MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

A Nonparametric Panel Model for Climate Data with Seasonal and Spatial Variation

Jiti Gao
Oliver Linton
Bin Peng

Working Paper No.9/22
June 2022

A Nonparametric Panel Model for Climate Data with Seasonal and Spatial Variation

Jiti Gao[⋆] and Oliver Linton[†] and Bin Peng[⋆]
Monash University[⋆] and University of Cambridge[†]

June 28, 2022

Abstract

In this paper, we consider a panel data model which allows for heterogeneous time trends at different locations. We propose a new estimation method for the panel data model before we establish an asymptotic theory for the proposed estimation method. For inferential purposes, we develop a bootstrap method for the case where weak correlation presents in both dimensions of the error terms. We examine the finite–sample properties of the proposed model and estimation method through extensive simulated studies. Finally, we use the newly proposed model and method to investigate rainfall, temperature and sunshine data of U.K. respectively. Overall, we find the weather of winter has changed dramatically over the past fifty years. Changes may vary with respect to locations for the other seasons.

Keywords: Bootstrap method; Interactive fixed–effect; Panel rainfall data, Time trend

JEL classification: Q50, C23

1 Introduction

Time series analysis of climate data has become a big part of statistics and econometrics (e.g., Vogelsang & Franses 2005, Kaufmann et al. 2017 and reference therein); many conferences have been organized on the topic and top journals have published special issues (e.g., Hillebrand et al. 2020, Phillips et al. 2020). Many different methods have been used to identify trends in the presence of seasonal variation over time and spatial comovement based on panel data of outcome variables. We propose a nonparametric panel data model that captures the trend, seasonal, and spatial variation aspect of climate data. Our model allows the trends to vary by location and season and to be of general functional form.

In what follows, we use

$$\{y_{it} \,|\, i \in [N], t \in [T]\}, \tag{1.1}$$

to represent the climate measure recorded by different locations at different times, where $[L] = \{1, \ldots, L\}$ for any positive integer L, different locations are indexed by i, and time periods are indexed by t. Moreover, there are $J+1$ periods[1] in the entire time span, so we group the indices over time as follows.

$$\cup_{j=1}^{J+1} S_j = [T] \quad \text{and} \quad S_j \cap S_{j^*} = \emptyset \text{ for } j \neq j^*.$$

As a data-driven study, J is fixed throughout the paper, and S_j's are known. Our goal is to fit the climate data of (1.1) in a suitable panel data model, so that some useful information can be recovered.

We allow the trends of the outcome variable to vary with respect to locations and season, and we introduce the following heterogeneous time-varying functions:

$$\{m_{ij}(\tau_t) \,|\, i \in [N], t \in [T]\}, \tag{1.2}$$

where $\tau_t = t/T$ is the rescaled time point, and j indexes the group. The weather measurements at different locations are usually more or less connected, and more often than not, are driven by the same heat/cold wave from time to time. Therefore, we turn to the so-called common factor structure (or interactive fixed effects) to mimic these common features. Mathematically, it is written as

$$\{\gamma_i^\top f_t \,|\, i \in [N], t \in [T]\}, \tag{1.3}$$

where both γ_i and f_t are $d_f \times 1$ unobservable vectors, and are to be determined by the data. In (1.3), d_f is a non-negative finite integer. In what follows, we first assume d_f is

[1]Depending on the data availability, we may let $J = 3$ or $J = 11$ to capture seasonality at different frequency.

known in Sections 2.1 and 2.2 for simplicity, and shall work on its estimation in Section 2.3.

In view of the aforementioned arguments, we specifically consider the following nonparametric panel data model:

$$y_{it} = m_{i0}(\tau_t) + \sum_{j=1}^{J} D_{jt} m_{ij}(\tau_t) + \gamma_i^\top f_t + \varepsilon_{it}, \tag{1.4}$$

where $D_{jt} = \mathbb{I}(t \in S_j)$ with $\mathbb{I}(\cdot)$ being the indicator function, $m_{i0}(\cdot)$ stands for the global trend[2] of the i^{th} unit, and $m_{ij}(\cdot)$ with $j \geq 1$ includes the j^{th} periodic trend of the i^{th} unit. We are interested in estimating the trending functions, and in establishing valid inference for the case where weak correlation presents over both dimensions of ε_{it}. In addition, we also work on the estimation of the factor structure, and pay attention to the selection of the number of factors (i.e., d_f). Once disentangling the information of the dataset, we are able to understand whether the following equation holds for $\forall i, j$.

$$m_{ij}(\tau_2^*) - m_{ij}(\tau_1^*) = 0,$$

in which $0 \leq \tau_1^* < \tau_2^* \leq 1$ are two given time points of interest over which we suspect the change has occurred. For example, we may set $\tau_1^* = 0$ and $\tau_2^* = 1$ in order to compare the present and the past. Notably, this is different from a constancy test such as those in Chen & Hong (2012) and Zhang & Wu (2012), but it provides certain flexibility when looking at the dataset spanned over a long time horizon. We acknowledge the vast literature of parameter stability tests, and refer interested readers to Gao & Hawthorne (2006), Chen & Hong (2012), Zhang & Wu (2012), Su & Wang (2017), and among others for extensive studies.

We next review the literature of panel data models with interactive fixed effects. Since the seminal works of Pesaran (2006) and Bai (2009), a variety of panel data models with interactive fixed effects have been proposed and studied. Among them, a large group concerning heterogeneous coefficients relies on the common correlated effects (CCE) approach, e.g., Pesaran & Tosetti (2011), Su & Jin (2012), Boneva & Linton (2017), to name a few. Some common limitations shared by these studies are: (1) a factor structure is always imposed on the regressors; (2) the usual rank limitation inherent in the CCE approach applies; (3) the CCE approach deals with the coefficients of the regressors only while the estimation of the factor structure is always left behind. Recently, another strand of literature starts to extend the principal component analysis (PCA) approach and the maximum likelihood approach of Bai (2009) and Bai & Li (2014) by incorporating heterogeneous coefficients, e.g., Huang et al. (2021), Li et al. (2020), Liu (2020), etc. By doing so, the

[2]Strictly speaking, $m_{i0}(\cdot)$ stands for the global trend plus the trend of the reference group.

unobservable factor structure together with the heterogeneous coefficients can be unveiled simultaneously. Among these studies, however, majority work focuses on parametric models, with two exceptions by Su & Jin (2012) and Liu (2020) that concern heterogeneous functions using sieve and kernel methods, respectively. Although asymptotic distributions are well established for the functions of interest in both papers, how to get some valid inference remains unclear if the unobservable error terms admit weak correlation over both dimensions.

Our contributions to the literature are four-fold. First, we consider a panel data model that allows for heterogeneous time trends at different locations. The newly proposed framework suits our empirical dataset. Second, we establish an asymptotic theory for the proposed estimation method, and develop a nonparametric dependent wild bootstrap (DWB) method to obtain valid inference for the case where weak correlation is present in both dimensions of the error terms. Third, we examine our theoretical findings through extensive simulations. Finally, we use the newly proposed model and method to investigate U.K. rainfall data. We find that overall, the rainfall in Autumn and Winter indeed increases over the past sixty years, however, there is no significant change in Spring and Summer. Notably, for Spring, Summer and Autumn, the majority of the stations show no significant change in rainfall data, which also implies that the increasing rainfall in Autumn is driven by a small number of locations on the west coast of the country.

