Multiscale Comparison
of Nonparametric Trend Curves

Marina Khismatullina

Erasmus School of Economics, Erasmus University Rotterdam

and

Michael Vogt*

Department of Mathematics and Economics, Ulm University

April 2, 2024

Abstract

We develop new econometric methods for the comparison of nonparametric time trends. In many applications, practitioners are interested in examining the trending behaviour of the observed time series. Among other things, they would like to know which trends have a different form and in which time intervals the differences occur. We design a multiscale test to formally approach these questions. Specifically, we develop a test which allows to make rigorous confidence statements about which time trends are different and where (i.e., in which time intervals) they differ. Based on our multiscale test, we further develop a clustering algorithm which allows to cluster the observed time series into groups with the same trending behaviour. We derive asymptotic theory for our test and clustering methods. The theory is complemented by a simulation study and two applications to house pricing and GDP growth data.

Keywords: Multiscale statistics; nonparametric regression; time series errors; strong approximations; anti-concentration bounds.

1 Introduction

The comparison of time trends is an important topic in time series analysis and has a wide range of applications in economics, finance and other fields such as climatology. In economics, one may wish to compare trends in real GDP (Grier and Tullock, 1989) or trends in long-term interest rates (Christiansen and Pigott, 1997) across different countries. In finance, it is of interest to compare the volatility trends of different stocks (Nyblom and Harvey, 2000). In climatology, researchers are interested in comparing temperature trends across different spatial locations (Karoly and Wu, 2005).

In this paper, we develop new econometric methods for the comparison of time trends. 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), Chen et al. (2012), Zhang et al. (2012) and Hidalgo and Lee (2014) among others. Following this recent development, we consider a general panel model with deterministic time trends. We observe a panel of n time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T for $1 \leq i \leq n$, where Y_{it} are real-valued random variables and $\boldsymbol{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ are d-dimensional random vectors. Each time series $\mathcal{T}_i$ is modelled by the equation

$$Y_{it} = m_i\left(\frac{t}{T}\right) + \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it} \tag{1.1}$$

for $1 \leq t \leq T$, where $m_i : [0,1] \to \mathbb{R}$ is a nonparametric (deterministic) trend function, $\boldsymbol{\beta}_i$ is a vector of d parameters, $\boldsymbol{X}_{it}$ is a vector of covariates, α_i is a fixed effect and ε_{it} is an idiosyncratic error term. As common in nonparametric regression, the trends m_i in model (1.1) depend on rescaled time t/T rather than on real time t and are thus defined on the unit interval $[0,1]$; see e.g. Robinson (1989), Dahlhaus (1997) and Vogt and Linton (2014) for a discussion of rescaled time. More details on the components of model (1.1) together with the technical assumptions imposed on them can be found in Section 2.

There are various statistical tests for the comparison of time trends in the literature. Most commonly, the null hypothesis $H_0 : m_i = m + c_i \ \forall i \in \{1, \ldots, n\}$ is tested under which the trends m_i are parallel to each other. By a simple normalization of the components in model (1.1), we can always achieve that the trends m_i are exactly the same under H_0 rather than only parallel. Hence, after normalization (the details of which are given in Section 2), the null hypothesis H_0 can be reformulated as follows:

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

The problem of testing H_0 has been studied in a number of articles. In a restricted version of (1.1) without covariates, it has been investigated in Härdle and Marron (1990), Hall and Hart (1990), Degras et al. (2012) and Chen and Wu (2019) among others. Test methods in a less restricted version of (1.1) with covariates and homogeneous parameters ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i) have been derived, e.g., in Zhang et al. (2012) and Hidalgo and Lee (2014).

Even though many different approaches for testing H_0 have been developed over the years, most of them have a serious shortcoming: they are very uninformative. They allow to say (with a certain statistical confidence) whether H_0 is violated or not, but they do not give any information on which types of violations occur. In particular, if H_0 is rejected, they only provide the information that some of the trend curves are not the same. Practitioners, however, would like to have much more specific information: they would like to know (i) *which* trends have a different form and (ii) *in which time intervals* the trends are different. The main aim of this paper is to construct a test procedure which provides exactly this information. More technically speaking, we construct a test which allows to make rigorous simultaneous confidence statements about (i) *which* trend curves are different and (ii) *where* (i.e., in which time intervals) they differ.

Very roughly speaking, our approach works as follows: Let $\{\mathcal{I}_r : 1 \le r \le R\}$ be a large family of time intervals $\mathcal{I}_r \subseteq [0, 1]$ and let $H_{0,r}^{[i,j]} : m_i(w) = m_j(w) \ \forall w \in \mathcal{I}_r$ be the hypothesis

3

that the two trend curves m_i and m_j are identical on the interval $\mathcal{I}_r$. Our procedure simultaneously tests $H_{0,r}^{[i,j]}$ for all r and all pairs of time series (i,j). As the intervals $\mathcal{I}_r$ have different lengths or scales, it is called a multiscale method. Our main theoretical result shows that under suitable technical conditions, the developed multiscale test controls the familywise error rate, i.e., the probability of wrongly rejecting at least one null hypothesis $H_{0,r}^{[i,j]}$. As we will see, this allows us to make simultaneous confidence statements of the following form for a given significance level $\alpha \in (0,1)$:

> *With probability at least $1-\alpha$, the trends m_i and m_j differ on any interval* $\mathcal{I}_r$ *and for any pair (i,j) for which $H_{0,r}^{[i,j]}$ is rejected.* (1.2)

Hence, as desired, we can make confidence statements about (i) *which* trend curves are different and (ii) *where* they differ.

To the best of our knowledge, there are no other test procedures available in the literature which allow to make such confidence statements in the context of the general model (1.1). We are only aware of two other multiscale tests for the comparison of nonparametric time trends, both of which are restricted to a strongly simplified version of model (1.1) without covariates. The first is a SiZer test by Park et al. (2009), for which theory has only been derived in the special case of $n = 2$ time series. The second is a multiscale test by **?**, which was developed to detect differences between epidemic time trends in the context of the COVID-19 pandemic. Being tailored to epidemic count data, their model is quite specific. In particular, it includes neither covariates nor fixed effects and it does not allow for any dependence structures across i and t. In contrast, our model is quite general and thus apt for a wide range of economic and financial applications. Notably, our methodology can be regarded as an extension of that in **?**. However, their theoretical results heavily depend on the specific setting and thus do not carry over easily to our general framework. Our theoretical analysis is thus utterly different from theirs. See Section **??** for a discussion of the main differences.

4

Our multiscale test is constructed step by step in Section 3 and its theoretical properties are laid out in Section 4. Based on this test, we further develop a clustering algorithm which allows to detect groups of time series with the same trend. This clustering algorithm is introduced and investigated in Section 5. The proofs of all theoretical results are relegated to the Supplementary Material. We complement the theoretical analysis of the paper by a simulation study in Section 6 and two application examples in Section 7 and Section A of the Supplement. In the first application example, we examine a long record of housing price data from different countries, whereas the second example is concerned with GDP growth data from a number of OECD countries.

2   Model details

2.1   Notation

For a vector $\mathbf{v} = (v_1, \ldots, v_m) \in \mathbb{R}^m$, we write $\|\mathbf{v}\|_q = \left(\sum_{i=1}^m v_i^q \right)^{1/q}$ to denote its ℓ_q-norm and use the shorthand $\|\mathbf{v}\| = \|\mathbf{v}\|_2$ in the special case $q = 2$. For a random variable V, we define its $\mathcal{L}^q$-norm by $\|V\|_q = (\mathbb{E}|V|^q)^{1/q}$ and write $\|V\| := \|V\|_2$ in the case $q = 2$. Let η_t ($t \in \mathbb{Z}$) be independent and identically distributed (i.i.d.) random variables, write $\mathcal{F}_t = (\ldots, \eta_{t-1}, \eta_t)$ and let $g : \mathbb{R}^\infty \to \mathbb{R}$ be a measurable function such that $g(\mathcal{F}_t) = g(\ldots, \eta_{t-1}, \eta_t)$ is a properly defined random variable. Following Wu (2005), we define the *physical dependence measure* of the process $\{g(\mathcal{F}_t)\}_{t=-\infty}^\infty$ by $\delta_q(g, t) = \|g(\mathcal{F}_t) - g(\mathcal{F}_t')\|_q$, where $\mathcal{F}_t' = (\ldots, \eta_{-1}, \eta_0', \eta_1, \ldots, \eta_t)$ is a coupled version of $\mathcal{F}_t$ with η_0' an i.i.d. copy of η_0.

2.2   Model

We now give some details on the components of model (1.1). We first have a closer look at the error structure which consists of the two components α_i and ε_{it}. We do not impose any

distributional assumptions on the terms α_i. In particular, they can be either deterministic or random and they may be correlated with the covariates $\boldsymbol{X}_{it}$ in an arbitrary way. Following the panel data literature in economics, we refer to them as fixed effects. The terms ε_{it} are idiosyncratic errors which are assumed to have zero mean. They are allowed to be correlated across t but are assumed to be independent across i. Precise technical conditions on them are given in Section 2.3 below. Whereas the idiosyncratic errors ε_{it} are restricted to be independent across i, we do not impose any restrictions on the dependence structure between the fixed effects α_i across i. Hence, we allow for cross-sectional dependence in the error structure via the fixed effects α_i. Similarly, the covariates $\boldsymbol{X}_{it}$ are allowed to be dependent across i in an arbitrary way. As a consequence, the n time series $\mathcal{T}_i$ in our panel can be correlated with each other in various ways.

The trend functions m_i are not identified in model (1.1). In particular, equation (1.1) is equivalent to $Y_{it} = m_i^\diamond(t/T) + \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + \alpha_i^\diamond + \varepsilon_{it}$, where $m_i^\diamond = m_i + c_i^\diamond$, $\alpha_i^\diamond = \alpha_i - c_i^\diamond$ and $c_i^\diamond$ is an arbitrary real constant. For identification, we impose the constraint $\int_0^1 m_i(u)du = 0$ for all i, which vertically shifts each curve m_i such that the overall mean level $\int_0^1 m_i(u)$ is equal to 0. Such a normalization is quite natural in many application contexts. Take e.g. the application on house price data in Section 7. The term $\int_0^1 m_i(u)du$ can be interpreted as the long-term average house price level in country i, which may be very different across countries. The normalization $\int_0^1 m_i(u)du = 0$ corrects for differences in these long-term price levels. Put differently, the normalized trends give the deviation from the long-term price levels.

