $1 \leq t \leq T$ , where  $m_i : [0, 1] \rightarrow \mathbb{R}$  is a nonparametric (deterministic) trend function,  $\mathbf{X}_{it}$  is a vector of regressors or controls and  $\beta_i$  is the corresponding parameter vector. Moreover,  $\alpha_i$  are so-called fixed effect error terms and  $\varepsilon_{it}$  are standard regression errors with  $\mathbb{E}[\varepsilon_{it} | \mathbf{X}_{it}] = 0$  for all  $t$ . Model (1.1) nests a number of panel models which have recently been considered in the literature. Special cases of model (1.1) 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 (1.1) with a parametric trend are studied in Vogelsang and Franses (2005), Sun (2011) and Xu (2012) among others. Within the general framework of model (1.1), we can formulate a number of interesting statistical questions concerning the set of trend functions  $\{m_i : 1 \leq i \leq n\}$ .

### (a) Testing for equality of nonparametric trend curves

In many application contexts, an important question is whether the time trends  $m_i$  in model (1.1) 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

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

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

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

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

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

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 (1.1). 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

$$Y_{it} = m_i\left(\frac{t}{T}\right) + \alpha_i + \varepsilon_{it} \quad (1 \leq t \leq T, 1 \leq i \leq n), \quad (1.2)$$

where  $\mathbb{E}[\varepsilon_{it}] = 0$  for all  $i$  and  $t$  and the terms  $\alpha_i$  are assumed to be deterministic. Obviously, (1.2) is a special case of (1.1) 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

$$\Delta_{n,T} = \sum_{i=1}^n \int_0^1 (\hat{m}_{i,h}(u) - \hat{m}_h(u))^2 du, \quad (1.3)$$

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 (1.1), where  $\beta_i = \beta$  for all  $i$ . The regression coefficients  $\beta_i$  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  $\beta$  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  $\Delta_i$  is a deterministic function with the property that  $\Delta_i \equiv 0$  under  $H_0$  and  $\eta_{it}$  denotes the error term. Testing  $H_0$  is thus equivalent to testing the hypothesis  $H'_0 : \Delta_i \equiv 0$  for all  $1 \leq i \leq n$ . The authors construct a test statistic for the hypothesis  $H'_0$  on the basis of nonparametric kernel estimators of the functions  $\Delta_i$  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  $\alpha$ ). 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 appropri-

ate 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  $\mathbb{P}(S_h > q_h(\alpha) \text{ for some } h \in \mathcal{H}) = (1 - \alpha) + o(1)$ .

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 (2018). 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 (1.1) 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 (1.1). 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  $H_0 : m_1 = \dots = m_n$  is violated in the general panel data model (1.1). 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 structure can be defined as follows within the framework of model (1.1): There exist sets or groups of time series  $G_1, \dots, G_{K_0}$  with  $\{1, \dots, n\} = \bigcup_{k=1}^{K_0} G_k$  such that for each  $1 \leq k \leq K_0$ ,

$$m_i = m_j \quad \text{for all } i, j \in G_k. \quad (1.4)$$

According to (1.4), 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  $G_1, \dots, G_{K_0}$  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 (1.1). Degras et al. (2012) used a repeated testing procedure based on  $L_2$ -type test statistics of the form (1.3) in order to estimate the unknown group structure in model (1.2). 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 (1.1) 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 Kinader (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 (1.1) 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  $1 \leq t \leq T$  and  $1 \leq i \leq n$ , where the coefficient vectors  $\beta_i$  are allowed to vary across individuals  $i$  and the error terms  $u_{it}$  may include fixed effects. Similar to the trend functions in model (1.1), the coefficients  $\beta_i$  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  $1 \leq k \leq K_0$ . 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 (1.1) 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 Vogt and Linton (2018). 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 (1.4) 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

Before we proceed any further, we need to introduce some notation used throughout the paper. For a vector  $\mathbf{v} = (v_1, \dots, v_m) \in \mathbb{R}^m$ , we write  $|\mathbf{v}| = (\sum_{i=1}^m v_i^2)^{1/2}$ . For a random vector  $\mathbf{V}$ , we define its  $\mathcal{L}^q, q > 1$  norm as  $\|\mathbf{V}\|_q = (\mathbb{E}|\mathbf{V}|^q)^{1/q}$ . For a particular case  $q = 2$ , we write  $\|\mathbf{V}\| := \|\mathbf{V}\|_2$ .

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

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

where  $\mathcal{F}_t = (\dots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \dots, \epsilon_{t-1}, \epsilon_t)$  and  $\mathcal{F}'_t = (\dots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \dots, \epsilon_{t-1}, \epsilon_t)$  is a coupled process of  $\mathcal{F}_t$  with  $\epsilon'_0$  being an i.i.d. copy of  $\epsilon_0$ .

The model setting is as follows. We observe a panel of  $n$  time series  $\mathcal{Z}_i = \{(Y_{it}, \mathbf{X}_{it}) : 1 \leq t \leq T\}$  of length  $T$  for  $1 \leq i \leq n$ . Each time series  $\mathcal{Z}_i$  satisfies the model equation

$$Y_{it} = \boldsymbol{\beta}_i^\top \mathbf{X}_{it} + m_i\left(\frac{t}{T}\right) + \alpha_i + \varepsilon_{it} \quad (2.1)$$

for  $1 \leq t \leq T$ , where  $\boldsymbol{\beta}_i$  is a  $d \times 1$  vector of unknown parameters,  $\mathbf{X}_{it}$  is a  $d \times 1$  vector of individual covariates,  $m_i$  is an unknown nonparametric trend function defined on  $[0, 1]$ ,  $\alpha_i$  is a (deterministic or random) intercept term and  $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$  is a zero-mean stationary error process. As usual in nonparametric regression, the trend functions  $m_i$  in model (2.1) depend on rescaled time  $t/T$  rather than on real time  $t$ ; cp. Robinson (1989), Dahlhaus (1997) and Vogt and Linton (2014) for the

use and some discussion of the rescaled time argument. The functions  $m_i$  are only identified up to an additive constant in model (2.1): One can reformulate the model as  $Y_{it} = [m_i(t/T) + c_i] + \beta_i^\top \mathbf{X}_{it} + [\alpha_i - c_i] + \varepsilon_{it}$ , that is, one can freely shift additive constants  $c_i$  between the trend  $m_i(t/T)$  and the error component  $\alpha_i$ . In order to obtain identification, one may impose different normalization constraints on the trends  $m_i$ . One possibility is to normalize them such that  $\int_0^1 m_i(u) du = 0$  for all  $i$ . In what follows, we take for granted that the trends  $m_i$  satisfy this constraint. The term  $\alpha_i$  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  $\mathcal{Z}_i$  which remain constant over time. We allow the error terms  $\alpha_i$  to be dependent across  $i$  in an arbitrary way. Hence, by including them in model equation (2.1), we allow the  $n$  time series  $\mathcal{Z}_i$  in our panel to be correlated with each other. Whereas the terms  $\alpha_i$  may be correlated, the error processes  $\mathcal{E}_i$  are assumed to be independent across  $i$ . Technical conditions regarding the model are discussed further in this section.

Finally, note that throughout the paper, we restrict attention to the case where the number of time series  $n$  in model (2.1) 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.

## 2.1 Assumptions

Each process  $\mathcal{E}_i$  is supposed to satisfy the following conditions:

- (C1) For each  $i$  the variables  $\varepsilon_{it}$  allow for the representation  $\varepsilon_{it} = G_i(\dots, \eta_{it-1}, \eta_{it})$ , where  $\eta_{it}$  are i.i.d. random variables across  $t$  and  $G_i : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}$  is a measurable function. Denote  $\mathcal{J}_{it} = (\dots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$ .
- (C2) For all  $i$  it holds that  $\mathbb{E}[\varepsilon_{it}] = 0$  and  $\|\varepsilon_{it}\|_q < \infty$  for some  $q > 4$ .