The structure of this paper is as follows. Section 2 proposes the estimation method and the corresponding asymptotic theory. Specifically, we establish the asymptotic distributions of the proposed estimators in Section 2.1, develop the corresponding theory of the nonparametric DWB method in Section 2.2, and finally consider the estimation of the number of factors in Section 2.3. Section 3 conducts extensive simulation studies to examine the finite–sample .properties of the proposed model and methods. Using the proposed model and method, Section 4 studies the rainfall data of U.K. We conclude in Section 5. The proofs are regulated in the online supplementary appendix.

Before proceeding further, we introduce some notations to facilitate the development. $\to_P$ and $\to_D$ stand for convergence in probability and convergence in distribution respectively; $K(\cdot)$ and h respectively stand for the kernel function and the bandwidth of the nonparametric regression; $K_h(\cdot) = K(\cdot/h)/h$; for a matrix A with full column rank, let $M_A = I - A(A^\top A)^{-1}A^\top$; for a matrix A, let $\|A\|$ and $\|A\|_2$ be the Frobenius norm and the Spectral norm respectively; for two matrices A and B having the same dimensions, let $A \circ B$ stand for the Hadamard product of A and B; for two random variables a and b, $a \asymp b$ stands for $a = O_P(b)$ and $b = O_P(a)$.

2 Estimation Method and Theory

In this section, we first present the estimation method, and establish the corresponding asymptotic properties. Specifically, assuming d_f is known, we derive the asymptotic distributions of the estimators in Section 2.1, and provide the bootstrap inference in Section 2.2. Finally, we consider the estimation of the number of factors in Section 2.3.

Remark 2.1. *Before proceeding further, we comment on an identification issue associated with the heterogeneous coefficients. For simplicity, suppose that f_t is independent and identically distributed (i.i.d.), and can be written as*

$$f_t = f + \xi_t,$$

in which $f = E[f_t]$ and $E[\xi_t] = 0$. In this case, we can rewrite the model (1.4) *as*

$$y_{it} = m_{i0}^*(\tau_t) + \sum_{j=1}^{J} D_{jt} m_{ij}(\tau_t) + \gamma_i^\top \xi_t + \varepsilon_{it},$$

where $m_{i0}^*(\tau_t) = m_{i0}(\tau_t) + f^\top \gamma_i$. It is now clear that one can only identify $m_{i0}(\tau_t)$ up to an unknown constant. Throughout the paper, we assume $E[f_t] = 0$ for simplicity, and refer interested readers to, for example, Linton (1997), Sperlich et al. (2002), Connor et al. (2012) and many references therein for extensive discussions on similar matters.

We now proceed, and write (1.4) in a vector form as follows:

$$Y_i = DM_i + F\gamma_i + \mathcal{E}_i, \tag{2.1}$$

where $D = (I_T, D_1, \ldots, D_J)$, $D_j = \text{diag}\{D_{j1}, \ldots, D_{jT}\}$ for $j \in [J]$, $M_i = (M_{i0}^\top, \ldots, M_{iJ}^\top)^\top$, and $M_{i\ell} = (m_{i\ell}(\tau_1), \ldots, m_{i\ell}(\tau_T))^\top$ for $\ell \in 0 \cup [J]$. In (2.1), the definitions of Y_i, F and $\mathcal{E}_i$ are obvious, they thus are omitted.

To recover $m_{ij}(\cdot)$'s, we briefly introduce the local linear method. The idea is that when τ_t is close to τ,

$$m_{ij}(\tau_t) \approx z_t^\top \widetilde{m}_{ij}(\tau), \tag{2.2}$$

where $z_t = (1, \frac{\tau_t - \tau}{h})^\top$ and $\widetilde{m}_{ij}(\tau) = (m_{ij}(\tau), h \cdot m_{ij}^{(1)}(\tau))^\top$. Thus, (2.1) can be parametrized as follows:

$$K_\tau Y_i \approx K_\tau \mathbb{Z} \widetilde{m}_i(\tau) + K_\tau F \gamma_i + K_\tau \mathcal{E}_i, \tag{2.3}$$

where $K_\tau = \text{diag}\{\sqrt{K_h(\tau_1 - \tau)}, \ldots, \sqrt{K_h(\tau_T - \tau)}\}$, $\widetilde{m}_i(\tau) = (\widetilde{m}_{i0}(\tau)^\top, \ldots, \widetilde{m}_{iJ}(\tau)^\top)^\top$, and $\mathbb{Z} = D(I_{J+1} \otimes (z_1, \ldots, z_T)^\top)$. In addition, let $K(\cdot)$ be a boundary adjusted kernel (Hong &

Li 2005) in order to avoid an extra constant term in the limit when τ is sufficiently close to 0 and 1:

$$K\left(\frac{\tau_t-\tau}{h}\right)=\begin{cases}\mathcal{K}\left(\frac{\tau_t-\tau}{h}\right)/\int_{-\tau/h}^{1}\mathcal{K}(w)dw, & \tau\in[0,h),\\ \mathcal{K}\left(\frac{\tau_t-\tau}{h}\right), & \tau\in[h,1-h],\\ \mathcal{K}\left(\frac{\tau_t-\tau}{h}\right)/\int_{-1}^{(1-\tau)/h}\mathcal{K}(w)dw, & \tau\in(1-h,1],\end{cases} \tag{2.4}$$

where $\mathcal{K}(w)$ is a typical kernel function (say, the Epanechnikov kernel). The use of (2.2) and (2.4) together ensures that the bias of the kernel estimation is the same magnitude everywhere on $[0,1]$ (Connor et al. 2012, Section 3.2).

In view of (2.3), for $\tau\in[0,1]$, we define the following objective function:

$$Q_\tau(\mathbb{B},\mathbb{F})=\sum_{i=1}^{N}(Y_i-\mathbb{Z}\beta_i)^\top K_\tau M_{\mathbb{F}}K_\tau(Y_i-\mathbb{Z}\beta_i), \tag{2.5}$$

where $\mathbb{B}=(\beta_1,\ldots,\beta_N)^\top$ with each $\beta_i=\widetilde{m}_i(\tau)$ being a $2(J+1)\times 1$–dimensional vector, $\mathbb{F}$ is a $T\times d_f$ matrix, and $M_{\mathbb{F}}=I-\mathbb{F}(\mathbb{F}^\top\mathbb{F})^{-1}\mathbb{F}^\top$. A commonly used assumption accompanied the local linear method is that the functions (i.e., $m_{ij}(\cdot)$'s) are twice continuously differentiable on $[0,1]$. Therefore, it is reasonable to assume $\max_i\|\beta_i\|<\infty$ (e.g., Li & Racine 2007). For $\mathbb{F}$, we suppose that $\frac{1}{T}\mathbb{F}^\top\mathbb{F}=I_{d_f}$ for the purpose of identification, which is well adopted in the literature of PCA (e.g., Jolliffe 2012). Therefore, to minimise (2.5), we introduce the following two sets:

$$\mathbb{S}_B=\{\mathbb{B}=(\beta_1,\ldots,\beta_N)^\top\mid\max_i\|\beta_i\|<\infty\}\quad\text{and}\quad\mathbb{S}_F=\left\{\mathbb{F}\mid\frac{1}{T}\mathbb{F}^\top\mathbb{F}=I_{d_f}\right\}. \tag{2.6}$$