Assuming that the trends m_i are parallel, i.e., $m_i = m + c_i$ with some constants c_i and a common trend m for all i, the constraint $\int_0^1 m_i(u)du = 0$ implies that $c_i = c := \int_0^1 m(u)du$ for all i. Put differently, it implies that the trends are exactly the same. Hence, under our identification constraint, the null hypothesis that the trends are parallel to each other coincides with the hypothesis that the trends are identical.

Finally note that throughout the paper, we restrict attention to the case where the number of time series n in model (1.1) is fixed. Extending our theoretical results to the case where n grows with T is a possible topic for further research.

2.3 Assumptions

The error processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ satisfy the following conditions.

(C1) The variables ε_{it} allow for the representation $\varepsilon_{it} = g_i(\mathcal{F}_{it})$, where $\mathcal{F}_{it} = (\ldots, \eta_{it-2},$
$\eta_{it-1}, \eta_{it})$, the variables η_{it} are i.i.d. across t, and $g_i : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function such that ε_{it} is well-defined. It holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q \leq C < \infty$ for some $q > 4$ and a sufficiently large constant C.

(C2) The processes $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ are independent across i.

Assumption (C1) implies that the error processes $\mathcal{E}_i$ are stationary and causal (in the sense that ε_{it} does not depend on future innovations η_{is} with $s > t$). The class of error processes that satisfy condition (C1) is very large. It includes linear processes, nonlinear transformations thereof, as well as a large variety of nonlinear processes such as Markov chain models and nonlinear autoregressive models (Wu and Wu, 2016). Following Wu (2005), we impose conditions on the dependence structure of the error processes $\mathcal{E}_i$ in terms of the physical dependence measure $\delta_q(g_i, t)$ defined in Section 2.1:

(C3) For each i, $\sum_{s \geq 0} \delta_q(g_i, s)$ is finite and $\sum_{s \geq t} \delta_q(g_i, s) = O(t^{-\gamma}(\log t)^{-A})$ with q from
(C1), where $A > \frac{2}{3}(1/q + 1 + \gamma)$ and $\gamma = \{q^2 - 4 + (q - 2)\sqrt{q^2 + 20q + 4}\}/8q$.

For fixed i and t, the expression $\sum_{s \geq t} \delta_q(g_i, s)$ measures the cumulative effect of the innovation η_0 on the variables $\varepsilon_{it}, \varepsilon_{it+1}, \ldots$ in terms of the $\mathcal{L}^q$-norm. Condition (C3) puts some restrictions on the decay of $\sum_{s \geq t} \delta_q(g_i, s)$ (as a function of t). It is fulfilled by a wide range of stationary processes $\mathcal{E}_i$. For a detailed discussion of (C1)–(C3) and some examples of error processes that satisfy these conditions, see Khismatullina and Vogt (2020).

The covariates $\boldsymbol{X}_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ are assumed to have the following properties.

(C4) The variables $X_{it,j}$ allow for the representation $X_{it,j} = h_{ij}(\mathcal{G}_{it,j})$, where $\mathcal{G}_{it,j} = (\ldots, \xi_{it-1,j}, \xi_{it,j})$, the random variables $\xi_{it,j}$ are i.i.d. across t and $h_{ij} : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function such that $X_{it,j}$ is well-defined. We use the notation $\boldsymbol{X}_{it} = \boldsymbol{h}_i(\mathcal{G}_{it})$ with $\boldsymbol{h}_i = (h_{i1}, \ldots, h_{id})^\top$ and $\mathcal{G}_{it} = (\mathcal{G}_{it,1}, \ldots, \mathcal{G}_{it,d})^\top$. It holds that $\mathbb{E}[X_{it,j}] = 0$ and $\|X_{it,j}\|_{q'} < \infty$ for all i and j, where $q' > \max\{4, \theta q\}$ with q from (C1) and θ specified in (C10) below.

(C5) The matrix $\mathbb{E}[\Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top]$ is invertible for each i.

(C6) For each i and j, it holds that $\sum_{s \geq 0} \delta_{q'}(h_{ij}, s)$ is finite and $\sum_{s \geq t} \delta_{q'}(h_{ij}, s) = O(t^{-\alpha})$ for some $\alpha > 1/2 - 1/q'$ with q' from (C4).

Assumption (C4) guarantees that the process $\{\boldsymbol{X}_{it} : 1 \leq t \leq T\}$ is stationary and causal for each i. Assumption (C6) restricts the serial dependence of the process $\{\boldsymbol{X}_{it} : 1 \leq t \leq T\}$ for each i in terms of the physical dependence measure.

We finally impose some assumptions on the relationship between the covariates and the errors and on the trend functions m_i.

(C7) The random variables $\Delta \boldsymbol{X}_{it} = \boldsymbol{X}_{it} - \boldsymbol{X}_{it-1}$ and $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$ are uncorrelated, i.e., $\mathrm{Cov}(\Delta \boldsymbol{X}_{it}, \Delta \varepsilon_{it}) = \mathbb{E}[\Delta \boldsymbol{X}_{it} \Delta \varepsilon_{it}] = 0$.

(C8) The trend functions m_i are Lipschitz continuous on $[0,1]$, i.e., $|m_i(v) - m_i(w)| \leq L|v-w|$ for all $v, w \in [0,1]$ and some constant $L < \infty$. Moreover, they are normalised such that $\int_0^1 m_i(u)du = 0$ for each i.

3 The multiscale test

The strategy to construct our multiscale test is as follows: Let $\mathcal{I}_{u,h} := [u - h, u + h] \subseteq [0,1]$ be the rescaled time interval with midpoint $u \in [0,1]$ and length $2h$. We call u the location and h the scale of the interval $\mathcal{I}_{u,h}$. For any given location-scale point (u, h), let

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

be the hypothesis that m_i and m_j are identical on the interval $\mathcal{I}_{u,h}$. $H_0^{[i,j]}(u,h)$ can be viewed as a local null hypothesis that characterises the behaviour of two trend functions locally on the interval $\mathcal{I}_{u,h}$, whereas $H_0 : m_1 = m_2 = \ldots = m_n$ is the global null hypothesis that is concerned with the comparison of all trends on the whole unit interval $[0,1]$. We consider a large family of time intervals $\mathcal{I}_{u,h}$ that fully cover the unit interval. More formally, we consider all intervals $\mathcal{I}_{u,h}$ with $(u,h) \in \mathcal{G}_T$, where $\mathcal{G}_T$ is a set of location-scale points (u,h) with the property that $\bigcup_{(u,h) \in \mathcal{G}_T} \mathcal{I}_{u,h} = [0,1]$. Technical conditions on the set $\mathcal{G}_T$ are given below. The global null H_0 can now be formulated as

H_0 : The hypothesis $H_0^{[i,j]}(u,h)$ holds true for all intervals $\mathcal{I}_{u,h}$ with $(u,h) \in \mathcal{G}_T$

and for all $1 \leq i < j \leq n$.

This formulation shows that we can test H_0 by a procedure which simultaneously tests the local hypothesis $H_0^{[i,j]}(u,h)$ for all $1 \leq i < j \leq n$ and all $(u,h) \in \mathcal{G}_T$.

In what follows, we design such a simultaneous test procedure step by step. To start with, we introduce some auxiliary estimators in Section 3.1. We then construct the test statistics in Section 3.2 and set up the test procedure in Section 3.3. Finally, Section 3.4 explains how to implement the procedure in practice.

3.1 Preliminary steps

If the fixed effects α_i and the coefficient vectors $\boldsymbol{\beta}_i$ were known, our testing problem would be greatly simplified. In particular, we could consider the model $Y_{it}^\circ = m_i(\frac{t}{T}) + \varepsilon_{it}$ with $Y_{it}^\circ := Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it}$, which is a standard nonparametric regression equation. As α_i and $\boldsymbol{\beta}_i$ are not observed in practice, we construct estimators $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$ and replace the unknown variables Y_{it}° by the approximations $\widehat{Y}_{it} := Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it}$. In what follows, we explain how to construct estimators of α_i and $\boldsymbol{\beta}_i$.

In order to estimate the parameter vector $\boldsymbol{\beta}_i$ for a given i, we consider the time series

$\Delta\mathcal{T}_i = \{(\Delta Y_{it}, \Delta \boldsymbol{X}_{it}) : 2 \le t \le T\}$ of the first differences $\Delta Y_{it} = Y_{it} - Y_{it-1}$ and $\Delta \boldsymbol{X}_{it} = \boldsymbol{X}_{it} - \boldsymbol{X}_{it-1}$. We can write $\Delta Y_{it} = Y_{it} - Y_{it-1} = \boldsymbol{\beta}_i^\top \Delta \boldsymbol{X}_{it} + (m_i(\frac{t}{T}) - m_i(\frac{t-1}{T})) + \Delta\varepsilon_{it}$, where $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{it-1}$. Since m_i is Lipschitz by Assumption (C8), we can use the fact that $|m_i(\frac{t}{T}) - m_i(\frac{t-1}{T})| = O(\frac{1}{T})$ to get that $\Delta Y_{it} = \boldsymbol{\beta}_i^\top \Delta \boldsymbol{X}_{it} + \Delta\varepsilon_{it} + O\left(\frac{1}{T}\right)$. We may estimate $\boldsymbol{\beta}_i$ by applying least squares methods to this equation, treating $\Delta \boldsymbol{X}_{it}$ as the regressors and ΔY_{it} as the response variable. As a result, we obtain the estimator

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^T \Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top\Big)^{-1} \sum_{t=2}^T \Delta \boldsymbol{X}_{it} \Delta Y_{it}. \tag{3.1}$$

In Lemma B.5 in the Supplement, we show that $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_p(T^{-1/2})$ under our assumptions. Given $\widehat{\boldsymbol{\beta}}_i$, we next estimate the fixed effect term α_i by

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^T \big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it}\big). \tag{3.2}$$

This is a reasonable estimator of α_i for the following reason: For each i, it holds that $\widehat{\alpha}_i - \alpha_i = \big(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i\big)^\top \frac{1}{T} \sum_{t=1}^T \boldsymbol{X}_{it} + \frac{1}{T} \sum_{t=1}^T m_i(\frac{t}{T}) + \frac{1}{T} \sum_{t=1}^T \varepsilon_{it} = O_p(T^{-1/2})$, since $\frac{1}{T} \sum_{i=1}^T \varepsilon_{it} = O_p(T^{-1/2})$ by the law of large numbers, $\frac{1}{T} \sum_{i=1}^T m_i(t/T) = O(T^{-1})$ due to Lipschitz continuity of m_i and the normalisation $\int_0^1 m_i(u)du = 0$, $\frac{1}{T} \sum_{t=1}^T \boldsymbol{X}_{it} = O_p(1)$ by Chebyshev's inequality and $\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_p(T^{-1/2})$.