Following Wu (2005), we impose conditions on the dependence structure of the error processes  $\mathcal{E}_i$  in terms of the physical dependence measure  $\delta_q(G_i, t)$ . In particular, we assume the following:

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

The conditions (C1)–(C3) are fulfilled by a wide range of stationary processes  $\mathcal{E}_i$ . For a detailed discussion of these properties, see Khismatullina and Vogt (2018).

Regarding the independent variables  $\mathbf{X}_{it}$ , we need the following additional assumptions for each  $i$ :



- (C4) The covariates  $\mathbf{X}_{it}$  allow for the representation  $\mathbf{X}_{it} = \mathbf{H}_i(\dots, u_{it-1}, u_{it})$  with  $u_{it}$  being i.i.d. random variables and  $\mathbf{H}_i := (H_{i1}, H_{i2}, \dots, H_{id})^\top : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^d$  is a measurable function such that  $\mathbf{H}_i(\mathcal{U}_{it})$  is well defined. Denote  $\mathcal{U}_{it} = (\dots, u_{it-1}, u_{it})$ .
- (C5) Let  $N_i$  be the  $d \times d$  matrix with  $kl$ -th entry  $n_{i,kl} = \mathbb{E}[H_{ik}(\mathcal{U}_{i0})H_{il}(\mathcal{U}_{i0})]$ . We assume that the smallest eigenvalue of  $N_i$  is strictly bigger than 0.
- (C6) Let  $\mathbb{E}[\mathbf{H}_i(\mathcal{U}_{i0})] = \mathbf{0}$  and  $\|\mathbf{H}_i(\mathcal{U}_{it})\|_4 < \infty$ .
- (C7)  $\sum_{s=0}^{\infty} \delta_4(\mathbf{H}_i, s) < \infty$ .

To be able to prove the main theorems in Section 3, we need additional assumptions on the relationship between the covariates and the error process.

- (C8)  $\mathbf{X}_{it}$  (elementwise) and  $\varepsilon_{is}$  are uncorrelated for each  $t, s \in \{1, \dots, T\}$ .
- (C9) Let  $\zeta_{i,t} = (u_{it}, \eta_{it})^\top$ . Define  $\mathcal{I}_{i,t} = (\dots, \zeta_{i,t-1}, \zeta_{i,t})$  and  $\mathbf{U}_i(\mathcal{I}_{i,t}) = \mathbf{H}_i(\mathcal{U}_{it})G_i(\mathcal{J}_{it})$ . Then,  $\sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) < \infty$ .

### 3 Testing for equality of time trends

In this section, we adapt the multiscale method developed in Khismatullina and Vogt (2018) to the problem of comparison of the trend curves  $m_i$  in model (2.1). 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 what follows, we describe the construction of the test statistic that addresses the question of comparing different trend curves. More specifically, we test the null hypothesis  $H_0 : m_1 = m_2 = \dots = m_n$  in model (2.1). We assume that all the trend functions  $m_i(\cdot)$  are continuously differentiable on  $[0, 1]$ .

It is obvious that if  $\alpha_i$  and  $\beta_i$  are known, the problem of testing for the common time trend would be greatly simplified. That is, we would test  $H_0 : m_1 = m_2 = \dots = m_n$  in the model

$$\begin{aligned} Y_{it} - \alpha_i - \beta_i^\top \mathbf{X}_{it} &=: Y_{it}^\circ \\ &= m_i\left(\frac{t}{T}\right) + \varepsilon_{it}, \end{aligned}$$

which is a standard nonparametric regression equation. The variables  $Y_{it}^\circ$  are not observed since the intercept  $\alpha_i$  and the coefficients  $\beta_i$  are not known. Given appropriate estimators  $\hat{\beta}_i$  and  $\hat{\alpha}_i$ , we can then consider

$$\hat{Y}_{it} := Y_{it} - \hat{\alpha}_i - \hat{\beta}_i^\top \mathbf{X}_{it} = (\beta_i - \hat{\beta}_i)^\top \mathbf{X}_{it} + m_i\left(\frac{t}{T}\right) + (\alpha_i - \hat{\alpha}_i) + \varepsilon_{it}.$$

Then our unobserved variables  $Y_{it}^\circ$  can be approximated by  $\hat{Y}_{it}$  and we compute our test statistic based on  $\hat{Y}_{it}$ . In what follows, we assume that an estimator with the property that  $\beta_i - \hat{\beta}_i = O_P(T^{-1/2})$  is given. Details on one of the possible ways to construct  $\hat{\beta}_i$  are deferred to Section 3.4.

Given  $\hat{\beta}_i$ , consider an appropriate estimator  $\hat{\alpha}_i$  for the intercept  $\alpha_i$  calculated by

$$\begin{aligned} \hat{\alpha}_i &= \frac{1}{T} \sum_{t=1}^T (Y_{it} - \hat{\beta}_i^\top \mathbf{X}_{it}) = \frac{1}{T} \sum_{t=1}^T (\beta_i^\top \mathbf{X}_{it} - \hat{\beta}_i^\top \mathbf{X}_{it} + \alpha_i + m_i(t/T) + \varepsilon_{it}) = \\ &= (\beta_i - \hat{\beta}_i)^\top \frac{1}{T} \sum_{t=1}^T \mathbf{X}_{it} + \alpha_i + \frac{1}{T} \sum_{i=1}^T m_i(t/T) + \frac{1}{T} \sum_{i=1}^T \varepsilon_{it}. \end{aligned} \quad (3.1)$$

Note that  $\frac{1}{T} \sum_{i=1}^T \varepsilon_{it} = O_P(T^{-1/2})$  and  $\frac{1}{T} \sum_{i=1}^T m_i(t/T) = O(T^{-1})$  due to Lipschitz continuity of  $m_i$  and normalization  $\int_0^1 m_i(u) du = 0$ . Furthermore,  $\frac{1}{T} \sum_{t=1}^T \mathbf{X}_{it} = O_P(1)$  by Chebyshev's inequality and  $\hat{\beta}_i - \beta_i = O_P(T^{-1/2})$ . Plugging all these results together in (3.1), we get that  $\hat{\alpha}_i - \alpha_i = O_P(T^{-1/2})$ . Thus, the unobserved variable  $Y_{it}^\circ := Y_{it} - \beta_i^\top \mathbf{X}_{it} - \alpha_i = m_i(t/T) + \varepsilon_{it}$  can be well approximated by  $\hat{Y}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_i^\top \mathbf{X}_{it} = Y_{it}^\circ + O_P(T^{-1/2})$ .

We further let  $\hat{\sigma}_i^2$  be an estimator of the long-run error variance  $\sigma_i^2 = \sum_{\ell=-\infty}^{\infty} \text{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$  which is computed from the constructed sample  $\{\hat{Y}_{it} : 1 \leq t \leq T\}$ . We thus regard  $\hat{\sigma}_i^2 = \hat{\sigma}_i^2(\hat{Y}_{i1}, \dots, \hat{Y}_{iT})$  as a function of the variables  $\hat{Y}_{it}$  for  $1 \leq t \leq T$ . Throughout the section, we assume that  $\hat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$  with  $\rho_T = o(1/\log T)$ . Details on how to construct estimators of  $\sigma_i^2$  are deferred to Section ??.

We are now ready to introduce the multiscale statistic for testing the hypothesis  $H_0 : m_1 = m_2 = \dots = m_n$ . For any pair of time series  $i$  and  $j$ , we define the kernel averages

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

where  $w_{t,T}(u, h)$  are the local linear kernel weights calculated by the following formula.

$$w_{t,T}(u, h) = \frac{\Lambda_{t,T}(u, h)}{\{\sum_{t=1}^T \Lambda_{t,T}(u, h)^2\}^{1/2}}, \quad (3.2)$$

where

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

$S_{T,\ell}(u, h) = (Th)^{-1} \sum_{t=1}^T K(\frac{t}{h} - u)(\frac{t}{h})^\ell$  for  $\ell = 0, 1, 2$  and  $K$  is a kernel function with the following properties:

- (C10) The kernel  $K$  is non-negative, symmetric about zero and integrates to one. Moreover, it has compact support  $[-1, 1]$  and is Lipschitz continuous, that is,  $|K(v) - K(w)| \leq C|v - w|$  for any  $v, w \in \mathbb{R}$  and some constant  $C > 0$ .