With (2.5) and (2.6) in hand, for $\forall\tau\in[0,1]$, the estimators are defined as follows:

$$(\widehat{\mathbb{B}}_\tau,\widehat{\mathbb{F}}_\tau)=\underset{\mathbb{B}\in\mathbb{S}_B,\mathbb{F}\in\mathbb{S}_F}{\operatorname{argmin}}\ Q_\tau(\mathbb{B},\mathbb{F}), \tag{2.7}$$

in which $\widehat{\mathbb{B}}_\tau=(\widehat{\beta}_{\tau,1},\ldots,\widehat{\beta}_{\tau,N})^\top$. Moreover, (2.7) admits the following expressions:

$$\widehat{\beta}_{\tau,i}=\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau Y_i, \tag{2.8}$$

for $i=1,\ldots,N$, and

$$\widehat{\mathbb{F}}_\tau\widehat{V}_\tau=K_\tau\widehat{\Sigma}(\widehat{\mathbb{B}}_\tau)K_\tau\widehat{\mathbb{F}}_\tau, \tag{2.9}$$

where $\widehat{\Sigma}(\mathbb{B})=\frac{1}{NT}\sum_{i=1}^{N}(Y_i-\mathbb{Z}\beta_i)(Y_i-\mathbb{Z}\beta_i)^\top$, and $\widehat{V}_\tau=\text{diag}\{\widehat{\lambda}_{\tau,1},\ldots,\widehat{\lambda}_{\tau,d_f}\}$ is formed by the largest d_f eigenvalues of $K_\tau\widehat{\Sigma}_\tau(\widehat{\mathbb{B}}_\tau)K_\tau$ arranged in descending order. Consequently, for $\forall i\in[N]$ and $\forall j\in 0\cup[J]$, the estimator of $m_{ij}(\tau)$ is defined by

$$\widehat{m}_{ij}(\tau)=\mathscr{S}_{m_j}\widehat{\beta}_{\tau,i}, \tag{2.10}$$

where $\mathscr{S}_{m_j}$'s are the selection matrices defined in an obvious manner.

Up to this point, we have presented the model and the estimation strategy. In what follows, we establish the corresponding asymptotic properties.

2.1 Asymptotic theory

In order to derive the asymptotic properties, we impose the following conditions to facilitate the development.

Assumption 1.

1. *All of* $m_{ij}(\cdot)$*'s are twice continuously differentiable on* $[0,1]$.

2. $\mathcal{K}(\cdot)$ *involved in (2.4) is a positive kernel function, is Lipschitz continuous on* $[-1,1]$, *and* $\int_{-1}^{1}\mathcal{K}(w)dw = 1$. *Also,* $(T\wedge N)h\to\infty$.

3. *(a). Let* $\|\mathcal{E}\|_2 = O_P(\sqrt{N}\vee\sqrt{T})$, *where* $\mathcal{E} = (\mathcal{E}_1,\ldots,\mathcal{E}_N)^\top$.

   (b). $\frac{1}{N}\Gamma^\top\Gamma = \Sigma_\gamma + O_P\left(\frac{1}{\sqrt{N}}\right)$, *where* $\Gamma = (\gamma_1,\ldots,\gamma_N)^\top$. *Also,* $\max_{i\geq 1} E\|\gamma_i\|^4 < \infty$.

   (c). Suppose that $\{f_t\}$ *is a stationary process such that* $E[f_1] = 0$, $E[f_1f_1^\top] = \Sigma_f$, $\sum_{t=1}^{\infty} E|f_1^\top f_{1+t}| = O(1)$, *and* $E\|f_1\|^4 < \infty$.

Assumptions 1.1 and 1.2 are standard conditions, which are widely used in the literature of nonparametric regression (e.g., Linton 1997, Sperlich et al. 2002). Due to the involvement of panel dataset, we have to require both quantities Nh and Th to diverge. Assumption 1.3(a) allows for cross–sectional dependence. The restrictions in Assumption 1.3(b)(c) are commonly used in the relevant literature of factor analysis.

Using Assumption 1, we present the first asymptotic result of this paper below.

Lemma 2.1. *Under Assumption 1, as* $(N,T)\to(\infty,\infty)$,

1. $\sup_{\tau\in[0,1]}\|P_{K_\tau F} - P_{\widehat{\mathbb{F}}_\tau}\| = o_P(1)$,

2. $\sup_{\tau\in[0,1]}\frac{1}{\sqrt{N}}\|\widehat{\mathbb{M}}_j(\tau) - \mathbb{M}_j(\tau)\| = o_P(1)$ *for* $j\in 0\cup[J]$,

3. $\sup_{\tau\in[0,1]}(\|\widehat{V}_\tau\|_2 + \|\widehat{V}_\tau^{-1}\|_2) = O_P(1)$,

where $\mathbb{M}_j(\tau) = (m_{1j}(\tau),\ldots,m_{Nj}(\tau))^\top$ *and* $\widehat{\mathbb{M}}_j(\tau) = (\widehat{m}_{1j}(\tau),\ldots,\widehat{m}_{Nj}(\tau))^\top$.

Lemma 2.1 establishes results on uniform convergence. The first result of Lemma 2.1 indicates that $\widehat{\mathbb{F}}_\tau$ can successfully recover the space spanned by $K_\tau F$, which also has full column rank in view of the fact that

$$\sup_{\tau\in[0,1]}\left\|\frac{1}{T}F^\top K_\tau^2 F - \Sigma_f\right\| = o_P(1).$$

In addition, the above equation infers that the asymptotic eigenvalues associated with $\frac{1}{T}F^\top K_\tau^2 F$ are identical to those yielded by $\frac{1}{T}F^\top F$. From the signal-noise ratio viewpoint, the PCA step (2.9) involved in our nonparametric regression yields a signal which is asymptotically equivalent to the situation when no nonparametric kernel involved. The second result of Lemma 2.1 establishes the consistency for the estimators of the functions, which should be expected. The third result of this lemma basically says that when studying (2.9), the eigenvalues of $K_\tau \widehat{\Sigma}(\widehat{\mathbb{B}}_\tau) K_\tau$ are bounded from both below and above uniformly in $\tau \in [0,1]$. It is worth mentioning that the establishment of Lemma 2.1 only requires very limited restrictions.

To establish asymptotic distributions, we impose the following conditions in Assumption 2.

Assumption 2.

1. *Let $\{\varepsilon_{it}\}$ be independent of $\{f_t\}$ and $\{\gamma_i\}$, and be stationary over t for each i. Moreover, $E[\varepsilon_{i1}] = 0$, $\max_{i\geq 1} \sum_{t=1}^{\infty} E|\varepsilon_{i1}\varepsilon_{i,t+1}| = O(1)$, and $\max_{i\geq 1} E|\varepsilon_{i1}|^\delta < \infty$ for $\delta \geq 4$.*
2. $\sup_{\tau\in[0,1]} \frac{1}{NT} \| \sum_{i=1}^N (\mathbb{Z}, F)^\top K_\tau^2 \mathcal{E}_i \| = O_P\left(\frac{\sqrt{\log(NT)}}{\sqrt{NTh}}\right)$.
3. *For $\forall \tau \in [0,1]$, let $\frac{1}{\sqrt{Th}} \mathbb{Z}^\top K_\tau^2 \mathcal{E}_i \to_D N(0, \Sigma_{2i,\tau})$, where $\Sigma_{2i,\tau} = \lim_T \frac{h}{T} \mathbb{Z}^\top K_\tau^2 E[\mathcal{E}_i \mathcal{E}_i^\top] K_\tau^2 \mathbb{Z}$.*

Assumption 2 requires ε_{it} to be stationary over time for each i, and allows for heteroskedasticity along the cross-sectional dimension. Assumption 2.2 can be simply justified if more conditions are imposed along the cross-sectional dimension of $\{\varepsilon_{it}\}$. We refer interested readers to (B.6) of Chen et al. (2012*b*) for a detailed development, and to discussions for Assumption A.4 of Su & Wang (2017) on a similar matter. Assumption 2.3 requires a central limit theory to hold for each i. It is noteworthy that $\widehat{\mathbb{F}}_\tau$ depends on the estimators of $m_{ij}(\tau)$'s in (2.9). After aggregating the information over i, the variables of the factor structure will completely vanish when establishing an asymptotic distribution for $\sqrt{Th}(\widehat{m}_{ij}(\tau) - m_{ij}(\tau))$. That is why Assumption 2.3 only involves $\mathbb{Z}$ and $\mathcal{E}_i$.