Besides $\widehat{\boldsymbol{\beta}}_i$ and $\widehat{\alpha}_i$, we also need an estimator of the long-run error variance $\sigma_i^2 = \sum_{\ell=-\infty}^\infty \text{Cov}(\varepsilon_{i0}, \varepsilon_{i\ell})$ for each i. Throughout the paper, we assume that the long-run error variance is the same for all i, i.e., $\sigma_i^2 = \sigma^2$ for all i. For technical reasons, we nevertheless use a different estimator of σ^2 for each i. In particular, we let $\widehat{\sigma}_i^2$ be an estimator of σ^2 which depends only on the i-th time series $\{\widehat{Y}_{it} : 1 \le t \le T\}$, i.e., $\widehat{\sigma}_i^2 = \widehat{\sigma}_i^2(\widehat{Y}_{i1}, \ldots, \widehat{Y}_{iT})$ is a function of the variables $\widehat{Y}_{it}$ for $1 \le t \le T$.[1] Our theory works with any estimator $\widehat{\sigma}_i^2$ which has the property that $\widehat{\sigma}_i^2 = \sigma^2 + o_p(\rho_T)$ for each i, where ρ_T slowly converges to zero, in particular, $\rho_T = o(1/\log T)$. We now briefly discuss two possible choices of $\widehat{\sigma}_i^2$. Following Kim (2016), we can estimate σ^2 by a variant of the subseries variance estimator, which

---

[1]This construction is for technical convenience. Specifically, it allows us to apply the strong approximation results from Berkes et al. (2014) separately to each time series i, which simplifies the proofs.

was first proposed by Carlstein (1986) and then extended by Wu and Zhao (2007). A full definition of the estimator can be found in Lemma B.6 of the Supplement where we derive its convergence rate. The subseries estimator is completely nonparametric: it does not presuppose a specific time series model for the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ but merely requires $\mathcal{E}_i$ to fulfill general weak dependence conditions. In practice, however, it often makes sense to impose additional structure on the error process $\mathcal{E}_i$. A very popular yet general error model is the class of AR(∞) processes. A difference-based estimator of the long-run error variance for this error model has been developed in Khismatullina and Vogt (2020) and can be easily adapted to the situation at hand. A detailed discussion of the estimator can be found in Section 4 of Khismatullina and Vogt (2020).

3.2   Construction of the test statistics

A statistic to test the local null hypothesis $H_0^{[i,j]}(u,h)$ for a given location-scale point (u,h) and a given pair of time series (i,j) can be constructed as follows: Let $\widehat{m}_{i,T}(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\widehat{Y}_{it}$ be the local linear estimator of m_i constructed from the pre-processed time series $\{\widehat{Y}_{it} : 1 \leq t \leq T\}$. Here, $w_{t,T}(u,h)$ are local linear kernel weights of the form

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^T \Lambda_{t,T}(u,h)^2\}^{1/2}}, \tag{3.3}$$

where

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

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

(C9) The kernel K is non-negative, symmetric about zero and integrates to one. Moreover, it has compact support $[-1,1]$ and is Lipschitz continuous, i.e., $|K(v) - K(w)| \leq C_K|v - w|$ for any $v, w \in \mathbb{R}$ and some constant $C_K > 0$.

We now consider the kernel average

$$\widehat{\psi}_{ij,T}(u,h) := \widehat{m}_{i,T}(u,h) - \widehat{m}_{j,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)(\widehat{Y}_{it} - \widehat{Y}_{jt}), \qquad (3.4)$$

which can be regarded as a measure of distance between the two trend curves m_i and m_j on the interval $\mathcal{I}_{u,h} = [u-h, u+h]$. We may thus use a normalized (absolute) version of $\widehat{\psi}_{ij,T}(u,h)$, in particular, the term

$$\left| \frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \right| \qquad (3.5)$$

as a statistic to test the local null $H_0^{[i,j]}(u,h)$. Our aim is to test $H_0^{[i,j]}(u,h)$ simultaneously for a wide range of location-scale points (u,h) and all possible pairs of time series (i,j). To take into account that we are faced with a simultaneous test problem, we replace the statistics from (3.5) by additively corrected versions of the form

$$\widehat{\psi}_{ij,T}^{0}(u,h) = \left| \frac{\widehat{\psi}_{ij,T}(u,h)}{(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2)^{1/2}} \right| - \lambda(h), \qquad (3.6)$$

where $\lambda(h) = \sqrt{2 \log\{1/(2h)\}}$. The scale-dependent correction term $\lambda(h)$ was first introduced in the multiscale approach of Dümbgen and Spokoiny (2001) and has been used in various contexts since then. We refer to Khismatullina and Vogt (2020) for a detailed discussion of the idea behind this additive correction.

In order to test the global null hypothesis H_0, we aggregate the test statistics $\widehat{\psi}_{ij,T}^{0}(u,h)$ by taking their maximum over all location-scale points $(u,h) \in \mathcal{G}_T$ and all pairs of time series (i,j) with $1 \le i < j \le n$. This leads to the multiscale test statistic

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

For our theoretical results, we suppose that the set of location-scale points $\mathcal{G}_T$ is a subset of $\mathcal{G}_T^{\text{full}} = \{(u,h) : \mathcal{I}_{u,h} = [u-h, u+h] \subseteq [0,1]$ with $u = t/T$ and $h = s/T$ for some $1 \le t, s \le T$ and $h \in [h_{\min}, h_{\max}]\}$ which fulfills the following conditions:

(C10) $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$, where $|\mathcal{G}_T|$ denotes the cardinality of $\mathcal{G}_T$.

(C11) $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$, i.e., $h_{\min}/\{T^{-(1-\frac{2}{q})} \log T\} \to \infty$ with $q > 4$ defined in (C1)

and $h_{\max} = o(1)$.

(C10) and (C11) place relatively mild restrictions on the set $\mathcal{G}_T$: (C10) allows $\mathcal{G}_T$ to be very large compared to the sample size T. In particular, $\mathcal{G}_T$ may grow as any polynomial of T. (C11) allows $\mathcal{G}_T$ to contain intervals $[u - h, u + h]$ of many different scales h, ranging from very small intervals of scale $h_{\min}$ to substantially larger intervals of scale $h_{\max}$. [Add example for grid!]

3.3   The test procedure

Let $\mathcal{M} = \{(u, h, i, j) : (u, h) \in \mathcal{G}_T \text{ and } 1 \leq i < j \leq n\}$ be the collection of all location-scale points (u, h) and all pairs of time series (i, j) under consideration. Moreover, let $\mathcal{M}_0 \subseteq \mathcal{M}$ be the set of tuples (u, h, i, j) for which $H_0^{[i,j]}(u, h)$ is true. For a given significance level $\alpha \in (0, 1)$, our multiscale test is carried out as follows:

(i) For each tuple $(u, h, i, j) \in \mathcal{M}$, we reject the local null hypothesis $H_0^{[i,j]}(u, h)$ if

$$\widehat{\psi}_{ij,T}^0(u, h) > q_{n,T}(\alpha),$$

where the critical value $q_{n,T}(\alpha)$ is constructed below such that the familywise error rate (FWER) is controlled at level α. As usual, the FWER is defined as the probability of wrongly rejecting $H_0^{[i,j]}(u, h)$ for at least one tuple $(u, h, i, j) \in \mathcal{M}$. More formally, it is defined as

$$\text{FWER}(\alpha) = \mathbb{P}\Big(\exists (u, h, i, j) \in \mathcal{M}_0 : \widehat{\psi}_{ij,T}^0(u, h) > q_{n,T}(\alpha)\Big)$$

for a given significance level $\alpha \in (0, 1)$ and we say that the FWER is controlled at level α if $\text{FWER}(\alpha) \leq \alpha$.

(ii) We reject the global null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ if at least one local null

hypothesis is rejected. Put differently, we reject H_0 if

$$\widehat{\Psi}_{n,T} > q_{n,T}(\alpha).$$

We now construct a critical value $q_{n,T}(\alpha)$ which controls the FWER at level α. Let $q_{n,T}(\alpha)$ be the $(1-\alpha)$-quantile of the multiscale statistic $\widehat{\Psi}_{n,T}$ under the global null $H_0 : m_1 = m_2 = \ldots = m_n$. Since

$$\begin{aligned}
\text{FWER}(\alpha) &= 1 - \mathbb{P}\Big(\forall (u,h,i,j) \in \mathcal{M}_0 : \widehat{\psi}^0_{ij,T}(u,h) \le q_{n,T}(\alpha)\Big) \\
&= 1 - \mathbb{P}\Big(\max_{(u,h,i,j) \in \mathcal{M}_0} \widehat{\psi}^0_{ij,T}(u,h) \le q_{n,T}(\alpha)\Big) \\
&\le 1 - \mathbb{P}_{H_0}\Big(\max_{(u,h,i,j) \in \mathcal{M}} \widehat{\psi}^0_{ij,T}(u,h) \le q_{n,T}(\alpha)\Big) \le \alpha
\end{aligned}$$

with $\mathbb{P}_{H_0}$ denoting the probability under H_0, this choice of $q_{n,T}(\alpha)$ indeed controls the FWER at the desired level. However, this choice is not computable in practice. We thus replace it by a suitable approximation: Suppose we could perfectly estimate the parameter vector $\boldsymbol{\beta}_i$ and the long-run error variance σ_i^2 for each i (where by assumption, $\sigma_i^2 = \sigma^2$ for all i), i.e., $\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i$ and $\widehat{\sigma}_i^2 = \sigma_i^2$ for each i. Then under H_0, the test statistics $\widehat{\psi}^0_{ij,T}(u,h)$ would simplify to

$$\widehat{\psi}^0_{ij,T}(u,h) = \left| \frac{\sum_{t=1}^{T} w_{t,T}(u,h)\{(\varepsilon_{it} - \bar{\varepsilon}_i) - (\varepsilon_{jt} - \bar{\varepsilon}_j)\}}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h), \qquad (3.7)$$

where $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ denotes the empirical average of the variables $\varepsilon_{i1}, \ldots, \varepsilon_{iT}$. Replacing the error terms ε_{it} in (3.7) by normally distributed variables leads to the statistics

$$\phi^0_{ij,T}(u,h) = \left| \frac{\phi_{ij,T}(u,h)}{(\sigma_i^2 + \sigma_j^2)^{1/2}} \right| - \lambda(h) \qquad (3.8)$$

with

$$\phi_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) \left\{ \sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j) \right\}, \qquad (3.9)$$

where Z_{it} are independent standard normal random variables for $1 \le t \le T$ and $1 \le i \le n$ and, as before, we use the shorthand $\bar{Z}_i = \bar{Z}_{i,T} := T^{-1} \sum_{t=1}^{T} Z_{it}$. With this notation, we

define

$$\Phi_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \phi_{ij,T}^0(u,h), \tag{3.10}$$

which can be regarded as a Gaussian version of the test statistic $\widehat{\Psi}_{n,T}$ under the null H_0 (in the idealized case with $\widehat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i$ and $\widehat{\sigma}_i^2 = \sigma_i^2$ for all i). We now choose $q_{n,T}(\alpha)$ to be the $(1 - \alpha)$-quantile of $\Phi_{n,T}$.