The kernel average  $\hat{\psi}_{ij,T}(u, h)$  can be regarded as measuring the distance between the two trend curves  $m_i$  and  $m_j$  on the interval  $[u - h, u + h]$ .

We now combine the test statistics  $\hat{\psi}_{ij,T}(u, h)$  for a wide range of different locations  $u$  and bandwidths or scales  $h$  in a following way:

$$\hat{\Psi}_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\hat{\psi}_{ij,T}(u, h)}{(\hat{\sigma}_i^2 + \hat{\sigma}_j^2)^{1/2}} \right| - \lambda(h) \right\},$$

where  $\lambda(h) = \sqrt{2 \log\{1/(2h)\}}$  and the set  $\mathcal{G}_T$  is the set of points  $(u, h)$  that are taken into consideration. The statistic  $\hat{\Psi}_{ij,T}$  can be interpreted as a global distance measure between the two curves  $m_i$  and  $m_j$ . Thus, the multiscale statistic  $\hat{\Psi}_{ij,T}$  simultaneously takes into account all locations  $u$  and bandwidths  $h$  with  $(u, h) \in \mathcal{G}_T$ . Throughout the paper, we suppose that  $\mathcal{G}_T$  is some subset of  $\mathcal{G}_T^{\text{full}} = \{(u, h) : u = t/T \text{ for some } 1 \leq t \leq T \text{ and } h \in [h_{\min}, h_{\max}]\}$ , where  $h_{\min}$  and  $h_{\max}$  denote some minimal and maximal bandwidth value, respectively. For our theory to work, we require the following conditions to hold:

- (C11)  $|\mathcal{G}_T| = O(T^\theta)$  for some arbitrarily large but fixed constant  $\theta > 0$ , where  $|\mathcal{G}_T|$  denotes the cardinality of  $\mathcal{G}_T$ .
- (C12)  $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$ , that is,  $h_{\min}/\{T^{-(1-\frac{2}{q})} \log T\} \rightarrow \infty$  with  $q > 4$  defined in (C2) and  $h_{\max} < 1/2$ .

We finally define the multiscale statistic for testing the null hypothesis  $H_0 : m_1 = m_2 = \dots = m_n$  as

$$\hat{\Psi}_{n,T} = \max_{1 \leq i < j \leq n} \hat{\Psi}_{ij,T},$$

that is, we define it as the maximal distance  $\hat{\Psi}_{ij,T}$  between any pair of curves  $m_i$  and  $m_j$  with  $i \neq j$ .

### 3.2 The test procedure

Let  $Z_{it}$  for  $1 \leq t \leq T$  and  $1 \leq i \leq n$  be independent standard normal random variables which are independent of the error terms  $\varepsilon_{it}$  and the covariates  $\mathbf{X}_{it}$ . Denote the empirical average of the variables  $Z_{i1}, \dots, Z_{iT}$  by  $\bar{Z}_{i,T} = T^{-1} \sum_{t=1}^T Z_{it}$ . To simplify notation, we

write  $\bar{Z}_i = \bar{Z}_{i,T}$  in what follows. For each  $i$  and  $j$ , we introduce the Gaussian statistic

$$\Phi_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{(\hat{\sigma}_i^2 + \hat{\sigma}_j^2)^{1/2}} \right| - \lambda(h) \right\},$$

where  $\phi_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \{ \hat{\sigma}_i(Z_{it} - \bar{Z}_i) - \hat{\sigma}_j(Z_{jt} - \bar{Z}_j) \}$ . Moreover, we define the statistic

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \Phi_{ij,T} \quad (3.3)$$

and denote its  $(1 - \alpha)$ -quantile by  $q_{n,T}(\alpha)$ . Our multiscale test of the hypothesis  $H_0 : m_1 = m_2 = \dots = m_n$  is defined as follows: For a given significance level  $\alpha \in (0, 1)$ , we reject  $H_0$  if  $\hat{\Psi}_{n,T} > q_{n,T}(\alpha)$ .

### 3.3 Theoretical properties of the test

To start with, we introduce the auxiliary statistic

$$\hat{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \hat{\Phi}_{ij,T}, \quad (3.4)$$

where

$$\hat{\Phi}_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\hat{\phi}_{ij,T}(u,h)}{\{\hat{\sigma}_i^2 + \hat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\}$$

and  $\hat{\phi}_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \{ (\varepsilon_{it} - \bar{\varepsilon}_i) + (\beta_i - \hat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j) - (\beta_j - \hat{\beta}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j) \}$  with  $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} = T^{-1} \sum_{t=1}^T \varepsilon_{it}$  and  $\bar{\mathbf{X}}_i = \bar{\mathbf{X}}_{i,T} = T^{-1} \sum_{t=1}^T \mathbf{X}_{it}$  respectively. Our first theoretical result characterizes the asymptotic behaviour of the statistic  $\hat{\Phi}_{n,T}$ .

**Theorem 3.1.** *Suppose that the error processes  $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$  are independent across  $i$  and satisfy (C1)–(C3) for each  $i$ . Moreover, let (C4)–(C12) be fulfilled and assume that  $\hat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$  with  $\rho_T = o(1/\log T)$  for each  $i$ . Then*

$$\mathbb{P}(\hat{\Phi}_{n,T} \leq q_{n,T}(\alpha)) = (1 - \alpha) + o(1)$$

Theorem 3.1 is the main stepping stone to derive the theoretical properties of our multiscale test. The proof of the theorem is provided in the Section ??.

### 3.4 Estimation of the parameters $\beta_i$

We now focus on finding an appropriate estimator  $\hat{\beta}_i$  of  $\beta_i$ . For that purpose, for each  $i$  we consider the time series  $\{\Delta Y_{it} : 2 \leq t \leq T\}$  of the differences  $\Delta Y_{it} = Y_{it} - Y_{it-1}$ .

We then have

$$\Delta Y_{it} = Y_{it} - Y_{it-1} = \beta_i^\top \Delta \mathbf{X}_{it} + \left( m_i\left(\frac{t}{T}\right) - m_i\left(\frac{t-1}{T}\right) \right) + \Delta \varepsilon_{it},$$

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

$$\Delta Y_{it} = \beta_i^\top \Delta \mathbf{X}_{it} + \Delta \varepsilon_{it} + O\left(\frac{1}{T}\right). \quad (3.5)$$

In particular, for each  $i$  we employ the least squares estimation method to estimate  $\beta_i$  in (3.5), treating  $\Delta \mathbf{X}_{it}$  as the regressors and  $\Delta Y_{it}$  as the response variable. That is, we propose the following differencing estimator:

$$\hat{\beta}_i = \left( \sum_{t=2}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^\top \right)^{-1} \sum_{t=2}^T \Delta \mathbf{X}_{it} \Delta Y_{it} \quad (3.6)$$

Then the asymptotic consistency for this differencing estimator is given by the following theorem:

**Theorem 3.2.** *Under Assumptions (C1) - (C9), we have*

$$\beta_i - \hat{\beta}_i = O_P\left(\frac{1}{\sqrt{T}}\right),$$

where  $\hat{\beta}_i$  is the differencing estimator given by (3.6).

## 4 Appendix

### 4.1 Proof of Theorem 3.1

In this section, we prove the theoretical results from Section 3. We use the following notation: The symbol  $C$  denotes a universal real constant which may take a different value on each occurrence. For  $a, b \in \mathbb{R}$ , we write  $a_+ = \max\{0, a\}$  and  $a \vee b = \max\{a, b\}$ . For any set  $A$ , the symbol  $|A|$  denotes the cardinality of  $A$ . The notation  $X \stackrel{\mathcal{D}}{=} Y$  means that the two random variables  $X$  and  $Y$  have the same distribution. Finally,  $f_0(\cdot)$  and  $F_0(\cdot)$  denote the density and the distribution function of the standard normal distribution, respectively.