Using Assumptions 1 and 2, we present the following theorem.

Theorem 2.1. *Let Assumptions 1 and 2 hold. If, in addition, $Th^5 \to c \in [0, \infty)$, we have as $(N,T) \to (\infty, \infty)$, for $\forall i \in [N]$, $\forall j \in 0 \cup [J]$, and $\forall \tau \in [0,1]$,*

$$\sqrt{Th}(\widehat{m}_{ij}(\tau) - m_{ij}(\tau) + O_P(h^2)) \to_D N(0, \Sigma_{ij}(\tau)),$$

where $\Sigma_{ij}(\tau) = \mathscr{S}_{m_j} \Sigma_{1,\tau}^{-1} \Sigma_{2i,\tau} \Sigma_{1,\tau}^{-1} \mathscr{S}_{m_j}^\top$, in which $\Sigma_{1,\tau} = \lim_T \frac{1}{T} \mathbb{Z}^\top K_\tau^2 \mathbb{Z}$ and $\mathscr{S}_{m_j}$ is the same as in (2.10).

It is worth mentioning that the term $O_P(h^2)$ is also related to other estimates of $m_{i^*j}(\tau)$ with $i^* \neq i$. It is not hard to see that each individual $\widehat{m}_{ij}(\tau)$ of (2.8) depends on $\widehat{\mathbb{F}}_\tau$, however, as explained above, $\widehat{\mathbb{F}}_\tau$ of (2.9) further depends on all of $\widehat{m}_{ij}(\tau)$'s. As a consequence, it is theoretically challenging to decompose the bias terms for each individual. In addition to solving the bias term issue, we develop a nonparametric DWB method in Section 2.2 below to deal with the infeasibility of $\Sigma_{ij}(\tau)$ as the variance component of Theorem 2.1 for valid inference.

Now, recall that we are interested in the difference between $m_{ij}(\tau_1^*)$ and $m_{ij}(\tau_2^*)$. Thus, consider $\Delta M_{ij}(\tau_2^*, \tau_1^*) = \widehat{m}_{ij}(\tau_2^*) - \widehat{m}_{ij}(\tau_1^*)$.

Corollary 2.1. *Let the conditions of Theorem 2.2 hold. If, in addition, $Th^5 \to 0$, we have as $(N, T) \to (\infty, \infty)$, for $\forall i \in [N]$, $\forall j \in 0 \cup [J]$,*

$$\sqrt{Th}\,[\Delta M_{ij}(\tau_2^*, \tau_1^*) - (m_{ij}(\tau_2^*) - m_{ij}(\tau_1^*))] \to_D N(0, \mathscr{S}_{m_j} \Sigma_{\tau_1^*, \tau_2^*} \mathscr{S}_{m_j}^\top),$$

where $\Sigma_{\tau_1^*, \tau_2^*} = \Sigma_{1,\tau_1^*}^{-1} \Sigma_{2i,\tau_1^*} \Sigma_{1,\tau_1^*}^{-1} + \Sigma_{1,\tau_2^*}^{-1} \Sigma_{2i,\tau_2^*} \Sigma_{1,\tau_2^*}^{-1}$.

The corollary is a direct application of Theorem 2.1. In what follows, we consider how to conduct the hypothesis testing based on a bootstrap procedure.

2.2 Bootstrap Inference

In this subsection, we develop a nonparametric DWB method for our panel data analysis. We briefly review the relevant literature before proceeding. The original DWB method is initially introduced in Shao (2010), wherein a comprehensive comparison between the DWB and some existing bootstrap methods can be found. The moving block bootstrap (MBB) approach for panel data models of Gonçalves (2011) shares a motivation very similar to what to be investigated below. By shuffling the sample along the time dimension randomly, the MBB method preserves the information of the cross-sectional dimension well. However, it destroys the time trends in the bootstrap draws. Therefore, we propose to extend the DWB method developed in Gao et al. (2022) for parametric panel data models for our nonparametric panel data regression model analysis.

By Section 2.1, we are able to obtain $\widehat{m}_{ij}(\tau_t)$ for all (i, j, t), which yields

$$\widehat{u}_{it} = y_{it} - \widehat{m}_{i0}(\tau_t) - \sum_{j=1}^{J} D_{jt} \widehat{m}_{ij}(\tau_t).$$

Let $\widehat{\mathbb{U}} = \{\widehat{u}_{it}\}_{N \times T}$, which serves as an estimate of $\Gamma F^\top + \mathcal{E}$.

We are now ready to introduce the following DWB method to our nonparametric panel data framework:

(i). Let $\xi = (\xi_1, \ldots, \xi_T)^\top$ be an ℓ-dependent time series for each bootstrap draw, and let ξ satisfy that

$$E[\xi_t] = 0, \quad E|\xi_t|^2 = 1, \quad E|\xi_t|^\delta < \infty, \quad E[\xi_t \xi_s] = a\left(\frac{t-s}{\ell}\right), \tag{2.11}$$

where $\ell \to \infty$ and $\frac{\ell}{\sqrt{Th}} \to 0$, δ is the same as that in Assumption 2.1, and $a(\cdot)$ is a symmetric kernel defined on $[-1, 1]$ satisfying that $a(0) = 1$ and $K_a(x) = \int_{\mathbb{R}} a(u) e^{-iux} du \geq 0$ for $x \in \mathbb{R}$.

(ii). For $\forall \tau$, construct a new set of dependent variables by $Y_i^* = \mathbb{Z}\widehat{\beta}_{\tau,i} + \widehat{\mathbb{U}}_i \circ \xi$, where $\circ$ defines the Hadamard product, and $\widehat{\mathbb{U}}_i$ is the i^{th} column of $\widehat{\mathbb{U}}$. Accordingly, the new estimates of $\widehat{m}_{ij}^*(\tau)$'s are obtained using

$$\widehat{\beta}_{\tau,i}^* = \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau Y_i^*. \tag{2.12}$$

(iii). We repeat the above procedure $\mathcal{L}$ times.

In practice, one can always generate ξ from $N(0, \Sigma_\xi)$, where $\Sigma_\xi = \{a\left(\frac{t-s}{\ell}\right)\}_{T\times T}$. Although the normal distribution is not really necessary in theory, it fulfils every aforementioned requirement. The restrictions imposed on $a(\cdot)$ are satisfied by a few commonly used kernels, such as the Bartlett and Parzen kernels. A variety of kernel functions satisfying (2.11) can be found in Andrews (1991) and Shao (2010) for example. In practice, one may simply adopt the Bartlett kernel as in Bai et al. (2020).

The above procedure provides a flexible way to construct the confidence interval, and also handles the correlation arose with $\{\varepsilon_{it}\}$. Formally, we summarize the asymptotic property of the DWB procedure in the following theorem.