Remark 3.1. *Importantly, this choice of $q_{n,T}(\alpha)$ can be computed by Monte Carlo simulations in practice: under our assumption that the long-run error variance σ_i^2 does not depend on i (i.e. $\sigma_i^2 = \sigma^2$ for all i), we can rewrite the statistics from (3.8) as*

$$\phi_{ij,T}^0(u,h) = \frac{1}{\sqrt{2}} \Big| \sum_{t=1}^T w_{t,T}(u,h) \left\{ (Z_{it} - \bar{Z}_i) - (Z_{jt} - \bar{Z}_j) \right\} \Big| - \lambda(h).$$

This shows that the distribution of these statistics and thus the distribution of the multiscale statistic $\Phi_{n,T}$ only depend on the Gaussian variables Z_{it} for $1 \leq i \leq n$ and $1 \leq t \leq T$. Consequently, we can approximate the distribution of $\Phi_{n,T}$ – and in particular its quantiles $q_{n,T}(\alpha)$ – by simulating values of the Gaussian random variables Z_{it}. In Section 3.4, we explain in detail how to compute Monte Carlo approximations of the quantiles $q_{n,T}(\alpha)$.

3.4   Implementation of the test in practice

In practice, we implement the test as follows for a given significance level $\alpha \in (0,1)$:

Step 1. Compute the $(1 - \alpha)$-quantile $q_{n,T}(\alpha)$ of the Gaussian statistic $\Phi_{n,T}$ by Monte Carlo simulations. Specifically, draw a large number L (say $L = 5000$) of samples of independent standard normal random variables $\{Z_{it}^{(\ell)} : 1 \leq t \leq T, 1 \leq i \leq n\}$ for $1 \leq \ell \leq L$. For each sample ℓ, compute the value $\Phi_{n,T}^{(\ell)}$ of the Gaussian statistic $\Phi_{n,T}$. Calculate the empirical $(1 - \alpha)$-quantile $\widehat{q}_{n,T}(\alpha)$ from the values $\{\Phi_{n,T}^{(\ell)} : 1 \leq \ell \leq L\}$ and use $\widehat{q}_{n,T}(\alpha)$ as an approximation of the quantile $q_{n,T}(\alpha)$.

Step 2. For each (i,j) with $1 \leq i < j \leq n$ and each $(u,h) \in \mathcal{G}_T$, reject the local null

hypothesis $H_0^{[i,j]}(u,h)$ if $\widehat{\psi}_{ij,T}^0(u,h) > \widehat{q}_{n,T}(\alpha)$. Reject the global null hypothesis H_0 if at least one local hypothesis $H_0^{[i,j]}(u,h)$ is rejected.

Step 3. Display the test results as follows. For each pair of time series (i,j), let $\mathcal{S}^{[i,j]}(\alpha)$ be the set of intervals $\mathcal{I}_{u,h}$ for which we reject $H_0^{[i,j]}(u,h)$. Produce a separate plot for each pair (i,j) which displays the intervals in $\mathcal{S}^{[i,j]}(\alpha)$. This gives a graphical overview over the intervals where our test finds a deviation from the null.

Remark 3.2. *In some cases, the number of intervals in $\mathcal{S}^{[i,j]}(\alpha)$ may be quite large, rendering the visual representation of the test results outlined in Step 3 cumbersome. To overcome this drawback, we replace $\mathcal{S}^{[i,j]}(\alpha)$ with the subset of minimal intervals $\mathcal{S}_{min}^{[i,j]}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$. As in Dümbgen (2002), we call an interval $\mathcal{I}_{u,h} \in \mathcal{S}^{[i,j]}(\alpha)$ minimal if there is no other interval $\mathcal{I}_{u',h'} \in \mathcal{S}^{[i,j]}(\alpha)$ such that $\mathcal{I}_{u',h'} \subset \mathcal{I}_{u,h}$. Notably, our theoretical results remain to hold true when the sets $\mathcal{S}^{[i,j]}(\alpha)$ are replaced by $\mathcal{S}_{min}^{[i,j]}(\alpha)$. See Section 4 for the details.*

4 Theoretical properties of the multiscale test

To start with, we investigate the auxiliary statistic

$$\widehat{\Phi}_{n,T} = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\phi}_{ij,T}^0(u,h), \tag{4.1}$$

where

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

and

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

with $\bar{\varepsilon}_i = \bar{\varepsilon}_{i,T} := T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$ and $\bar{\boldsymbol{X}}_i = \bar{\boldsymbol{X}}_{i,T} := T^{-1} \sum_{t=1}^{T} \boldsymbol{X}_{it}$. By construction, it holds that $\widehat{\phi}_{ij,T}(u,h) = \widehat{\psi}_{ij,T}(u,h)$ under $H_0^{[i,j]}(u,h)$. Hence, the auxiliary statistic $\widehat{\Phi}_{n,T}$ is

identical to the multiscale test statistic $\widehat{\Psi}_{n,T}$ under the global null H_0. Consequently, in order to determine the distribution of the test statistic $\widehat{\Psi}_{n,T}$ under H_0, it suffices to study the auxiliary statistic $\widehat{\Phi}_{n,T}$. The following theorem shows that the distribution of $\widehat{\Phi}_{n,T}$ is close to the distribution of the Gaussian statistic $\Phi_{n,T}$ introduced in (3.10).

Theorem 4.1. *Let (C1)–(C11) be fulfilled. Moreover, assume that $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$ and $\sigma_i^2 = \sigma^2$ for all i. Then*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(\widehat{\Phi}_{n,T} \le x) - \mathbb{P}(\Phi_{n,T} \le x) \right| = o(1).$$

Theorem 4.1 is key for deriving theoretical properties of our multiscale test. Its proof is provided in the Supplementary Material.

Remark 4.2. *The proof of Theorem 4.1 builds on two important theoretical results: strong approximation theory for dependent processes (Berkes et al., 2014) and anti-concentration bounds for Gaussian random vectors (Nazarov, 2003). It can be regarded as a further development of the proof strategy in Khismatullina and Vogt (2020) who developed multiscale methods to test for local increases/decreases of the nonparametric trend function m in the univariate time series model $Y_t = m(t/T) + \varepsilon_t$. We extend their theoretical results in several directions: we consider the case of multiple time series, we allow for a flexible fixed effect error structure and we include covariates, which produces additional approximation error that is quite technical to handle (see Step 4 in the proof of Theorem 4.1 for the details).*

Remark 4.3. *Even though the multiscale test in ? is similar in spirit to ours, their theoretical analysis is completely different from ours. In particular, their proofs build on high-dimensional central limit theorems for hyperrectangles (Chernozhukov et al., 2017), while our proofs rely on strong approximations. Hence, even without the inclusion of covariates and fixed effects, our proof strategy would be completely different. The issue is that the arguments in ? rely heavily on the i.i.d. structure of the data and thus do not carry over*

to the model setting at hand which involves complicated dependence structures both across

i and t.

With the help of Theorem 4.1, we now examine the theoretical properties of our multiscale test developed in Section 3. The following proposition shows that our test of the global null H_0 has correct (asymptotic) size.

Proposition 4.4. *Let the conditions of Theorem 4.1 be satisfied. Then under H_0,*

$$\mathbb{P}\big(\widehat{\Psi}_{n,T} \leq q_{n,T}(\alpha)\big) = (1 - \alpha) + o(1).$$

The next proposition characterises the power of the test against a certain class of local alternatives. To formulate it, we consider a sequence of pairs of functions $m_i := m_{i,T}$ and $m_j := m_{j,T}$ that depend on the time series length T and that are locally sufficiently far from each other.

Proposition 4.5. *Let the conditions of Theorem 4.1 be satisfied. Moreover, assume that for some pair of indices i and j, the functions $m_i = m_{i,T}$ and $m_j = m_{j,T}$ have the following property: There exists $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$ such that $m_{i,T}(w) - m_{j,T}(w) \geq c_T \sqrt{\log T / (Th)}$ for all $w \in [u - h, u + h]$ or $m_{j,T}(w) - m_{i,T}(w) \geq c_T \sqrt{\log T / (Th)}$ for all $w \in [u - h, u + h]$, where $\{c_T\}$ is any sequence of positive numbers with $c_T \to \infty$. Then*

$$\mathbb{P}\big(\widehat{\Psi}_{n,T} \leq q_{n,T}(\alpha)\big) = o(1).$$

We now turn attention to the local null hypotheses $H_0^{[i,j]}(u, h)$. As already defined above, let $\mathcal{M} = \{(u, h, i, j) : (u, h) \in \mathcal{G}_T \text{ and } 1 \leq i < j \leq n\}$ be the collection of location-scale points (u, h) and pairs of time series (i, j) under consideration. Moreover, let $\mathcal{M}_0 \subseteq \mathcal{M}$ be the set of tuples (u, h, i, j) for which $H_0^{[i,j]}(u, h)$ is true. Since we test $H_0^{[i,j]}(u, h)$ simultaneously for all $(u, h, i, j) \in \mathcal{M}$, we would like to bound the probability of making at least one false

discovery. More formally, we would like to control the family-wise error rate

$$\text{FWER}(\alpha) = \mathbb{P}\Big(\exists\,(u,h,i,j) \in \mathcal{M}_0 : \widehat{\psi}^0_{ij,T}(u,h) > q_{n,T}(\alpha)\Big)$$

$$= \mathbb{P}\Big(\exists\,(u,h,i,j) \in \mathcal{M}_0 : \mathcal{I}_{u,h} \in \mathcal{S}^{[i,j]}(\alpha)\Big)$$

at level α, where $\mathcal{S}^{[i,j]}(\alpha)$ is the set of intervals $\mathcal{I}_{u,h}$ for which the null hypothesis $H_0^{[i,j]}(u,h)$ is rejected. The following result shows that our test procedure asymptotically controls the FWER at level α.