### Introducing additional intermediate statistic

The auxiliary statistic  $\hat{\Phi}_{n,T}$  defined in Section 3.3 (which is basically the version of the main statistic  $\hat{\Psi}_{n,T}$  under the null hypothesis) heavily depends on the known covariates

$\mathbf{X}_{it}$ , whereas the Gaussian version  $\Phi_{n,T}$  does not. This is the reason why we need to introduce additional intermediate test statistic that connects  $\widehat{\Phi}_{n,T}$  and  $\Phi_{n,T}$ . Thus, for each  $i$  and  $j$ , consider the following kernel averages:

$$\widehat{\phi}_{ij,T}(u, h) = \sum_{t=1}^T w_{t,T}(u, h) \{(\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j)\}.$$

The intermediate statistic is then defined as

$$\begin{aligned} \widehat{\Phi}_{n,T} &= \max_{1 \leq i < j \leq n} \widehat{\Phi}_{ij,T} \quad \text{with} \\ \widehat{\Phi}_{ij,T} &= \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\widehat{\phi}_{ij,T}(u, h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \right| - \lambda(h) \right\}. \end{aligned} \tag{4.1}$$

This statistic can thus be regarded as an approximation of the statistic  $\widehat{\Phi}_{n,T}$ . Heuristically, the kernel averages  $\widehat{\phi}_{ij,T}(u, h)$  are close to the kernel averages  $\phi_{ij,T}(u, h)$  because of the properties of our estimators  $\widehat{\beta}_i$  and assumptions on  $\mathbf{X}_{it}$  but we will prove it later in this Section.

## Auxiliary results using strong approximation theory

The main purpose of this section is to prove that there is a version of the multiscale statistic  $\widehat{\Phi}_{n,T}$  defined in (4.1) which is close to a Gaussian statistic whose distribution is known. More specifically, we prove the following result.

**Proposition A.1.** *Under the conditions of Theorem 3.1, there exist statistics  $\widetilde{\Phi}_{n,T}$  for  $T = 1, 2, \dots$  with the following two properties: (i)  $\widetilde{\Phi}_{n,T}$  has the same distribution as  $\widehat{\Phi}_{n,T}$  for any  $T$ , and (ii)*

$$|\widetilde{\Phi}_{n,T} - \Phi_{n,T}| = o_p\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \frac{\rho_T}{\sqrt{h_{\min}}} \sqrt{\log T}\right),$$

where  $\Phi_{n,T}$  is a Gaussian statistic as defined in (3.3).

**Proof of Proposition A.1.** For the proof, we draw on strong approximation theory for each stationary process  $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$  that fulfill the conditions (C1)–(C3). By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), the following strong approximation result holds true: On a richer probability space, there exist a standard Brownian motion  $\mathbb{B}$  and a sequence  $\{\widetilde{\varepsilon}_t : t \in \mathbb{N}\}$  such that  $[\widetilde{\varepsilon}_1, \dots, \widetilde{\varepsilon}_T] \stackrel{\mathcal{D}}{=} [\varepsilon_1, \dots, \varepsilon_T]$  for each  $T$  and

$$\max_{1 \leq t \leq T} \left| \sum_{s=1}^t \widetilde{\varepsilon}_s - \sigma \mathbb{B}(t) \right| = o(T^{1/q}) \quad \text{a.s.}, \tag{4.2}$$

where  $\sigma^2 = \sum_{k \in \mathbb{Z}} \text{Cov}(\varepsilon_0, \varepsilon_k)$  denotes the long-run error variance.

We apply this result for each stationary process  $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$  separately. In order to do this, we define

$$\tilde{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \tilde{\Phi}_{ij,T}, \quad (4.3)$$

where

$$\tilde{\Phi}_{ij,T} = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\tilde{\phi}_{ij,T}(u,h)}{\{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\},$$

where  $\tilde{\phi}_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \{(\tilde{\varepsilon}_{it} - \tilde{\varepsilon}_i) - (\tilde{\varepsilon}_{jt} - \tilde{\varepsilon}_j)\}$  and  $\tilde{\sigma}_i^2$  are the same estimators as  $\hat{\sigma}_i^2$  with  $\hat{Y}_{it} = (\beta_i - \hat{\beta}_i)^\top \mathbf{X}_{it} + m_i(t/T) + (\alpha_i - \hat{\alpha}_i) + \varepsilon_{it}$  replaced by  $\tilde{Y}_{it} = (\beta_i - \hat{\beta}_i)^\top \mathbf{X}_{it} + m_i(t/T) + (\alpha_i - \hat{\alpha}_i) + \tilde{\varepsilon}_{it}$  for  $1 \leq t \leq T$ . In addition, we let

$$\begin{aligned} \Phi_{n,T} &= \max_{1 \leq i < j \leq n} \Phi_{ij,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{\{\hat{\sigma}_i^2 + \hat{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\} \\ \Phi_{n,T}^\diamond &= \max_{1 \leq i < j \leq n} \Phi_{ij,T}^\diamond = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_{ij,T}(u,h)}{\{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2\}^{1/2}} \right| - \lambda(h) \right\} \end{aligned}$$

with

$$\phi_{ij,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h) \{\hat{\sigma}_i(Z_{it} - \bar{Z}_i) - \hat{\sigma}_j(Z_{jt} - \bar{Z}_j)\}$$

and  $Z_{it} = \mathbb{B}_i(t) - \mathbb{B}_i(t-1)$ . With this notation, we can write

$$|\tilde{\Phi}_{n,T} - \Phi_{n,T}| \leq |\tilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond| + |\Phi_{n,T}^\diamond - \Phi_{n,T}|. \quad (4.4)$$

First consider  $|\tilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond|$ . Straightforward calculations yield that

$$|\tilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond| \leq \{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2\}^{-1/2} \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} |\tilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)|. \quad (4.5)$$

Using summation by parts,  $(\sum_{i=1}^n a_i b_i = \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + A_n b_n$  with  $A_j = \sum_{j=1}^i a_j$ ) we further obtain that

$$\begin{aligned} |\tilde{\phi}_{ij,T}(u,h) - \phi_{ij,T}(u,h)| &= \\ &= \left| \sum_{t=1}^T w_{t,T}(u,h) \{(\tilde{\varepsilon}_{it} - \tilde{\varepsilon}_i) - (\tilde{\varepsilon}_{jt} - \tilde{\varepsilon}_j) - \hat{\sigma}_i(Z_{it} - \bar{Z}_i) + \hat{\sigma}_j(Z_{jt} - \bar{Z}_j)\} \right| = \\ &= \left| \sum_{t=1}^{T-1} A_{ij,t}(w_{t,T}(u,h) - w_{t+1,T}(u,h)) + A_{ij,T} w_{T,T}(u,h) \right|, \end{aligned}$$

where

$$A_{ij,t} = \sum_{s=1}^t \{(\tilde{\varepsilon}_{is} - \tilde{\varepsilon}_i) - (\tilde{\varepsilon}_{js} - \tilde{\varepsilon}_j) - \hat{\sigma}_i(Z_{is} - \bar{Z}_i) + \hat{\sigma}_j(Z_{js} - \bar{Z}_j)\}.$$

Note that by construction  $A_{ij,T} = 0$  for all pairs  $(i, j)$ . Denoting

$$W_T(u, h) = \sum_{t=1}^{T-1} |w_{t+1,T}(u, h) - w_{t,T}(u, h)|,$$

we have

$$|\tilde{\phi}_{ij,T}(u, h) - \phi_{ij,T}(u, h)| = \left| \sum_{t=1}^{T-1} A_{ij,t} (w_{t,T}(u, h) - w_{t+1,T}(u, h)) \right| \leq W_T(u, h) \max_{1 \leq t \leq T} |A_{ij,t}|. \quad (4.6)$$