Theorem 2.2. *Let the conditions of Theorem 2.1 hold. If, in addition, suppose* (2.11) *holds and $Th^5 \to 0$, we then have as $(N, T) \to (\infty, \infty)$, for $\forall \tau \in [0, 1]$, $\forall i \in [N]$, $\forall j \in 0 \cup [J]$,*

$$\sup_{w\in\mathbb{R}} \left| \Pr\left(\sqrt{Th}(\widehat{m}_{ij}(\tau) - m_{ij}(\tau)) \leq w\right) - \Pr\left(\sqrt{Th}(\widehat{m}_{ij}^*(\tau) - \widehat{m}_{ij}(\tau)) \leq w\right) \right| = o_P(1),$$

where $\widehat{m}_{ij}^*(\tau)$ has been defined in the above bootstrap procedure.

Recall that we are interested in the following test for $\forall i \in [N]$ and $\forall j \in 0 \cup [J]$ at the given time points τ_1^* and τ_2^*:

$$H_0 : m_{ij}(\tau_2^*) - m_{ij}(\tau_1^*) = 0.$$

A simple way of testing this hypothesis is, for each generated ξ, calculating

$$\Delta M_{ij}^*(\tau_2^*, \tau_1^*) = \widehat{m}_{ij}^*(\tau_2^*) - \widehat{m}_{ij}^*(\tau_1^*)$$

using the bootstrap draws. Accordingly, we reject the null at the 5% significance level, if the corresponding confidence interval from the bootstrap procedure does not include 0.

2.3 Selection of the Number of Factors

To close our theoretical investigation, we finally consider the selection of the factor number. To facilitate the development, we define a new set

$$\mathbb{S}_F^* = \left\{ \mathbb{F} \mid \frac{1}{T}\mathbb{F}^\top \mathbb{F} = I_{d_{\max}} \right\},$$

where $d_{\max}$ ($\geq d_f$) is a user-specified large constant. Accordingly, we define

$$(\widetilde{\mathbb{B}}_\tau, \widetilde{\mathbb{F}}_\tau) = \underset{\mathbb{B}\in\mathbb{S}_B, \mathbb{F}\in\mathbb{S}_F^*}{\operatorname{argmin}} \; Q_\tau(\mathbb{B}, \mathbb{F}), \tag{2.13}$$

Although the number of factors is over-specified in (2.13), we still achieve a slow rate of convergence without imposing many assumptions. Specifically, we summarize the conclusion in the following corollary.

Corollary 2.2. *Under Assumption 1, as* $(N, T) \to (\infty, \infty)$,

$$\sup_{\tau\in[0,1]} \frac{1}{\sqrt{N}} \|\widetilde{\mathbb{M}}_j(\tau) - \mathbb{M}_j(\tau)\| = O_P\left(h^2 + \frac{1}{\sqrt{(N \wedge T)h}} \right),$$

in which $\widetilde{\mathbb{M}}_j(\tau)$ *is defined in the same way as* $\widehat{\mathbb{M}}_j(\tau)$ *in Lemma 2.1 but using* $\widetilde{\mathbb{B}}_\tau$.

To select the number of factors, we consider the covariance matrix $K_\tau \widehat{\Sigma}_\tau(\widetilde{\mathbb{B}}_\tau) K_\tau$, where $\widehat{\Sigma}_\tau(\mathbb{B})$ has been defined in (2.9). Further we define a mock eigenvalue $\lambda_{\tau,0} \equiv 1$ in order to cover the case with $d_f = 0$. Finally, we estimate d_f by

$$\widehat{d}_f = \sup_{\tau\in[0,1]} \underset{0\leq \ell \leq d_{\max}}{\operatorname{argmin}} \left\{ \frac{\lambda_{\tau,\ell+1}}{\lambda_{\tau,\ell}} \cdot \mathbb{I}(\lambda_{\tau,\ell} \geq \epsilon_{NT}) + \mathbb{I}(\lambda_{\tau,\ell} < \epsilon_{NT}) \right\}, \tag{2.14}$$

where $\epsilon_{NT} = \{\log(Nh \vee Th)\}^{-1}$, and $\lambda_{\tau,1}, \ldots, \lambda_{\tau,d_{\max}}$ are the largest $d_{\max}$ eigenvalues of $K_\tau \widehat{\Sigma}_\tau(\widetilde{\mathbb{B}}_\tau) K_\tau$ arranged in descending order.

In (2.14), the reason of having indicators function is to bypass a difficulty raised by Lam & Yao (2012), in which the authors point out the challenge of studying the term $\frac{\lambda_{\tau,\ell+1}}{\lambda_{\tau,\ell}}$ for $\ell > d_f$, and conjecture $\frac{\lambda_{\tau,\ell+1}}{\lambda_{\tau,\ell}} \asymp 1$ when $\ell > d_f$. In other words, whether simply using $\frac{\lambda_{\tau,\ell+1}}{\lambda_{\tau,\ell}}$ for $\ell \in [d_{\max}]$ in (2.14) guarantees a U-shape objective function remains unknown in theory. Therefore, to account for feasibility, we introduce the indicator function so that the modified objective function has a U-shape.

Theorem 2.3. *Under Assumption 1, as* $(N, T) \to (\infty, \infty)$, $\Pr(\widehat{d}_f = d_f) \to 1$.

In Section 3, we examine the finite sample performance of the proposed estimation methods.

2.4 Further Discussions

Of interest, another possible test to be conducted is evaluating the overall change across all stations. One way of doing it is to rewrite the model as follows:

$$y_{it} = m_0(\tau_t) + \sum_{j=1}^{J} D_{jt} m_j(\tau_t) + \gamma_i^\top f_t + \varepsilon_{it}, \tag{2.15}$$

in which we no longer have heterogeneous coefficients. We then conduct the hypothesis for each:

$$H_0 : m_j(\tau_2^*) - m_j(\tau_1^*) = 0. \tag{2.16}$$

By doing so, we automatically aggregate the information over both dimensions using the homogeneity design. The development of (2.15) will be similar to the online supplementary Appendix B of Dong et al. (2021), so we omit the details.

In what follows, we take another route by assuming $\forall j \in 0 \cup [J]$

$$m_{ij}(\tau) = m_j(\tau) + \nu_{ij}, \tag{2.17}$$

which is widely adopted in the literature, such as, Pesaran (2006, Eq. (13)), Fan et al. (2016, Eq. (1.3)), Boneva & Linton (2017, Eq. (6)), to name a few. In (2.17), ν_{ij} models the randomness associated with each individual station at the period j. Specifically, we require the following assumption to hold.

Assumption 3. *Suppose that for each j, ν_{ij} is independent and identically distributed over i, and satisfies that $E[\nu_{ij}] = 0$ and $E[\nu_{ij}^2] = \sigma_{\nu,j}^2$.*

Under (3), we test (2.16), and construct the following statistic

$$\Delta M_j(\tau_2^*, \tau_1^*) = \frac{1}{N} \sum_{i=1}^{N} [\widehat{m}_{ij}(\tau_2^*) - \widehat{m}_{ij}(\tau_1^*)]. \tag{2.18}$$

By the development of Theorem 2.1, the following corollary hold immediately.