Proposition 4.6. *Under the conditions of Theorem 4.1, for any given $\alpha \in (0,1)$,*

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

The statement of Proposition 4.6 can obviously be reformulated as follows:

$$\mathbb{P}\Big(\forall\,(u,h,i,j) \in \mathcal{M} \text{ for which } H_0^{[i,j]}(u,h) \text{ is true, } H_0^{[i,j]}(u,h) \text{ is not rejected}\Big)$$

$$\ge 1 - \alpha + o(1).$$

Consequently, we can make the following simultaneous confidence statement:

> *With (asymptotic) probability at least $1 - \alpha$, the trends m_i and m_j differ on any interval $\mathcal{I}_{u,h}$ and for all pairs (i,j) for which $H_0^{[i,j]}(u,h)$ is rejected.* (4.2)

Put differently:

> *With (asymptotic) probability at least $1 - \alpha$, for any pair (i,j) with $1 \le i < j \le n$, the two trends m_i and m_j differ on all intervals $\mathcal{I}_{u,h} \in \mathcal{S}^{[i,j]}(\alpha)$.* (4.3)

Remark 4.7. *According to (4.3), the graphical device introduced in Section 3.4 which depicts the intervals in $\mathcal{S}^{[i,j]}(\alpha)$ comes with the following statistical guarantee: we can claim with (asymptotic) confidence at least $1 - \alpha$ that the trends m_i and m_j differ on all the depicted intervals. As $\mathcal{S}^{[i,j]}_{min}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ for all i and j, this guarantee remains to hold true when the sets $\mathcal{S}^{[i,j]}(\alpha)$ are replaced by the sets of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$.*

5 Clustering

Consider a situation in which the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is violated. Even though some of the trend functions are different in this case, part of them may still be the same. Put differently, there may be groups of time series which have the same time trend. Formally speaking, we define a group structure as follows: There exist sets or groups of time series $G_1, \ldots, G_N$ with $N \le n$ and $\{1, \ldots, n\} = \dot{\bigcup}_{\ell=1}^{N} G_\ell$ such that for each $1 \le \ell \le N$,

$$m_i = f_\ell \quad \text{for all } i \in G_\ell,$$

where f_ℓ are group-specific trend functions. Hence, the time series which belong to the group G_ℓ all have the same time trend f_ℓ. Throughout the section, we suppose that the group-specific trend functions f_ℓ have the following properties:

(C12) For each ℓ, $f_\ell = f_{\ell,T}$ is a Lipschitz continuous function with $\int_0^1 f_{\ell,T}(w)dw = 0$. In particular, $|f_{\ell,T}(v) - f_{\ell,T}(w)| \le L|v - w|$ for all $v, w \in [0, 1]$ and some constant $L < \infty$ that does not depend on T. Moreover, for any $\ell \ne \ell'$, the trends $f_{\ell,T}$ and $f_{\ell',T}$ differ in the following sense: There exists $(u, h) \in \mathcal{G}_T$ with $\mathcal{I}_{u,h} \subseteq [0, 1]$ such that $f_{\ell,T}(w) - f_{\ell',T}(w) \ge c_T \sqrt{\log T/(Th)}$ for all $w \in \mathcal{I}_{u,h}$ or $f_{\ell',T}(w) - f_{\ell,T}(w) \ge c_T \sqrt{\log T/(Th)}$ for all $w \in \mathcal{I}_{u,h}$, where $0 < c_T \to \infty$.

In many applications, it is natural to suppose that there is a group structure in the data. In this case, a particular interest lies in estimating the unknown groups from the data at hand. In what follows, we combine our multiscale methods with a clustering algorithm to achieve this. More specifically, we use the multiscale statistics $\max_{(u,h) \in \mathcal{G}_T} \hat{\psi}_{ij,T}^0(u, h)$ calculated for each i and j as distance measures which are fed into a hierarchical clustering algorithm.

To describe the algorithm, we first need to introduce the notion of a dissimilarity measure: Let $S \subseteq \{1, \ldots, n\}$ and $S' \subseteq \{1, \ldots, n\}$ be two sets of time series from our sample. We

define a dissimilarity measure between S and S' by setting

$$\widehat{\Delta}(S, S') = \max_{\substack{i \in S, \\ j \in S'}} \left(\max_{(u,h) \in \mathcal{G}_T} \hat{\psi}_{ij,T}^0(u, h) \right). \tag{5.1}$$

This is called a complete linkage measure of dissimilarity. Alternatively, we may work with an average or a single linkage measure. We now combine the dissimilarity measure $\widehat{\Delta}$ with a hierarchical agglomerative clustering (HAC) algorithm which proceeds as follows:

Step 0 (Initialisation): Let $\widehat{G}_i^{[0]} = \{i\}$ denote the i-th singleton cluster for $1 \leq i \leq n$ and define $\{\widehat{G}_1^{[0]}, \ldots, \widehat{G}_n^{[0]}\}$ to be the initial partition of time series into clusters.

Step r (Iteration): Let $\widehat{G}_1^{[r-1]}, \ldots, \widehat{G}_{n-(r-1)}^{[r-1]}$ be the $n - (r - 1)$ clusters from the previous step. Determine the pair of clusters $\widehat{G}_\ell^{[r-1]}$ and $\widehat{G}_{\ell'}^{[r-1]}$ for which

$$\widehat{\Delta}(\widehat{G}_\ell^{[r-1]}, \widehat{G}_{\ell'}^{[r-1]}) = \min_{1 \leq k < k' \leq n-(r-1)} \widehat{\Delta}(\widehat{G}_k^{[r-1]}, \widehat{G}_{k'}^{[r-1]})$$

and merge them into a new cluster.

Iterating this procedure for $r = 1, \ldots, n-1$ yields a tree of nested partitions $\{\widehat{G}_1^{[r]}, \ldots, \widehat{G}_{n-r}^{[r]}\}$, which can be graphically represented by a dendrogram. Roughly speaking, the HAC algorithm merges the n singleton clusters $\widehat{G}_i^{[0]} = \{i\}$ step by step until we end up with the cluster $\{1, \ldots, n\}$. In each step, the closest two clusters are merged, where the distance between clusters is measured in terms of the dissimilarity $\widehat{\Delta}$. We refer the reader to Section 14.3.12 in Hastie et al. (2009) for an overview of hierarchical clustering methods.

When the number of groups N is known, we estimate the group structure $\{G_1, \ldots, G_N\}$ by the N-partition $\{\widehat{G}_1^{[n-N]}, \ldots, \widehat{G}_N^{[n-N]}\}$ produced by the HAC algorithm. When N is unknown, we estimate it by the $\widehat{N}$-partition $\{\widehat{G}_1^{[n-\widehat{N}]}, \ldots, \widehat{G}_{\widehat{N}}^{[n-\widehat{N}]}\}$, where $\widehat{N}$ is an estimator of N. The latter is defined as

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

where we write $\widehat{\Delta}(S) = \widehat{\Delta}(S, S)$ for short and $q_{n,T}(\alpha)$ is the $(1-\alpha)$-quantile of $\Phi_{n,T}$ defined

in Section 3.3.

The following proposition summarises the theoretical properties of the estimators $\widehat{N}$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$, where we use the shorthand $\widehat{G}_\ell = \widehat{G}_\ell^{[n-\widehat{N}]}$ for $1 \leq \ell \leq \widehat{N}$.

Proposition 5.1. *Let the conditions of Theorem 4.1 and (C12) be satisfied. Then*

$$\mathbb{P}\Big(\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, \ldots, G_N\}\Big) \geq (1 - \alpha) + o(1)$$

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

This result allows us to make statistical confidence statements about the estimated clusters $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ and their number $\widehat{N}$. In particular, we can claim with asymptotic confidence at least $1 - \alpha$ that the estimated group structure is identical to the true group structure. The proof of Proposition 5.1 can be found in the Supplement.

Our multiscale methods do not only allow us to compute estimators of the unknown groups $G_1, \ldots, G_N$ and their number N. They also provide information on the locations where two group-specific trend functions f_ℓ and $f_{\ell'}$ differ from each other. To turn this claim into a mathematically precise statement, we need to introduce some notation. First of all, note that the indexing of the estimators $\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}$ is completely arbitrary. We could, for example, change the indexing according to the rule $\ell \mapsto \widehat{N} - \ell + 1$. In what follows, we suppose that the estimated groups are indexed such that $P(\widehat{G}_\ell = G_\ell$ for all $\ell) \geq (1 - \alpha) + o(1)$. Proposition 5.1 implies that this is possible without loss of generality. Keeping this convention in mind, we define the sets

$$\mathcal{A}_{n,T}^{[\ell,\ell']}(\alpha) = \Big\{(u, h) \in \mathcal{G}_T : \widehat{\psi}_{ij,T}^0(u, h) > q_{n,T}(\alpha) \text{ for some } i \in \widehat{G}_\ell, j \in \widehat{G}_{\ell'}\Big\}$$

and

$$\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha) = \Big\{\mathcal{I}_{u,h} = [u - h, u + h] : (u, h) \in \mathcal{A}_{n,T}^{[\ell,\ell']}(\alpha)\Big\}$$

for $1 \leq \ell < \ell' \leq \widehat{N}$. An interval $\mathcal{I}_{u,h}$ is contained in $\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ if our multiscale test indicates a significant difference between the trends m_i and m_j on the interval $\mathcal{I}_{u,h}$ for some $i \in \widehat{G}_\ell$

and $j \in \widehat{G}_{\ell'}$. Put differently, $\mathcal{I}_{u,h} \in \mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ if the test suggests a significant difference between the trends of the ℓ-th and the ℓ'-th group on the interval $\mathcal{I}_{u,h}$. We further let

$$E_{n,T}^{[\ell,\ell']}(\alpha) = \left\{ \forall \mathcal{I}_{u,h} \in \mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha) : f_{\ell}(v) \neq f_{\ell'}(v) \text{ for some } v \in \mathcal{I}_{u,h} = [u-h, u+h] \right\}$$

be the event that the group-specific time trends f_{ℓ} and $f_{\ell'}$ differ on all intervals $\mathcal{I}_{u,h} \in \mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$. With this notation at hand, we can make the following formal statement whose proof is given in the Supplement.