Now consider  $\max_{1 \leq t \leq T} |A_{ij,t}|$ :

$$\begin{aligned} \max_{1 \leq t \leq T} |A_{ij,t}| &\leq \max_{1 \leq t \leq T} \left| \sum_{s=1}^t (\tilde{\varepsilon}_{is} - \tilde{\varepsilon}_i) - \hat{\sigma}_i \sum_{s=1}^t \{Z_{is} - \bar{Z}_i\} \right| + \max_{1 \leq t \leq T} \left| \sum_{s=1}^t (\tilde{\varepsilon}_{js} - \tilde{\varepsilon}_j) - \hat{\sigma}_j \sum_{s=1}^t \{Z_{js} - \bar{Z}_j\} \right| \leq \\ &\leq \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \hat{\sigma}_i \sum_{s=1}^t Z_{is} \right| + \max_{1 \leq t \leq T} |t(\tilde{\varepsilon}_i - \hat{\sigma}_i \bar{Z}_i)| + \\ &\quad + \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \hat{\sigma}_j \sum_{s=1}^t Z_{js} \right| + \max_{1 \leq t \leq T} |t(\tilde{\varepsilon}_j - \hat{\sigma}_j \bar{Z}_j)| = \\ &= \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \hat{\sigma}_i \sum_{s=1}^t Z_{is} \right| + T \left| \frac{1}{T} \left( \sum_{s=1}^T \tilde{\varepsilon}_{is} - \hat{\sigma}_i \sum_{s=1}^T Z_{is} \right) \right| + \\ &\quad + \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \hat{\sigma}_j \sum_{s=1}^t Z_{js} \right| + T \left| \frac{1}{T} \left( \sum_{s=1}^T \tilde{\varepsilon}_{js} - \hat{\sigma}_j \sum_{s=1}^T Z_{js} \right) \right| \leq \\ &\leq 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \hat{\sigma}_i \sum_{s=1}^t Z_{is} \right| + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \hat{\sigma}_j \sum_{s=1}^t Z_{js} \right| = \\ &= 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \hat{\sigma}_i \sum_{s=1}^t (\mathbb{B}_i(s) - \mathbb{B}_i(s-1)) \right| + \\ &\quad + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \hat{\sigma}_j \sum_{s=1}^t (\mathbb{B}_j(s) - \mathbb{B}_j(s-1)) \right| = \\ &= 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \hat{\sigma}_i \mathbb{B}_i(t) \right| + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \hat{\sigma}_j \mathbb{B}_j(t) \right| \leq \\ &\leq 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \right| + 2 \max_{1 \leq t \leq T} \left| \sigma_i \mathbb{B}_i(t) - \hat{\sigma}_i \mathbb{B}_i(t) \right| + \\ &\quad + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \sigma_j \mathbb{B}_j(t) \right| + 2 \max_{1 \leq t \leq T} \left| \sigma_j \mathbb{B}_j(t) - \hat{\sigma}_j \mathbb{B}_j(t) \right| \end{aligned}$$

Note that for all  $1 \leq i \leq n$  we have  $\hat{\sigma}_i^2 - \sigma_i^2 = o_P(\rho_T)$  which leads to  $\hat{\sigma}_i - \sigma_i = o_P(\rho_T)$ . Furthermore, regarding the standard Brownian motion  $\mathbb{B}_i(t)$  as a sum of standard



normal random variables, we get the following result:

$$\begin{aligned} \max_{1 \leq t \leq T} |\mathbb{B}_i(t)| &= \max_{1 \leq t \leq T} \left| \sum_{s=1}^t Z_{is} \right| = \max_{1 \leq t \leq T} \left| \sqrt{t} \frac{1}{\sqrt{t}} \sum_{s=1}^t Z_{is} \right| \leq \\ &\leq \max_{1 \leq t \leq T} \sqrt{t} \max_{1 \leq t \leq T} \left| \frac{1}{\sqrt{t}} \sum_{s=1}^t Z_{is} \right| = \sqrt{T} \max_{1 \leq t \leq T} \left| \frac{1}{\sqrt{t}} \sum_{s=1}^t Z_{is} \right| = O_P(\sqrt{T \log T}), \end{aligned}$$

where we used the bound for the maximum of  $T$  standard normal random variables and the fact that for each  $t$ ,  $1 \leq t \leq T$ ,  $\frac{1}{\sqrt{t}} \sum_{s=1}^t Z_{is}$  is a standard normal random variable. Therefore, applying the strong approximation result (4.2), we can infer that

$$\begin{aligned} \max_{1 \leq t \leq T} |A_{ij,t}| &\leq 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{is} - \sigma_i \mathbb{B}_i(t) \right| + 2 |\sigma_i - \hat{\sigma}_i| \max_{1 \leq t \leq T} |\mathbb{B}_i(t)| + \\ &\quad + 2 \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_{js} - \sigma_j \mathbb{B}_j(t) \right| + 2 |\sigma_j - \hat{\sigma}_j| \max_{1 \leq t \leq T} |\mathbb{B}_j(t)| = \\ &= o_P(T^{1/q} + \rho_T \sqrt{T \log T}). \end{aligned} \tag{4.7}$$

Standard arguments show that  $\max_{(u,h) \in \mathcal{G}_T} W_T(u, h) = O(1/\sqrt{Th_{\min}})$ . Plugging (4.7) in (4.6) and then in (4.5), we can thus infer that

$$\begin{aligned} |\tilde{\Phi}_{n,T} - \Phi_{n,T}^\diamond| &\leq \{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2\}^{-1/2} \max_{(u,h) \in \mathcal{G}_T} W_T(u, h) \max_{1 \leq i < j \leq n} \max_{1 \leq t \leq T} |A_{ij,t}| = \\ &= o_P\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \frac{\rho_T}{\sqrt{h_{\min}}} \sqrt{\log T}\right). \end{aligned} \tag{4.8}$$

Now consider  $|\Phi_{n,T}^\diamond - \Phi_{n,T}|$ . Since  $\phi_{ij,T}(u, h)$  is distributed as  $N(0, \hat{\sigma}_i^2 + \hat{\sigma}_j^2)$  for all  $(u, h) \in \mathcal{G}_T$  and all  $1 \leq i < j \leq n$ ,  $|\mathcal{G}_T| = O(T^\theta)$  for some large but fixed constant  $\theta$  by Assumption (C11),  $n$  is fixed and  $\hat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$  as well as  $\tilde{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ , we can establish that

$$|\Phi_{n,T}^\diamond - \Phi_{n,T}| \leq \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \left| \frac{\phi_{ij,T}(u, h)}{\{\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2\}^{1/2}} - \frac{\phi_{ij,T}(u, h)}{\{\hat{\sigma}_i^2 + \hat{\sigma}_j^2\}^{1/2}} \right| = o_P(\rho_T \sqrt{\log T}). \tag{4.9}$$

Plugging (4.8) and (4.9) in (4.4) completes the proof.  $\square$

## Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic  $\Phi_{n,T}$  defined in (3.3). We in particular show that  $\Phi_{n,T}$  does not concentrate too strongly in small regions of the form  $[x - \delta_T, x + \delta_T]$  with  $\delta_T$  converging to zero.

**Proposition A.2.** *Under the conditions of Theorem 3.1, it holds that*

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

where  $\delta_T = T^{1/q} / \sqrt{Th_{\min}} + \rho_T \sqrt{\log T} / h_{\min}$ .

**Proof of Proposition A.2.** We write  $x = (u, h)$  along with  $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \dots, x_p\}$ , where  $p := |\mathcal{G}_T| \leq O(T^\theta)$  for some large but fixed  $\theta > 0$  by our assumptions. Moreover, for  $k = 1, \dots, p$ , we set

$$U_{ij,2k-1} = \frac{\phi_{ij,T}(x_{k1}, x_{k2})}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} - \lambda(x_{k2})$$

$$U_{ij,2k} = -\frac{\phi_{ij,T}(x_{k1}, x_{k2})}{\{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{1/2}} - \lambda(x_{k2})$$

with  $x_k = (x_{k1}, x_{k2})$ . This notation allows us to write

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \max_{1 \leq k \leq 2p} U_{ij,k},$$

where  $(U_{12,1}, \dots, U_{(n-1)n,2p})^\top \in \mathbb{R}^{n(n-1)p}$  is a Gaussian random vector.

The main technical tool for proving Proposition A.2 are anti-concentration bounds for Gaussian random vectors. The following proposition slightly generalizes anti-concentration results derived in Chernozhukov et al. (2015), in particular Theorem 3 therein.