Corollary 2.3. *Under the conditions of Theorem 2.2, suppose further that $Nh^4 \to 0$ and $\frac{\log(NT)}{Th} \to 0$. As $(N, T) \to (\infty, \infty)$, $\forall j \in 0 \cup [J]$,*

$$\sqrt{N}[\Delta M_j(\tau_2^*, \tau_1^*) - (m_j(\tau_2^*) - m_j(\tau_1^*))] \to_D N(0, 2\sigma_{\nu,j}^2).$$

Given J does not diverge, we can further average $\Delta M_j(\tau_2^*, \tau_1^*)$ over j. The limiting distribution is obvious in view of Lemma A.3. The asymptotic distribution can be obtained by taking average over i using the bootstrap draws of Section 2.2.

Finally, to close the theoretical investigation, we address one practical issue — the unbalanceness of the dataset. In our dataset, the available observations of each station actually start from different time periods, and a few stations do not have observations in very recent years. Therefore, we actually observe

$$y_{it}^* = \begin{cases} y_{it} & t \in [\underline{T}_i, \overline{T}_i] \\ 0 & \text{otherwise} \end{cases}, \tag{2.19}$$

where y_{it} is defined in (1.4), and $\max_i \overline{T}_i = T$. Suppose that $\underline{T}_i/T$ and $\overline{T}_i/T$ converge to $\underline{\tau}_i$ and $\overline{\tau}_i$ respectively. Then it is easy to know that if

$$\frac{\sum_{i=1}^N \underline{T}_i + T - \overline{T}_i}{NT} \simeq h^2, \tag{2.20}$$

the aforementioned results will still be valid by restricting τ on the set $[\underline{\tau}_i, \overline{\tau}_i]$ wherever necessary.

Notably, if we restrict our study on the time period $[\max_i \underline{T}_i, \min_i \overline{T}_i]$, it will be somewhat similar to the analysis on the so-call "Tall" (or "Wide") matrix in Bai & Ng (2021). The difference is that in our dataset, we do not need to rotate the data to generate the "Tall" (or "Wide") matrix, and Bai & Ng (2021) focus on the case when missing happens in a random fashion. In another paper, Jin et al. (2021) also discuss how to use EM algorithm to tackle the issue with random missing, which does not suit our dataset.

3 Simulation

We now examine the finite–sample performance of the proposed estimation and DWB methods of Section 2. Before proceeding further, we comment on the numerical implementation. Throughout the paper, $\mathcal{K}(\cdot)$ is always set as the Epanechnikov kernel. As we require undersmoothing (i.e., $Th^5 \to 0$) in Theorem 2.2, let $h = T^{-1/4}$ for simplicity in what follows. To examine the sensitivity of the numerical procedure, we also consider $h_L = 0.8T^{-1/4}$ and $h_R = 1.2T^{-1/4}$ below. One may also follow Section 5.2 of Connor et al. (2012) or Section 5.2 of Su & Wang (2017) for alternative choices of the bandwidth. In general, we find the results are not sensitive to the bandwidth. For the DWB procedure, we adopt the Bartlett kernel, and the bandwidth ℓ is set as $\lceil 1.75(Th)^{1/3} \rceil$ following Palm et al. (2011). It is worth mentioning that Shao (2010) and Gao et al. (2022) discuss the optimal bandwidth ℓ under different settings in length. In summary, when the Bartlett kernel involved, the optimal bandwidth ℓ is at the order $(Th)^{1/3}$ up to an unknown constant for our nonparametric regression. In addition, Gao et al. (2022) conduct extensive simulations to show the DWB procedure is not sensitive to the choice of ℓ and the function form of $a(\cdot)$, so we will not further extend the following simulation studies along this line of research.

The data generating process is as follows:

$$y_{it} = m_{i0}(\tau_t) + \sum_{j=1}^{J} D_{jt} m_{ij}(\tau_t) + \gamma_i^\top f_t + \varepsilon_{it},$$

where $f_t \sim N(0, I_{d_f})$, $\gamma_i \sim 0.5 + N(0, I_{d_f})$, $D_{jt} = \mathbb{I}(\text{mod}(t, J+1) = j)$, and $\text{mod}(\cdot,\cdot)$ defines the modulo operation. Without loss of generality, we let $d_f = 2$ and $J = 3$, so there are 4 periods in the simulated data. To introduce weak correlation over both dimensions of ε_{it}, we let

$$\mathcal{E}_t = 0.5\mathcal{E}_{t-1} + N(0, \Sigma_\varepsilon),$$

where $\mathcal{E}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$ and $\Sigma_\varepsilon = \{0.2^{|i-j|}\}_{N\times N}$. Finally, for the trending functions, we consider two cases:

Case 1: $m_{ij}(\tau) = \frac{i}{N} + \frac{j}{J+1}$,

Case 2: $m_{ij}(\tau) = \frac{i}{N} + \frac{j}{J+1} + \exp(\tau)$,

where $j \in 0 \cup [J]$. Case 1 is designed to demonstrate the aforementioned procedure works for the model with constant coefficients as well. Also, it helps us examine the size of $H_0 : m_{ij}(\tau_2^*) - m_{ij}(\tau_1^*) = 0$. Case 2 introduces a smooth evolvement for each (i, j) over the entire period, so is designed to reflect the power of rejecting $H_0 : m_{ij}(\tau_2^*) - m_{ij}(\tau_1^*) = 0$.

For each generated dataset, we implement the following procedure:

1. Assuming d_f is known, we conduct the estimation on the unknown functions for the points $\tau \in \{0, 0.1, 0.2, \ldots, 1\}$. For each point, we also conduct 199 bootstrap replications in order to obtain a 95% confidence interval. Then we can evaluate the results of Sections 2.1 and 2.2.

2. Assuming d_f is unknown, we implement the procedure of Section 2.3 to estiamte d_f in order to evaluate the result of Section 2.3.

We repeat the above procedure 500 times[3].

To evaluate the performance, we introduce the following measures. First, we examine the estimates of the unknown functions, and define the following root mean squared errors for $\forall j \in 0 \cup [J]$:

[3]One certainly can increase the number of bootstrap replications and the number of Monte Carlo simulation replications. However, we have to bear with the amount of time cost due to the availability of the computational power. The current settings are sufficient to examine the theoretical findings of Section 2.

$$\text{RMSE}_j = \left\{ \frac{1}{500} \sum_{r=1}^{500} \frac{1}{11 \cdot N} \sum_{i=1}^{N} \sum_{s=0}^{10} (\widehat{m}_{ij}^{(r)}(\tau_s) - m_{ij}(\tau_s))^2 \right\}^{1/2},$$

where $j \in 0 \cup [J]$, and $\widehat{m}_{ij}^{(r)}(\tau_s)$ stands for the estimate of $m_{ij}(\tau_s)$ at the r^{th} simulation replication. Second, we report the coverage rate of the bootstrap procedure:

$$\text{CR}_j = \frac{1}{500} \sum_{r=1}^{500} \frac{1}{11 \cdot N} \sum_{i=1}^{N} \sum_{s=0}^{10} \mathbb{I}(\widehat{m}_{ij}^{(r)}(\tau_s) - m_{ij}(\tau_s) \in \text{CI}_{ijs}^{(r)}),$$

where $j \in 0 \cup [J]$, and $\text{CI}_{ijs}^{(r)}$ stands for the 95% confidence interval yielded by the bootstrap procedure for the quantity $\widehat{m}_{ij}^{*,(r)}(\tau_s) - \widehat{m}_{ij}^{(r)}(\tau_s)$ at the r^{th} simulation replication. Finally, to examine the estimate of the number of factors, we let

$$\text{NF} = \frac{1}{500} \sum_{r=1}^{500} \mathbb{I}(\widehat{d}_f^{(r)} = d_f),$$

where $\widehat{d}_f^{(r)}$ is the estimate of d_f at the r^{th} simulation replication.