Proposition 5.2. *Under the conditions of Proposition 5.1, the event*

$$E_{n,T}(\alpha) = \left\{ \bigcap_{1 \leq \ell < \ell' \leq \widehat{N}} E_{n,T}^{[\ell,\ell']}(\alpha) \right\} \cap \left\{ \widehat{N} = N \text{ and } \widehat{G}_{\ell} = G_{\ell} \text{ for all } \ell \right\}$$

asymptotically occurs with probability at least $1 - \alpha$, i.e.,

$$\mathbb{P}\big(E_{n,T}(\alpha)\big) \geq (1 - \alpha) + o(1).$$

According to Proposition 5.2, the sets $\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ allow us to locate, with a pre-specified confidence, time intervals where the group-specific trend functions f_{ℓ} and $f_{\ell'}$ differ from each other. In particular, we can claim with asymptotic confidence at least $1 - \alpha$ that the trend functions f_{ℓ} and $f_{\ell'}$ differ on all intervals $\mathcal{I}_{u,h} \in \mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$.[2]

6 Simulations

We now investigate our testing and clustering methods by means of a simulation study. The simulation design is as follows: We generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \beta_i X_{it} + \alpha_i + \varepsilon_{it}$, where the number of time series is set to $n = 15$ and we consider different time series lengths T. For simplicity, we assume that the fixed effect α_i is equal to 0 in all time series and we only include a single regressor X_{it} in the model, setting the unknown parameter β_i to 1 for all i. For each i, the errors ε_{it} follow the AR(1) model $\varepsilon_{it} = a\varepsilon_{i,t-1} + \eta_{it}$,

---

[2]This statement remains to hold true when the sets of intervals $\mathcal{S}_{n,T}^{[\ell,\ell']}(\alpha)$ are replaced by the corresponding sets of minimal intervals.

where $a = 0.25$ and the innovations η_{it} are i.i.d. normally distributed with mean 0 and variance 0.25. Similarly, for each i, the covariates X_{it} follow an AR(1) process of the form $X_{it} = a_x X_{i,t-1} + \zeta_{it}$, where $a_x = 0.5$ and the innovations ζ_{it} are i.i.d. normally distributed with mean 0 and variance 1. We assume that the covariates X_{it} and the error terms ε_{js} are independent from each other for all $1 \leq i, j \leq n$ and $1 \leq t, s \leq T$. To generate data under the null $H_0 : m_1 = \ldots = m_n$, we let $m_i = 0$ for all i without loss of generality. To produce data under the alternative, we define $m_1(u) = b(u - 0.5)$ with $b \in \{0.75, 1, 1.25\}$ and set $m_i = 0$ for all $i \neq 1$. Hence, all trend functions are the same except for m_1 which is an increasing linear function. Note that the normalisation constraint $\int_0^1 m_1(u)du = 0$ is directly satisfied in this case. For each simulation exercise, we simulate 5000 samples.

Our multiscale test is implemented as follows: The estimators $\widehat{\beta}_i$ and $\widehat{\alpha}_i$ of the unknown parameters β_i and α_i are computed as described in Section 3.1. Since the errors ε_{it} follow an AR(1) process, we estimate the long-run error variance σ_i^2 by the difference-based estimator proposed in Khismatullina and Vogt (2020), setting the tuning parameters q and r to 25 and 10, respectively. In order to construct our test statistics, we use an Epanechnikov kernel and the grid $\mathcal{G}_T = U_T \times H_T$ with $U_T = \left\{ u \in [0,1] : u = \frac{5t}{T} \text{ for some } t \in \mathbb{N} \right\}$ and $H_T = \left\{ h \in \left[\frac{\log T}{T}, \frac{1}{4} \right] : h = \frac{5t-3}{T} \text{ for some } t \in \mathbb{N} \right\}$. We thus consider intervals $\mathcal{I}_{u,h} = [u - h, u + h]$ which contain $5, 15, 25, \ldots$ data points. The critical value $q_{n,T}(\alpha)$ of our test is computed by Monte Carlo methods as described in Section 3.4, where we set $L = 5000$.

The simulation results for our multiscale test can be found in Table 1, which reports its actual size under H_0, and in Table 2, which reports its power against different alternatives. Both the actual size and the power are computed as the number of simulations in which the test rejects the global null H_0 divided by the total number of simulations. Inspecting Table 1, one can see that the actual size is fairly close to the nominal target α in all the considered scenarios. Hence, the test has approximately the correct size. Inspecting Table 2, one can further see that the test has reasonable power properties. For the smallest slope

24

Table 1: Size of the multiscale test for different sample sizes T and nominal sizes α.

	nominal size α		
T	0.01	0.05	0.1
100	0.009	0.045	0.087
250	0.013	0.063	0.117
500	0.013	0.057	0.112

Table 2: Power of the multiscale test for different sample sizes T and nominal sizes α. Each panel corresponds to a different slope parameter b.

(a) $b = 0.75$				(b) $b = 1.00$				(c) $b = 1.25$		
	nominal size α				nominal size α				nominal size α	
T	0.01	0.05	0.1	T	0.01	0.05	0.1	T	0.01	0.05	0.1
100	0.033	0.122	0.199	100	0.105	0.270	0.376	100	0.275	0.512	0.628
250	0.209	0.434	0.549	250	0.635	0.840	0.901	250	0.933	0.986	0.993
500	0.741	0.891	0.947	500	0.994	0.999	0.999	500	1.000	1.000	1.000

$b = 0.75$ and the smallest time series length $T = 100$, the power is only moderate, reflecting the fact that the alternative with $b = 0.75$ is not very far away from the null. However, as we increase the slope b and the sample size T, the power increases quickly. Already for the slope value $b = 1.00$, we reach a power of 0.99 for $T = 500$ and for all nominal sizes α.

We next investigate the finite sample performance of the clustering algorithm from Section 5. To do so, we consider a very simple scenario: we generate data from the model $Y_{it} = m_i(\frac{t}{T}) + \varepsilon_{it}$, i.e., we assume that there are no fixed effects and no covariates. The error terms ε_{it} are specified as above. Moreover, as before, we set the number of time series to $n = 15$ and we consider different time series lengths T. We partition the $n = 15$ time series into $N = 3$ groups, each containing 5 time series. Specifically, we set $G_1 = \{1, \ldots, 5\}$, $G_2 = \{6, \ldots, 10\}$ and $G_3 = \{11, \ldots, 15\}$, and we assume that $m_i = f_l$ for all $i \in G_l$ and all $l = 1, 2, 3$. The group-specific trend functions f_1, f_2 and f_3 are defined as $f_1(u) = 0$, $f_2(u) = 1 \cdot (u - 0.5)$ and $f_3(u) = (-1) \cdot (u - 0.5)$. In order to estimate the groups G_1, G_2, G_3 and their number $N = 3$, we use the same implementation as before followed by the clustering procedure from Section 5.

The simulation results are reported in Table 3. The entries in Table 3a are computed as

Table 3: Clustering results for different sample sizes T and nominal sizes α.

(a) Empirical probabilities that
$\widehat{N} = N$

	nominal size α		
T	0.01	0.05	0.1
100	0.055	0.188	0.298
250	0.713	0.922	0.939
500	0.994	0.979	0.956

(b) Empirical probabilities that
$\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, G_2, G_3\}$

	nominal size α		
T	0.01	0.05	0.1
100	0.009	0.045	0.077
250	0.640	0.825	0.845
500	0.992	0.978	0.956

(a) T = 100 (b) T = 250 (c) T = 500
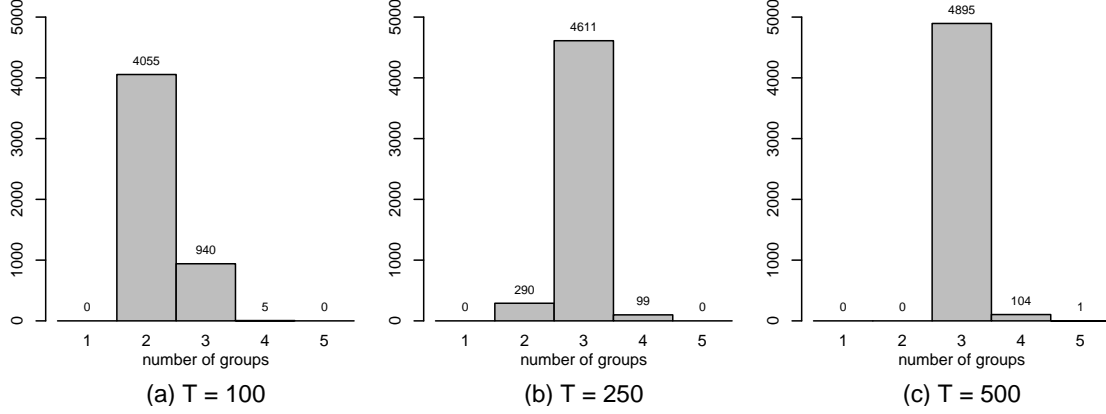
Figure 1: Estimated number of groups $\widehat{N}$ for nominal size $\alpha = 0.05$.
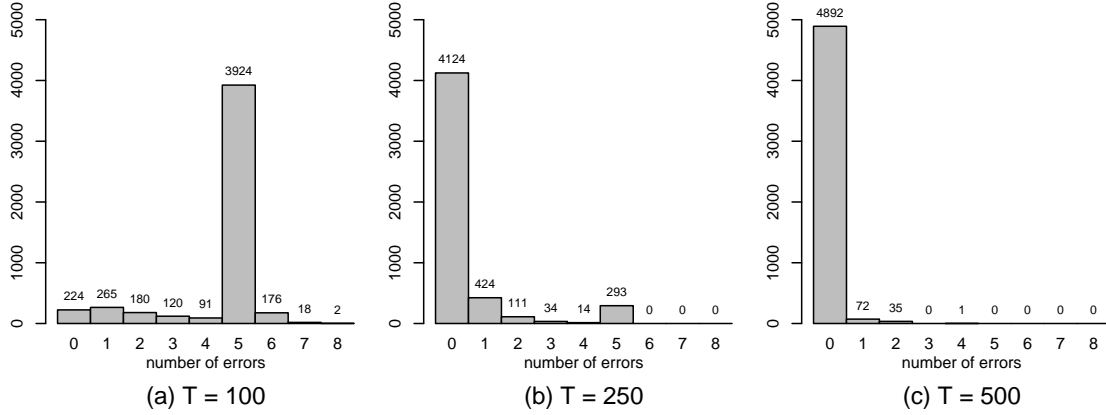

(a) T = 100 (b) T = 250 (c) T = 500

Figure 2: Number of classification errors for nominal size $\alpha = 0.05$.

the number of simulations for which $\widehat{N} = N$ divided by the total number of simulations. They thus specify the empirical probabilities with which the estimate $\widehat{N}$ is equal to the true number of groups $N = 3$. Analogously, the entries of Table 3b give the empirical probabilities with which the estimated group structure $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ equals the true one $\{G_1, G_2, G_3\}$. The results in Table 3 nicely illustrate the theoretical properties of our

clustering algorithm. According to Proposition 5.1, the probability that $\widehat{N} = N$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\} = \{G_1, G_2, G_3\}$ should be at least $(1 - \alpha)$ asymptotically. For the largest sample size $T = 500$, the empirical probabilities reported in Table 3 can indeed be seen to exceed the value $(1 - \alpha)$ as predicted by Proposition 5.1. For the smaller sample sizes $T = 100$ and $T = 250$, in contrast, the empirical probabilities are substantially smaller than $(1 - \alpha)$. This reflects the asymptotic nature of Proposition 5.1 and is not very surprising. It simply mirrors the fact that for the smaller sample sizes $T = 100$ and $T = 250$, the effective noise level in the simulated data is quite high.