**Proposition A.3.** *Let  $(X_1, \dots, X_p)^\top$  be a Gaussian random vector in  $\mathbb{R}^p$  with  $\mathbb{E}[X_j] = \mu_j$  and  $\text{Var}(X_j) = \sigma_j^2 > 0$  for  $1 \leq j \leq p$ . Define  $\bar{\mu} = \max_{1 \leq j \leq p} |\mu_j|$  together with  $\underline{\sigma} = \min_{1 \leq j \leq p} \sigma_j$  and  $\bar{\sigma} = \max_{1 \leq j \leq p} \sigma_j$ . Moreover, set  $a_p = \mathbb{E}[\max_{1 \leq j \leq p} (X_j - \mu_j)/\sigma_j]$  and  $b_p = \mathbb{E}[\max_{1 \leq j \leq p} (X_j - \mu_j)]$ . For every  $\delta > 0$ , it holds that*

$$\sup_{x \in \mathbb{R}} \mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) \leq C\delta\{\bar{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\},$$

where  $C > 0$  depends only on  $\underline{\sigma}$  and  $\bar{\sigma}$ .

The proof of Proposition A.3 is provided in Khismatullina and Vogt (2018).  $\square$

## Auxiliary results concerning the intermediate statistic

In this section, we establish some properties of the intermediate test statistic  $\widehat{\Phi}_{n,T}$  defined in (4.1) and its relationship with  $\widehat{\Phi}_{n,T}$ . First of all, we prove the proposition that relates the distribution of these two statistics.

**Proposition A.4.** For any  $x \in \mathbb{R}$  and  $\gamma_{n,T} > 0$ , we have

$$\begin{aligned} \mathbb{P}(\widehat{\Phi}_{n,T} \leq x - \gamma_{n,T}) - \mathbb{P}(|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}) &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x) \leq \\ &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x + \gamma_{n,T}) + \mathbb{P}(|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}). \end{aligned}$$

**Proof of Proposition A.4.** From the law of total probability and the monotonic property of the probability function, we have

$$\begin{aligned} \mathbb{P}(\widehat{\Phi}_{n,T} \leq x) &= \mathbb{P}(\widehat{\Phi}_{n,T} \leq x, |\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| \leq \gamma_{n,T}) + \mathbb{P}(\widehat{\Phi}_{n,T} \leq x, |\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}) \leq \\ &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x, \widehat{\Phi}_{n,T} - \gamma_{n,T} \leq \widehat{\Phi}_{n,T} \leq \widehat{\Phi}_{n,T} + \gamma_{n,T}) + \mathbb{P}(|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}) \leq \\ &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x + \gamma_{n,T}) + \mathbb{P}(|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}). \end{aligned}$$

Analogously,

$$\begin{aligned} \mathbb{P}(\widehat{\Phi}_{n,T} \leq x - \gamma_{n,T}) &= \mathbb{P}(\widehat{\Phi}_{n,T} \leq x - \gamma_{n,T}, |\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| \leq \gamma_{n,T}) + \\ &\quad + \mathbb{P}(\widehat{\Phi}_{n,T} \leq x - \gamma_{n,T}, |\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}) \leq \\ &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x - \gamma_{n,T}, \widehat{\Phi}_{n,T} - \gamma_{n,T} \leq \widehat{\Phi}_{n,T} \leq \widehat{\Phi}_{n,T} + \gamma_{n,T}) + \\ &\quad + \mathbb{P}(|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}) \leq \\ &\leq \mathbb{P}(\widehat{\Phi}_{n,T} \leq x) + \mathbb{P}(|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| > \gamma_{n,T}). \end{aligned}$$

Now consider  $|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}|$ . Analogously as in the proof of Proposition A.1, we can establish that

$$|\widehat{\Phi}_{n,T} - \widehat{\Phi}_{n,T}| \leq \{\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\}^{-1/2} \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} |\widehat{\phi}_{ij,T}(u, h) - \widehat{\phi}_{ij,T}(u, h)|,$$

where

$$\widehat{\phi}_{ij,T}(u, h) = \sum_{t=1}^T w_{t,T}(u, h) \{(\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j)\}$$

and

$$\widehat{\phi}_{ij,T}(u, h) = \sum_{t=1}^T w_{t,T}(u, h) \{(\varepsilon_{it} - \bar{\varepsilon}_i) + (\beta_i - \widehat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j) - (\beta_j - \widehat{\beta}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\}.$$

Thus,

$$|\widehat{\phi}_{ij,T}(u, h) - \widehat{\phi}_{ij,T}(u, h)| = \left| \sum_{t=1}^T w_{t,T}(u, h) \{(\beta_i - \widehat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\beta_j - \widehat{\beta}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)\} \right| \leq$$

$$\begin{aligned}
&\leq \sum_{t=1}^T |w_{t,T}(u, h)| |(\beta_i - \hat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i) - (\beta_j - \hat{\beta}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)| \leq \\
&\leq \sum_{t=1}^T |(\beta_i - \hat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)| + \sum_{t=1}^T |(\beta_j - \hat{\beta}_j)^\top (\mathbf{X}_{jt} - \bar{\mathbf{X}}_j)|
\end{aligned}$$

Then

$$\begin{aligned}
|\hat{\Phi}_{n,T} - \Phi_{n,T}| &\leq 2\{\hat{\sigma}_i^2 + \hat{\sigma}_j^2\}^{-1/2} \max_{1 \leq i \leq n} \max_{(u,h) \in \mathcal{G}_T} \sum_{t=1}^T |(\beta_i - \hat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)| = \\
&= 2\{\hat{\sigma}_i^2 + \hat{\sigma}_j^2\}^{-1/2} \max_{1 \leq i \leq n} \sum_{t=1}^T |(\beta_i - \hat{\beta}_i)^\top (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)|
\end{aligned}$$

Furthermore, we know that for all  $i \in \{1, \dots, n\}$  we have  $\beta_i - \hat{\beta}_i = O_P(T^{-1/2})$  and  $\hat{\sigma}_i^2 = \sigma_i^2 + o_P(\rho_T)$ .

Statement of the proposition directly follows from it.  $\square$

### Proof of Theorem 3.1

To prove Theorem 3.1, we make use of the two auxiliary results derived above. By Proposition A.1, there exist statistics  $\tilde{\Phi}_{n,T}$  for  $T = 1, 2, \dots$  which are distributed as  $\hat{\Phi}_T$  for any  $T \geq 1$  and which have the property that

$$|\tilde{\Phi}_{n,T} - \Phi_{n,T}| = o_p\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \frac{\rho_T}{\sqrt{h_{\min}}} \sqrt{\log T}\right), \quad (4.10)$$

where  $\Phi_{n,T}$  is a Gaussian statistic as defined in (3.3). The approximation result (4.10) allows us to replace the multiscale statistic  $\hat{\Phi}_{n,T}$  by an identically distributed version  $\tilde{\Phi}_{n,T}$  which is close to the Gaussian statistic  $\Phi_{n,T}$ .

In the next step, we show that

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\hat{\Phi}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)| = o(1). \quad (4.11)$$

The proof of (4.11) proceeds as follows. Suppose that  $\mathbb{P}(\hat{\Phi}_{n,T}) \geq \mathbb{P}(\Phi_{n,T})$ . Then by Proposition A.4 we have

$$\begin{aligned}
0 &\leq |\mathbb{P}(\hat{\Phi}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)| = \mathbb{P}(\hat{\Phi}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x) \leq \\
&\leq \mathbb{P}(\hat{\Phi}_{n,T} \leq x + \gamma_{n,T}) + \mathbb{P}(|\hat{\Phi}_{n,T} - \Phi_{n,T}| > \gamma_{n,T}) - \mathbb{P}(\Phi_{n,T} \leq x) = \\
&= \mathbb{P}(\hat{\Phi}_{n,T} \leq x + \gamma_{n,T}) - \mathbb{P}(\Phi_{n,T} \leq x) + o(1) =
\end{aligned}$$

$$= \mathbb{P}\left(\widehat{\Phi}_{n,T} \leq x + \gamma_{n,T}\right) - \mathbb{P}\left(\Phi_{n,T} \leq x + \gamma_{n,T}\right) + \mathbb{P}\left(\Phi_{n,T} \leq x + \gamma_{n,T}\right) - \mathbb{P}\left(\Phi_{n,T} \leq x\right) + o(1) \leq \quad (4.12)$$

which immediately implies the statement of Theorem 3.1.