We summarize the results of RMSE and CR in Table 1. A few facts emerge. First, we comment on RMSE_j's. As shown in both panels of Table 1, the RMSE_j's decrease as T goes up. The pattern is pretty consistent over j. It is not supervising that RMSE_0 is always smaller than RMSE_j with $j \geq 1$, as the available observations of each periodic trend are only a fraction of the entire dataset. The value of N has no impacts on RMSE_j's, which is also consistent with the theoretical findings and the construction of RMSE_j's. Overall, the results of RMSE_j's are not sensitive to the choice of bandwidth.

Second, we comment on the coverage rate. Although the simulated data have weak cross-sectional dependence and time series autocorrelation due to the DGP of ε_{it}, in both tables, the values of CR_j's are approaching the nominal rate (i.e., 95%) as T increases. Notably, the proposed bootstrap method tends to generate a narrower confidence interval when the value T is relative small. That is why the values of CR_j's are always lower than the nominal rate (i.e., 95%). The results are fairly acceptable in view of the facts that a) for nonparametric regression the effective sample size is at order Th, and b) the proposed bootstrap methods needs to further truncate the effective sample using a bandwidth ℓ. Finally, the results are very similar for both Cases 1 and 2, so it infers a good size and power for the hypothesis test $H_0 : m_{ij}(\tau_2^*) - m_{ij}(\tau_1^*) = 0$ in practice.

Finally, we comment on the results of the estimation on the number of factors. In Table 2, the results are accurate overall. Two exceptions are the cases when $(N, T) = (50, 100)$. It seems to suggest that the proposed method is less sensitive to the choice of bandwidth when the sample size is relatively large. Traditionally, when estimating the number of factors, one may start from $T = 25$ (e.g., Lam & Yao 2012), so the value of NF can be much

Table 1: Results of RMSE and CR for Cases 1 & 2

			Case 1						Case 2					
			RMSE			CR			RMSE			CR		
		j	h_L	h	h_R	h_L	h	h_R	h_L	h	h_R	h_L	h	h_R
$N=50$	$T=100$	0	0.958	0.880	0.817	0.803	0.822	0.837	0.959	0.876	0.813	0.809	0.830	0.843
		1	1.211	1.106	1.029	0.840	0.858	0.869	1.225	1.111	1.031	0.843	0.862	0.874
		2	1.244	1.126	1.046	0.841	0.859	0.872	1.221	1.110	1.036	0.848	0.865	0.877
		3	1.185	1.086	1.010	0.842	0.859	0.871	1.198	1.090	1.014	0.844	0.862	0.874
	$T=200$	0	0.719	0.652	0.607	0.858	0.871	0.881	0.727	0.660	0.613	0.862	0.874	0.882
		1	0.916	0.829	0.769	0.883	0.893	0.902	0.913	0.831	0.771	0.884	0.895	0.902
		2	0.932	0.845	0.789	0.886	0.897	0.904	0.939	0.849	0.786	0.891	0.902	0.909
		3	0.910	0.825	0.763	0.887	0.899	0.907	0.905	0.821	0.762	0.889	0.900	0.907
	$T=400$	0	0.551	0.499	0.463	0.894	0.902	0.903	0.547	0.496	0.460	0.899	0.906	0.908
		1	0.691	0.625	0.577	0.914	0.921	0.920	0.684	0.618	0.573	0.920	0.926	0.926
		2	0.719	0.652	0.604	0.911	0.918	0.917	0.706	0.640	0.593	0.916	0.923	0.922
		3	0.687	0.623	0.579	0.916	0.922	0.920	0.678	0.615	0.571	0.916	0.923	0.922
	$T=600$	0	0.472	0.425	0.394	0.903	0.910	0.912	0.471	0.426	0.394	0.909	0.915	0.917
		1	0.599	0.540	0.499	0.916	0.922	0.922	0.590	0.535	0.496	0.922	0.928	0.928
		2	0.609	0.550	0.510	0.921	0.926	0.926	0.604	0.543	0.502	0.924	0.929	0.929
		3	0.585	0.527	0.488	0.923	0.928	0.928	0.591	0.538	0.497	0.921	0.926	0.926
$N=100$	$T=100$	0	0.976	0.889	0.824	0.799	0.821	0.835	0.964	0.878	0.811	0.812	0.833	0.846
		1	1.241	1.128	1.049	0.837	0.854	0.866	1.213	1.104	1.026	0.845	0.862	0.873
		2	1.241	1.127	1.052	0.838	0.856	0.868	1.243	1.128	1.042	0.844	0.863	0.876
		3	1.224	1.113	1.036	0.838	0.855	0.868	1.213	1.105	1.021	0.844	0.861	0.872
	$T=200$	0	0.720	0.651	0.605	0.853	0.868	0.878	0.734	0.668	0.622	0.861	0.874	0.882
		1	0.925	0.838	0.777	0.875	0.887	0.896	0.940	0.854	0.796	0.883	0.894	0.900
		2	0.944	0.855	0.795	0.877	0.889	0.897	0.955	0.865	0.807	0.884	0.894	0.900
		3	0.892	0.810	0.754	0.878	0.890	0.899	0.915	0.834	0.775	0.887	0.897	0.904
	$T=400$	0	0.556	0.502	0.464	0.893	0.903	0.908	0.549	0.497	0.461	0.897	0.906	0.907
		1	0.696	0.628	0.580	0.910	0.918	0.922	0.696	0.626	0.579	0.916	0.923	0.921
		2	0.714	0.645	0.597	0.913	0.920	0.924	0.711	0.641	0.594	0.915	0.922	0.921
		3	0.695	0.628	0.584	0.913	0.920	0.922	0.680	0.613	0.569	0.917	0.924	0.924
	$T=600$	0	0.467	0.424	0.393	0.909	0.913	0.916	0.472	0.428	0.396	0.907	0.912	0.914
		1	0.596	0.539	0.500	0.922	0.924	0.926	0.595	0.537	0.495	0.921	0.926	0.927
		2	0.608	0.550	0.510	0.926	0.928	0.929	0.602	0.545	0.505	0.922	0.927	0.927
		3	0.588	0.531	0.490	0.927	0.928	0.930	0.584	0.526	0.487	0.921	0.927	0.927
$N=200$	$T=100$	0	0.977	0.890	0.822	0.803	0.824	0.838	0.968	0.888	0.819	0.805	0.825	0.839
		1	1.233	1.117	1.037	0.838	0.855	0.866	1.220	1.111	1.028	0.845	0.862	0.875
		2	1.264	1.147	1.062	0.836	0.854	0.867	1.235	1.128	1.050	0.842	0.859	0.870
		3	1.230	1.115	1.028	0.840	0.859	0.871	1.215	1.111	1.028	0.842	0.860	0.871
	$T=200$	0	0.729	0.664	0.618	0.851	0.865	0.875	0.718	0.654	0.610	0.866	0.879	0.887
		1	0.916	0.835	0.775	0.874	0.885	0.894	0.919	0.832	0.773	0.885	0.897	0.903
		2	0.948	0.862	0.802	0.874	0.886	0.895	0.938	0.854	0.796	0.889	0.899	0.906
		3	0.909	0.826	0.765	0.878	0.890	0.897	0.902	0.822	0.764	0.889	0.900	0.907
	$T=400$	0	0.542	0.490	0.454	0.894	0.903	0.907	0.555	0.504	0.469	0.896	0.903	0.904
		1	0.676	0.608	0.563	0.914	0.921	0.924	0.700	0.634	0.588	0.913	0.920	0.918
		2	0.703	0.635	0.589	0.913	0.919	0.922	0.708	0.644	0.598	0.916	0.922	0.920
		3	0.684	0.618	0.573	0.912	0.919	0.922	0.688	0.625	0.582	0.915	0.921	0.919
	$T=600$	0	0.470	0.425	0.393	0.907	0.912	0.914	0.470	0.425	0.393	0.910	0.915	0.917
		1	0.586	0.529	0.489	0.924	0.926	0.928	0.586	0.529	0.489	0.927	0.929	0.930
		2	0.602	0.546	0.504	0.927	0.929	0.929	0.602	0.546	0.504	0.928	0.930	0.931
		3	0.585	0.528	0.488	0.926	0.929	0.930	0.585	0.528	0.488	0.927	0.929	0.931