Figures 1 and 2 give more insight into what happens for $T = 100$ and $T = 250$. Figure 1 shows histograms of the 5000 simulated values of $\widehat{N}$, while Figure 2 depicts histograms of the number of classification errors produced by our algorithm. By the number of classification errors, we simply mean the number of incorrectly classified time series, which is formally calculated as $\min_{\pi \in S_{\widehat{N}}} \{|G_1 \setminus \widehat{G}_{\pi(1)}| + |G_2 \setminus \widehat{G}_{\pi(2)}| + |G_3 \setminus \widehat{G}_{\pi(3)}|\}$ with $S_{\widehat{N}}$ being the set of all permutations of $\{1, 2, \ldots, \widehat{N}\}$. The histogram of Figure 1 for $T = 100$ clearly shows that our method underestimates the number of groups ($\widehat{N} = 2$ in 4055 cases out of 5000). In particular, it fails to detect the difference between two out of three groups in a large number of simulations. This is reflected in the corresponding histogram of Figure 2 which shows that there are exactly 5 classification errors in 3924 of the 5000 simulation runs. In most of these cases, the estimated group structure $\{\widehat{G}_1, \widehat{G}_2\}$ coincides with either $\{G_1 \cup G_2, G_3\}$, $\{G_1, G_2 \cup G_3\}$ or $\{G_1 \cup G_3, G_2\}$. In summary, we can conclude that the small empirical probabilities for $T = 100$ in Table 3 are due to the algorithm underestimating the number of groups. Inspecting the histograms for $T = 250$, one can see that the performance of the estimators $\widehat{N}$ and $\{\widehat{G}_1, \ldots, \widehat{G}_{\widehat{N}}\}$ improves significantly, even though the corresponding empirical probabilities in Table 3 are still somewhat below the target $(1 - \alpha)$.

7 Empirical application

We analyse a historical dataset on nominal annual house prices from Knoll et al. (2017) that contains data for 14 advanced economies covering 143 years from 1870 to 2012. A second application on GDP growth is provided in the Supplement. In our analysis, we consider 8 countries (Australia, Belgium, Denmark, France, Netherlands, Norway, Sweden and USA) over the time period 1890–2012. The data for all these countries except Belgium contain no missing values. The data for Belgium have five missing values (spanning five years during WWI) which we impute by linear interpolation. The time series of the other 6 countries have more than 10 missings each, which is why we exclude them from the analysis.

We deflate the nominal house prices with the corresponding consumer price index (CPI) to obtain real house prices (HP). Variables that can potentially influence the average house prices are numerous, and there seems to be no general consensus about what the main determinants are. Possible determinants include, but are not limited to, demographic factors such as population growth (Holly et al. 2010, Wang and Zhang 2014, Churchill et al. 2021); fundamental economic factors such as real GDP (Huang et al. 2013, Churchill et al. 2021), interest rate and inflation (Abelson et al. 2005, Otto 2007, Huang et al. 2013, Jordà et al. 2015); urbanisation (Chen et al. 2011, Wang et al. 2017); government subsidies and regulations (Malpezzi 1999); stock markets (Gallin 2006); etc. In our analysis, we focus on the following determinants of the average house prices: real GDP (GDP), population size (POP), long-term interest rate (I) and inflation ($INFL$) which is measured as change in CPI. Most other factors (such as government regulations, construction costs, and real wages) vary rather slowly over time and can be captured by time trend, fixed effects and slope heterogeneity. Data for CPI, real GDP, population size and long-term interest rate are taken from the Jordà-Schularick-Taylor Macrohistory Database[3], which is freely available at `http://www.macrohistory.net/data/` (accessed on 13 January 2022).

---

[3]See Jordà et al. (2017) for a detailed description of the variable construction.

In summary, we observe a panel of $n = 8$ time series $\mathcal{T}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length $T = 123$ for each country $i \in \{1, \dots, 8\}$, where $Y_{it} = \ln HP_{it}$ and $\boldsymbol{X}_{it} = (\ln GDP_{it}, \ln POP_{it}, I_{it}, INFL_{it})^\top$. For each i, the time series $\mathcal{T}_i$ is assumed to follow the model $Y_{it} = m_i(t/T) + \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it}$, or equivalently,

$$\ln HP_{it} = m_i\left(\frac{t}{T}\right) + \beta_{i,1} \ln GDP_{it} + \beta_{i,2} \ln POP_{it} + \beta_{i,3} I_{it} + \beta_{i,4} INFL_{it} + \alpha_i + \varepsilon_{it} \quad (7.1)$$

for $1 \leq t \leq T$, where $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, \beta_{i,3}, \beta_{i,4})^\top$ is a vector of unknown parameters, m_i is a country-specific unknown nonparametric time trend and α_i is a fixed effect.

The inclusion of a nonparametric trend function m_i in model (7.1) is supported by the literature. Ugarte et al. (2009), for example, model the trend in average Spanish house prices by means of splines. Winter et al. (2022) include a long-term stochastic trend component when describing the dynamic behaviour of real house prices in 8 advanced economies. Including a nonparametric trend function when modelling the evolution of house prices is also the main conclusion in Zhang et al. (2016), where it is shown that the time series of logarithmic US house prices is trend-stationary, i.e., can be transformed into a stationary series by subtracting a deterministic trend.

We implement our multiscale test as follows: (i) We choose K to be the Epanechnikov kernel and consider the set of location-scale points $\mathcal{G}_T = U_T \times H_T$, where $U_T = \big\{u \in [0,1] : u = \frac{t}{T} \text{ for some } t \in \mathbb{N}\big\}$ and $H_T = \big\{h \in \big[\frac{\log T}{T}, \frac{1}{4}\big] : h = \frac{5t-3}{T} \text{ for some } t \in \mathbb{N}\big\}$. We thus take into account all locations u on an equidistant grid U_T with step length $1/T$ and all scales $h = 2/T, 7/T, 12/T, \dots$ with $\log T/T \leq h \leq 1/4$. This implies that each interval $\mathcal{I}_{u,h} = [u - h, u + h]$ with $(u, h) \in \mathcal{G}_T$ spans $5, 15, 25, \dots$ years. The lower bound $\log T/T$ is motivated by Assumption (C11). (ii) We assume that for each i, the error process $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ follows an AR(p_i) model and we estimate the long-run variances σ_i^2 by the difference-based estimator from Khismatullina and Vogt (2020) with tuning parameters q and r equal to 15 and 10, respectively. We choose p_i by minimizing

BIC. For 7 out of 8 countries the order p_i determined by BIC[4] is equal to 1. For the sake of simplicity, we thus assume that $p_i = 1$ for all i. (iii) The critical values $q_{n,T}(\alpha)$ are computed by Monte Carlo methods as described in Section 3.4, where we set $L = 5000$. Besides these choices, we implement the test exactly as described in Section 3.

We are now ready to apply the test to the data. The overall null H_0 is rejected at levels $\alpha = 0.05$ and $\alpha = 0.1$. The detailed test results for $\alpha = 0.05$ are presented in Figures 3–6. Each figure corresponds to a specific pair of countries (i, j) for which our test detects differences between the trends. In each figure, panel (a) shows the augmented time series $\{\widehat{Y}_{it} : 1 \leq t \leq T\}$ and $\{\widehat{Y}_{jt} : 1 \leq t \leq T\}$ for the two countries i and j that are compared. Panel (b) presents smoothed versions of the time series from (a), in particular, it shows local linear kernel estimates of the two trend functions m_i and m_j, where the bandwidth window covers 15 years and an Epanechnikov kernel is used. Panel (c) presents the results produced by our test: it depicts in grey the set $\mathcal{S}^{[i,j]}(\alpha)$ of all the intervals for which the test rejects the local null $H_0^{[i,j]}(u, h)$. The set of minimal intervals $\mathcal{S}_{min}^{[i,j]}(\alpha) \subseteq \mathcal{S}^{[i,j]}(\alpha)$ is highlighted in black. According to (4.3), we can make the following simultaneous confidence statement about the intervals plotted in panels (c) of Figures 3–6: we can claim, with confidence of about 95%, that there is a difference between the functions m_i and m_j on each of these intervals.

Overall, our findings are in line with the observations in Knoll et al. (2017) who find considerable cross-country heterogeneity in the house price trends. The authors note that before World War II, the countries exhibit similar trends in real house prices, while the trends start to diverge sometime after World War II. This fits with our findings in Figures 3–4 which show that our test detects differences between the trends of Australia and the Netherlands (of Belgium and the Netherlands) starting only from 1968 (1966) onwards. Contrary to Knoll et al. (2017), however, our test also finds significant differences in the