## 4.2 Proof of Theorem 3.2

We define the first-differenced regressors as follows.

$$\Delta \mathbf{X}_{it} = \mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1}) := \Delta \mathbf{H}_i(\mathcal{U}_{it}).$$

Similarly,

$$\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1} = G_i(\mathcal{J}_{it}) - G_i(\mathcal{J}_{it-1}) = \Delta G_i(\mathcal{J}_{it}).$$

With these assumptions we can prove the following propositions.

**Proposition 4.1.** *Under Assumptions (C4) and (C6),  $\|\Delta \mathbf{H}_i(\mathcal{U}_{it})\|_4 < \infty$ .*

**Proof of Proposition 4.1.** By Assumption (C6),

$$\|\Delta \mathbf{H}_i(\mathcal{U}_{it})\|_4 \leq \|\mathbf{H}_i(\mathcal{U}_{it})\|_4 + \|\mathbf{H}_i(\mathcal{U}_{it-1})\|_4 < \infty.$$

□

**Proposition 4.2.** *Under Assumption (C7),  $\Delta \mathbf{X}_{it}$  (elementwise) and  $\Delta \varepsilon_{it}$  are uncorrelated for each  $t \in \{1, \dots, T\}$ .*

**Proof of Proposition 4.2.** By Assumption (C7),

$$\begin{aligned} \mathbb{E}[\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}] &= \mathbb{E}[(\mathbf{X}_{it} - \mathbf{X}_{it-1})(\varepsilon_{it} - \varepsilon_{it-1})] = \\ &= \mathbb{E}[\mathbf{X}_{it} \varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it-1} \varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it} \varepsilon_{it-1}] + \mathbb{E}[\mathbf{X}_{it-1} \varepsilon_{it-1}] = \\ &= \mathbb{E}[\mathbf{X}_{it}] \mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it-1}] \mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\mathbf{X}_{it}] \mathbb{E}[\varepsilon_{it-1}] + \mathbb{E}[\mathbf{X}_{it-1}] \mathbb{E}[\varepsilon_{it-1}] = \\ &= (\mathbb{E}[\mathbf{X}_{it}] - \mathbb{E}[\mathbf{X}_{it-1}]) (\mathbb{E}[\varepsilon_{it}] - \mathbb{E}[\varepsilon_{it-1}]) = \mathbb{E}[\Delta \mathbf{X}_{it}] \mathbb{E}[\Delta \varepsilon_{it}] \end{aligned}$$

□

**Proposition 4.3.** *Define*

$$\Delta \mathbf{U}_i(\mathcal{I}_{i,t}) := \Delta \mathbf{H}_i(\mathcal{U}_{it}) \Delta G_i(\mathcal{J}_{it}).$$

*Under Assumptions (C2) - (C3), (C8) - (C9), we have that  $\sum_{s=0}^{\infty} \delta_2(\Delta \mathbf{U}_i, s) < \infty$ .*

**Proof of Proposition 4.3.** Note the following

$$\begin{aligned}
\delta_2(\Delta \mathbf{U}_i, t) &= \|\Delta \mathbf{U}_i(\mathcal{I}_{i,t}) - \Delta \mathbf{U}_i(\mathcal{I}'_{i,t})\|_2 = \\
&= \|\Delta \mathbf{H}_i(\mathcal{U}_{it})\Delta G_i(\mathcal{J}_{it}) - \Delta \mathbf{H}_i(\mathcal{U}'_{it})\Delta G_i(\mathcal{J}'_{it})\|_2 = \\
&= \|(\mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1}))(\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{G}_i(\mathcal{J}_{it-1})) - (\mathbf{H}_i(\mathcal{U}'_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1}))(\mathbf{G}_i(\mathcal{J}'_{it}) - \mathbf{G}_i(\mathcal{J}'_{it-1}))\|_2 = \\
&= \|\mathbf{H}_i(\mathcal{U}_{it})\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}_{it-1})\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}_{it})\mathbf{G}_i(\mathcal{J}_{it-1}) + \mathbf{H}_i(\mathcal{U}_{it-1})\mathbf{G}_i(\mathcal{J}_{it-1}) - \\
&\quad - \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}'_{it}) + \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}'_{it}) + \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}'_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}'_{it-1})\|_2 \leq \\
&\leq \|\mathbf{H}_i(\mathcal{U}_{it})\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}'_{it})\|_2 + \|\mathbf{H}_i(\mathcal{U}_{it-1})\mathbf{G}_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}'_{it-1})\|_2 + \\
&\quad + \|\mathbf{H}_i(\mathcal{U}_{it-1})\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}'_{it})\|_2 + \|\mathbf{H}_i(\mathcal{U}_{it})\mathbf{G}_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}'_{it-1})\|_2 = \\
&= \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) + \\
&\quad + \|\mathbf{H}_i(\mathcal{U}_{it-1})\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}_{it}) + \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it-1})\mathbf{G}_i(\mathcal{J}'_{it})\|_2 + \\
&\quad + \|\mathbf{H}_i(\mathcal{U}_{it})\mathbf{G}_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}_{it-1}) + \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it})\mathbf{G}_i(\mathcal{J}'_{it-1})\|_2 \leq \\
&\leq \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) + \\
&\quad + \|(\mathbf{H}_i(\mathcal{U}_{it-1}) - \mathbf{H}_i(\mathcal{U}'_{it-1}))\mathbf{G}_i(\mathcal{J}_{it})\|_2 + \|\mathbf{H}_i(\mathcal{U}'_{it-1})(\mathbf{G}_i(\mathcal{J}_{it}) - \mathbf{G}_i(\mathcal{J}'_{it}))\|_2 + \\
&\quad + \|(\mathbf{H}_i(\mathcal{U}_{it}) - \mathbf{H}_i(\mathcal{U}'_{it}))\mathbf{G}_i(\mathcal{J}_{it-1})\|_2 + \|\mathbf{H}_i(\mathcal{U}'_{it})(\mathbf{G}_i(\mathcal{J}_{it-1}) - \mathbf{G}_i(\mathcal{J}'_{it-1}))\|_2 \leq \\
&\leq \delta_2(\mathbf{U}_i, t) + \delta_2(\mathbf{U}_i, t-1) + (\delta_2(\mathbf{H}_i, t-1) + \delta_2(\mathbf{H}_i, t))\|\mathbf{G}_i\|_2 + (\delta_2(\mathbf{G}_i, t-1) + \delta_2(\mathbf{G}_i, t))\|\mathbf{H}_i\|_2
\end{aligned}$$

Here  $\mathcal{U}'_{it} = (\dots, u_{i(-1)}, u'_{i0}, u_{i1}, \dots, u_{it-1}, u_{it})$ ,  $\mathcal{U}'_{it-1} = (\dots, u_{i(-1)}, u'_{i0}, u_{i1}, \dots, u_{it-1})$ ,  $\mathcal{J}'_{it} = (\dots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \dots, \eta_{it-1}, \eta_{it})$ ,  $\mathcal{J}'_{it-1} = (\dots, \eta_{i(-1)}, \eta'_{i0}, \eta_{i1}, \dots, \eta_{it-1})$  are coupled processes with  $u'_{i0}$  being an i.i.d. copy of  $u_{i0}$  and  $\eta'_{i0}$  being an i.i.d. copy of  $\eta_{i0}$ .