---

[4]Applying other information criteria such as FPE, AIC and HQ yields exactly the same results in these cases.
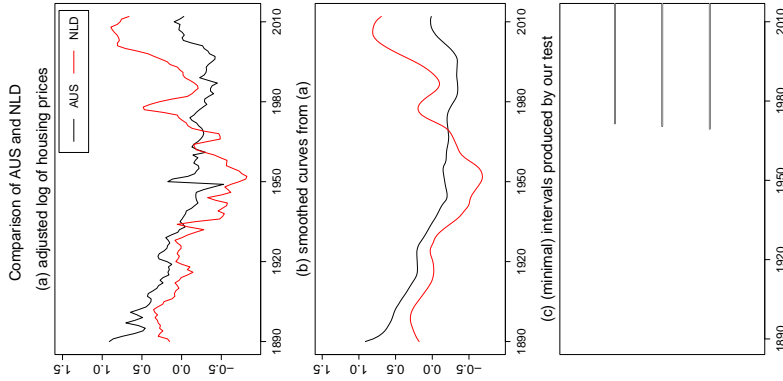
Figure 3: Test results for the comparison of the house prices in Australia and the Netherlands.
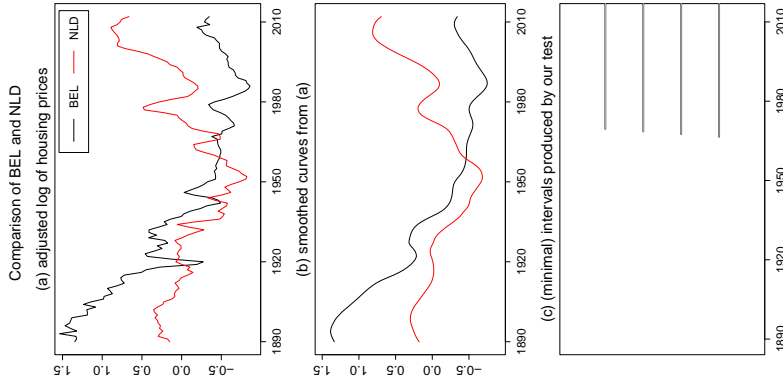
Figure 4: Test results for the comparison of the house prices in Belgium and the Netherlands.
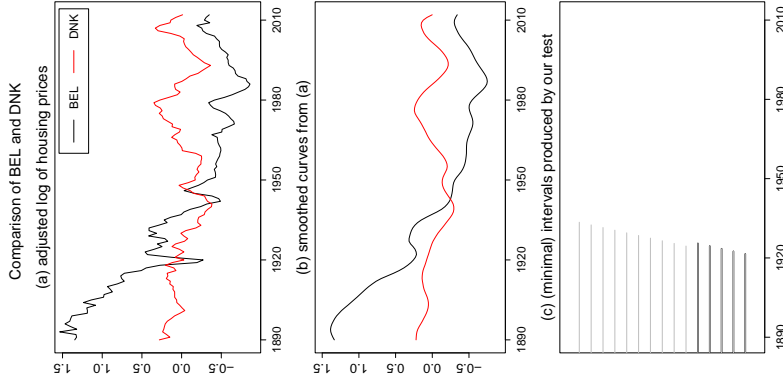
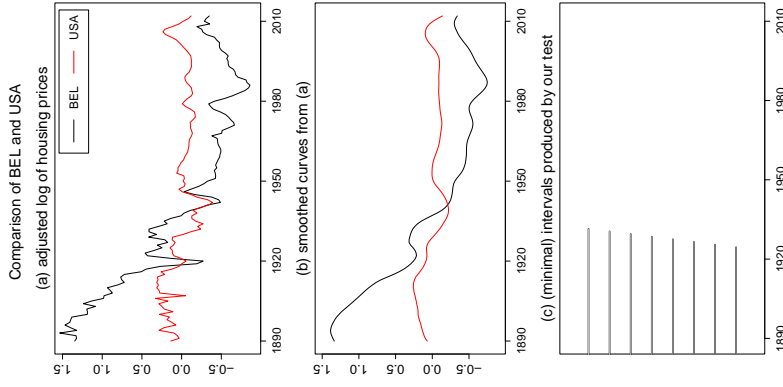Figure 5: Test results for the comparison of the house prices in Belgium and Denmark.

Figure 6: Test results for the comparison of the house prices in Belgium and the USA.

Note: In each figure, panel (a) shows the two augmented time series of the house prices, panel (b) presents smoothed versions of the augmented time series, and panel (c) depicts the set of intervals $\mathcal{S}^{[i,j]}(\alpha)$ in grey and the subset of minimal intervals $\mathcal{S}^{[i,j]}_{min}(\alpha)$ in black.
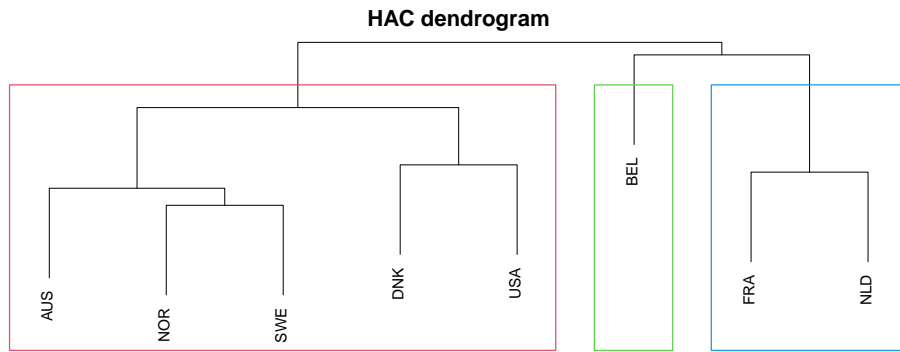
HAC dendrogram

Figure 7: Dendrogram of the HAC algorithm. Each coloured rectangle corresponds to one of the clusters.
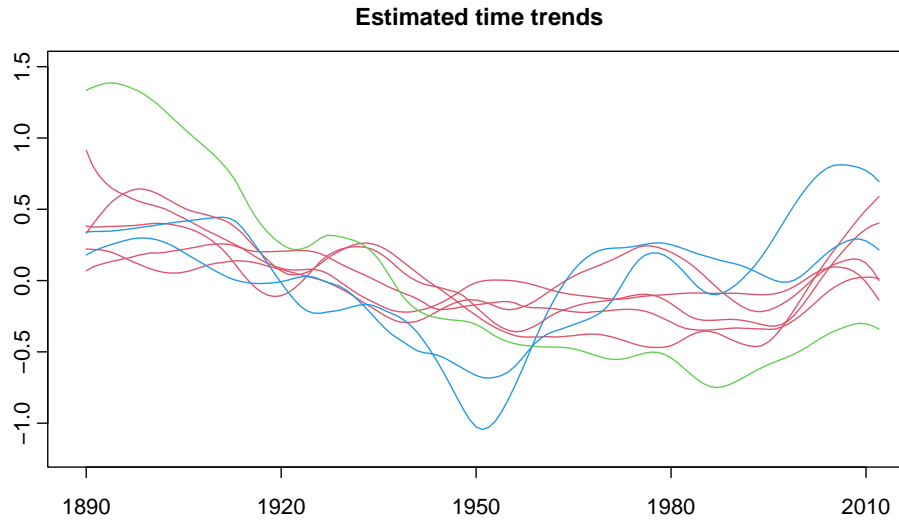
Estimated time trends

Figure 8: Local linear estimates of the $n = 8$ time trends (calculated from the augmented time series $\widehat{Y}_{it}$ with a bandwidth window covering 15 years and an Epanechikov kernel). Each trend estimate is coloured according to the cluster that it is assigned to.

first half of the observed time period, specifically, between the time trends of Belgium and Denmark and of Belgium and the USA (Figures 5 and 6, respectively). This discrepancy in the results is most certainly due to the fact that the method used in Knoll et al. (2017) does not account for effects of other factors such as GDP or population growth, while our test allows to include various determinants of the average house prices in model (7.1).

We next apply the clustering procedure from Section 5 to the data, where we set $\alpha = 0.05$. The results are displayed in Figures 7 and 8. Specifically, Figure 7 shows the dendrogram with the results of the HAC algorithm and Figure 8 presents local linear kernel estimates

of the trend curves (calculated with a bandwidth window of 15 years and an Epanechnikov kernel). The number of clusters is estimated to be $\widehat{N} = 3$. The coloured rectangles in Figure 7 are drawn around the countries that belong to the same cluster and the same colours are used to display the trend estimates in Figure 8.

Inspecting the results, we can see that there is one cluster consisting only of Belgium (plotted in green). Figure 8 suggests that the Belgian time trend indeed evolves somewhat differently from the other trends in the first 30 years of the observed time period. The algorithm further detects a cluster that consists of two countries: France and the Netherlands (plotted in blue). The time trends of these two countries display some kind of dip around 1950, which is not present in the time trends of the other countries. Hence, overall, our algorithm appears to produce a reasonable grouping of the house price trends.

8 Conclusion and extensions

Discuss:

- locally stationary covariates and errors, in particular time-varying error variances

- FWER vs. FDR control

- connection to uniform confidence bands

References

ABELSON, P., JOYEUX, R., MILUNOVICH, G. and CHUNG, D. (2005). Explaining house prices in Australia: 1970–2003. *Economic Record*, **81** S96–S103.

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.

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

CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, **14** 1171–1179.

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

CHEN, J., GUO, F. and WU, Y. (2011). One decade of urban housing reform in China: urban housing price dynamics and the role of migration and urbanization, 1995–2005. *Habitat International*, **35** 1–8.

CHEN, L. and WU, W. B. (2019). Testing for trends in high-dimensional time series. *Journal of the American Statistical Association*, **114** 869–881.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, **45** 2309–2352.

CHRISTIANSEN, H. and PIGOTT, C. (1997). Long-term interest rates in globalised markets.

CHURCHILL, S. A., BAAKO, K. T., MINTAH, K. and ZHANG, Q. (2021). Transport infrastructure and house prices in the long run. *Transport Policy*, **112** 1–12.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, **25** 1–37.

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.

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.

GALLIN, J. (2006). The long-run relationship between house prices and income: evidence from local housing markets. *Real Estate Economics*, **34** 417–438.

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.

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

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*. New York, Springer.

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.

Holly, S., Pesaran, M. H. and Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, **158** 160–173.

Huang, M., Li, R. and Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, **108** 929–941.

Jordà, Ò., Schularick, M. and Taylor, A. M. (2015). Betting the house. *Journal of International Economics*, **96** S2–S18.

Jordà, Ò., Schularick, M. and Taylor, A. M. (2017). Macrofinancial history and the new business cycle facts. *NBER Macroeconomics Annual*, **31** 213–263.

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*, **82** 5–37.

Kim, K. H. (2016). Inference of the trend in a partially linear model with locally stationary regressors. *Econometric Reviews*, **35** 1194–1220.

Knoll, K., Schularick, M. and Steger, T. (2017). No price like home: global house prices, 1870-2012. *American Economic Review*, **107** 331–53.

Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of Housing Economics*, **8** 27–62.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*. Springer, 169–187.

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

Otto, G. (2007). The growth of house prices in Australian capital cities: what do economic fundamentals explain? *Australian Economic Review*, **40** 225–238.

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.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*. Springer, 253–264.

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

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

Ugarte, M. D., Goicoa, T., Militino, A. F. and Durbán, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Computational Statistics & Data Analysis*, **53** 3616–3629.

VOGT, M. and LINTON, O. (2014). Nonparametric estimation of a periodic sequence in the presence of a smooth trend. *Biometrika*, **101** 121–140.

WANG, X.-R., HUI, E. C.-M. and SUN, J.-X. (2017). Population migration, urbanization and housing prices: evidence from the cities in China. *Habitat International*, **66** 49–56.

WANG, Z. and ZHANG, Q. (2014). Fundamental factors in the housing markets of China. *Journal of Housing Economics*, **25** 53–61.

WINTER, J. D., KOOPMAN, S. J. and HINDRAYANTO, I. (2022). Joint decomposition of business and financial cycles: evidence from eight advanced economies. *Oxford Bulletin of Economics and Statistics*, **84** 57–79.

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

WU, W. B. and WU, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics*, **10** 352–379.

WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B*, **69** 391–410.

ZHANG, J., DE JONG, R. and HAURIN, D. (2016). Are US real house prices stationary? New evidence from univariate and panel data. *Studies in Nonlinear Dynamics & Econometrics*, **20** 1–18.

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.