This leads us to

$$\begin{aligned}
\sum_{s=0}^{\infty} \delta_2(\Delta \mathbf{U}_i, s) &\leq \sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) + \sum_{s=1}^{\infty} \delta_2(\mathbf{U}_i, s-1) + \\
&\quad + \sum_{s=1}^{\infty} (\delta_2(\mathbf{H}_i, s-1) + \delta_2(\mathbf{H}_i, s))\|\mathbf{G}_i\|_2 + \sum_{s=1}^{\infty} (\delta_2(\mathbf{G}_i, s-1) + \delta_2(\mathbf{G}_i, s))\|\mathbf{H}_i\|_2 < \infty
\end{aligned}$$

□

**Proposition 4.4.** Under Assumptions (C1) - (C9),

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

**Proof of Proposition 4.4.** We need the following notation:

$$\begin{aligned}
\mathcal{P}_{i,t}(\cdot) &:= \mathbb{E}[\cdot | \mathcal{I}_{i,t}] - \mathbb{E}[\cdot | \mathcal{I}_{i,t-1}], \\
\kappa_i &:= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \varepsilon_{it},
\end{aligned}$$

$$\kappa_{i,s}^{\mathcal{P}} := \frac{1}{T} \sum_{t=1}^T \mathcal{P}_{i,t-s}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}).$$

Then,

$$\begin{aligned} \|\kappa_{i,s}^{\mathcal{P}}\|^2 &= \left\| \frac{1}{T} \sum_{t=1}^T \mathcal{P}_{i,t-s}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}) \right\|^2 \leq \\ &\leq \frac{1}{T^2} \sum_{t=1}^T \left\| \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s-1}) \right\|^2 = \\ &= \frac{1}{T^2} \sum_{t=1}^T \left\| \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s} | \mathcal{I}_{i,t-s}) \right\|^2, \end{aligned}$$

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

$$\begin{aligned} \|\kappa_{i,s}^{\mathcal{P}}\|^2 &\leq \frac{1}{T^2} \sum_{t=1}^T \left\| \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s} | \mathcal{I}_{i,t-s}) \right\|^2 \leq \\ &\leq \frac{1}{T^2} \sum_{t=1}^T \left\| \Delta \mathbf{X}_{it} \Delta \varepsilon_{it} - \Delta \mathbf{X}'_{it,s} \Delta \varepsilon'_{it,s} \right\|^2 = \\ &= \frac{1}{T^2} \sum_{t=1}^T \left\| \Delta \mathbf{H}_i(\mathcal{U}_{it}) \Delta G_i(\mathcal{J}_{it}) - \Delta \mathbf{H}_i(\mathcal{U}'_{it,s}) \Delta G_i(\mathcal{J}'_{it,s}) \right\|^2 = \\ &= \frac{1}{T^2} \sum_{t=1}^T \left\| \Delta \mathbf{U}_i(\mathcal{I}_{i,t}) - \Delta \mathbf{U}_i(\mathcal{I}'_{i,t,s}) \right\|^2 \leq \frac{1}{T^2} \sum_{t=1}^T \delta_2^2(\Delta \mathbf{U}_i, s) = \frac{1}{T} \delta_2^2(\Delta \mathbf{U}_i, s) \end{aligned}$$

with  $\mathcal{U}'_{it,s} = (\dots, u_{it-s-1}, u'_{it-s}, u_{it-s+1}, \dots, u_{it})$ ,  $\mathcal{J}'_{it,s} = (\dots, \eta_{it-s-1}, \eta'_{it-s}, \eta_{it-s+1}, \dots, \eta_{it})$ ,  $\zeta'_{it} = (u'_{it}, \eta'_{it})^\top$  and  $\mathcal{I}'_{i,t,s} = (\dots, \zeta_{it-s-1}, \zeta'_{it-s}, \zeta_{it-s+1}, \dots, \zeta_{it})$ .

Moreover,

$$\begin{aligned} \kappa_i - \mathbb{E} \kappa_i &= \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \varepsilon_{it} - \mathbb{E} \kappa_i = \frac{1}{T} \sum_{t=1}^T \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t}) - \mathbb{E} \kappa_i = \\ &= \frac{1}{T} \sum_{t=1}^T (\mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t}) - \mathbb{E}(\mathbf{X}_{it} \Delta \varepsilon_{it})) = \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{s=0}^{\infty} (\mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s}) - \mathbb{E}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it} | \mathcal{I}_{i,t-s-1})) = \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{s=0}^{\infty} \mathcal{P}_{i,t-s}(\Delta \mathbf{X}_{it} \Delta \varepsilon_{it}) = \sum_{s=0}^{\infty} \kappa_{i,s}^{\mathcal{P}}. \end{aligned}$$

Thus, by Proposition 4.3,

$$\|\kappa_i - \mathbb{E}\kappa_i\| \leq \sum_{s=0}^{\infty} \|\kappa_{i,s}^{\mathcal{P}}\| \leq \frac{1}{\sqrt{T}} \sum_{s=0}^{\infty} \delta_2(\Delta \mathbf{U}_i, s) = O\left(\frac{1}{\sqrt{T}}\right)$$

Since  $\mathbb{E}\kappa_i = 0$  by Proposition 4.2, we conclude that

$$\left\| \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \varepsilon_{it} \right\| = O\left(\frac{1}{\sqrt{T}}\right).$$

Therefore, the proposition follows.  $\square$

**Proof of Theorem 3.2.** Recall the differencing estimator  $\hat{\beta}_i$ :

$$\begin{aligned} \hat{\beta}_i &= \left( \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^{\top} \right)^{-1} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta Y_{it} = \\ &= \left( \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^{\top} \right)^{-1} \sum_{t=1}^T \Delta \mathbf{X}_{it} \left( \Delta \mathbf{X}_{it}^{\top} \beta_i + \Delta \varepsilon_{it} + O\left(\frac{1}{T}\right) \right) = \\ &= \beta_i + \left( \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^{\top} \right)^{-1} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \varepsilon_{it} + O\left(\frac{1}{T}\right) \left( \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^{\top} \right)^{-1} \sum_{t=1}^T \Delta \mathbf{X}_{it}. \end{aligned}$$

This leads to

$$\begin{aligned} \sqrt{T}(\hat{\beta}_i - \beta_i) &= \left( \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^{\top} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \varepsilon_{it} + \\ &+ O\left(\frac{1}{\sqrt{T}}\right) \left( \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^{\top} \right)^{-1} \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it}. \end{aligned}$$

Since

$$\mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^T \Delta H_{ij}(\mathcal{U}_{it}) \right] = 0$$

and

$$\text{Var} \left[ \frac{1}{T} \sum_{t=1}^T \Delta H_{ij}(\mathcal{U}_{it}) \right] \leq \frac{4}{T^2} \mathbb{E} [H_{ij}^2(\mathcal{U}_{it})],$$

by Chebyshev's inequality we have that  $\left| \frac{1}{T} \sum_{t=1}^T \Delta H_{ij}(\mathcal{U}_{it}) \right| = O_P(1)$  for each  $j \in \{1, \dots, d\}$ . And this in turn implies that

$$\left| \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{H}_i(\mathcal{U}_{it}) \right| = \left| \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \right| = O_P(1). \quad (4.13)$$

Similarly, by Proposition 4.1 and Chebyshev's inequality, we have that for each  $j, k \in$



$\{1, \dots, d\}$

$$\left| \frac{1}{T} \sum_{t=1}^T \Delta H_{ij}(\mathcal{U}_{it}) \Delta H_{ik}(\mathcal{U}_{it}) \right| = O_P(1),$$

which leads to

$$\left\| \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{H}_i(\mathcal{U}_{it}) \Delta \mathbf{H}_i(\mathcal{U}_{it})^\top \right\| = \left\| \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^\top \right\| = O_P(1), \quad (4.14)$$

where  $\|A\|$  with  $A$  being a matrix is any matrix norm.

By Assumption (C5), we know that  $\mathbb{E}[\Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^\top] = \mathbb{E}[\Delta \mathbf{X}_{i0} \Delta \mathbf{X}_{i0}^\top]$  is invertible, thus,

$$\left\| \left( \frac{1}{T} \sum_{t=1}^T \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^\top \right)^{-1} \right\| = O_P(1).$$

By applying Proposition 4.4, (4.13) and (4.14), the statement of the theorem follows.  $\square$