lower than 1 for small T. Due to the nonparametric nature of the proposed framework, we do not explore the cases with T less than 100, which is also consistent the simulation setup of Su & Wang (2017). As discussed under Lemma 2.1, from the signal-noise ratio viewpoint, the PCA step (2.9) involved in our nonparametric regression yields a signal which is asymptotically equivalent to the situation when no nonparametric kernel involved. That is why Table 2 is filled with 1's when we estimate the number of factors.

Table 2: Results of NF for Case 1 and Case 2

		Case 1			Case 2		
		h_L	h	h_R	h_L	h	h_R
$N = 50$	$T = 100$	1	1	1	1	0.98	0.98
	$T = 200$	1	1	1	1	1	1
	$T = 400$	1	1	1	1	1	1
	$T = 600$	1	1	1	1	1	1
$N = 100$	$T = 100$	1	1	1	1	1	1
	$T = 200$	1	1	1	1	1	1
	$T = 400$	1	1	1	1	1	1
	$T = 600$	1	1	1	1	1	1
$N = 200$	$T = 100$	1	1	1	1	1	1
	$T = 200$	1	1	1	1	1	1
	$T = 400$	1	1	1	1	1	1
	$T = 600$	1	1	1	1	1	1

4 An Empirical Study

In this section, we demonstrate the usefulness of the newly proposed framework using rainfall, temperature and sunshine data of U.K. respectively. Most climate studies are concerned with temperature and conclude that there has been an increasing trend in global and regional temperature. Rather less has been written about precipitation, but the U.K. Meteorological Office recently writes (UKMet 2021)

> *Are we experiencing more heavy rainfall and flooding events ? Several indicators in the latest U.K. State of the Climate report show that the U.K.'s climate is becoming wetter. For example the highest rainfall totals over a five day period are 4% higher during the most recent decade (2008-2017) compared to 1961-1990. Furthermore, the amount of rain from extremely wet days has increased by 17% when comparing the same time periods. In addition, there is a slight increase in the longest sequence of consecutive wet days for the U.K. The change in rainfall depends on your location — for example, changes are largest for Scotland and not significant for most southern and eastern areas of England.*

It is our purpose to investigate whether these claims are satisfied. The data collected from U.K. Meteorological Office include 37 stations (i.e., $N = 37$). Figure 1 presents the geographic locations of these stations. Clearly, they are widely spread out across U.K.

4.1 Data Cleaning

The monthly data cover the period from January of 1853 to July of 2021. In the original dataset, temperature is recorded as min and max temperature for each month, so we take a simple average as the final observation. The units for rainfall, temperature and sunshine are mm, celsius, and hours respectively.

Figure 1: The Geographic Locations of 37 Stations

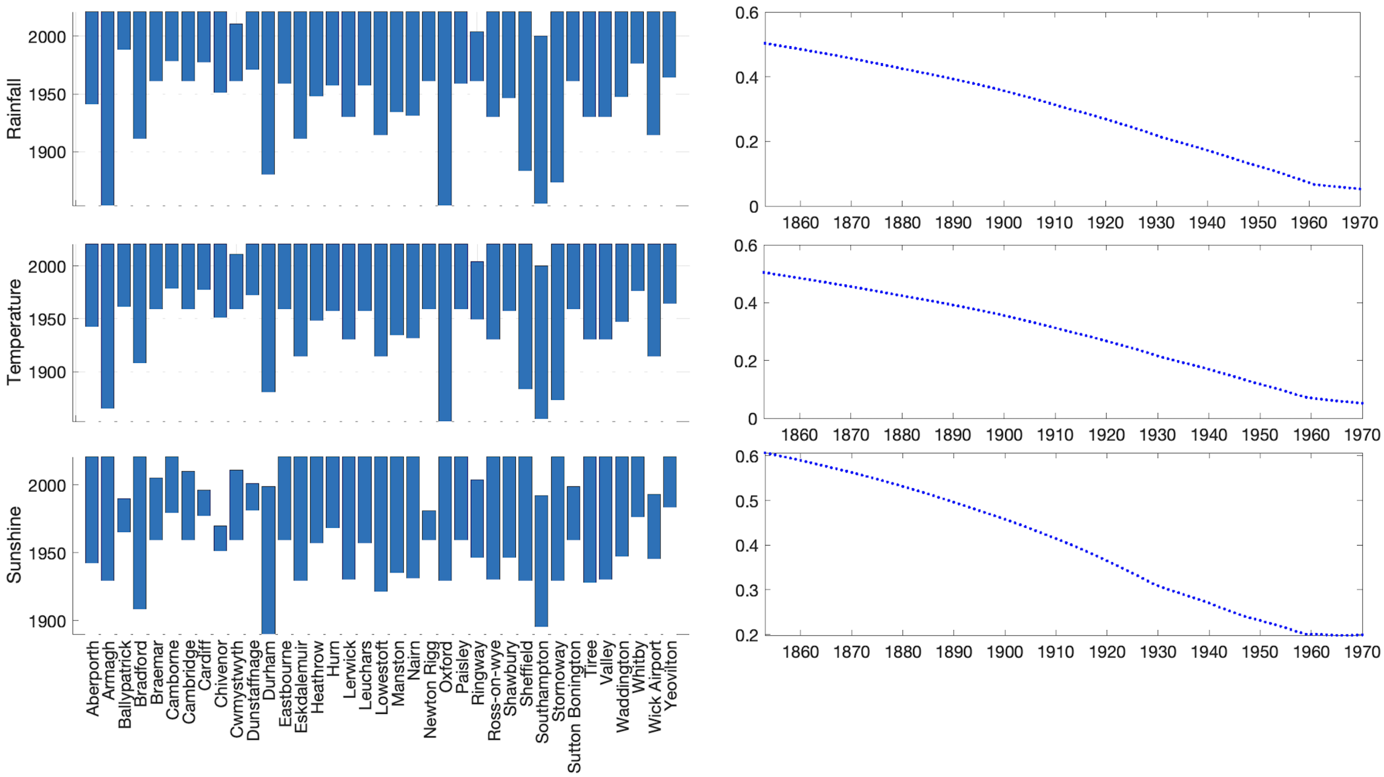
Although the first record from Oxford dates back to January of 1853, for majority of stations, data are not available until early 1900s. Therefore, we have to throw away some information to ensure the unbalanceness does not create serious biases for our analysis. We pick a year within the range of 1853-1970 as the initial year of our analysis. Simple calculation shows that $T = 2023$ for the period from January of 1853 to July of 2021, while $T = 619$ for the period from January of 1970 to July of 2021. In view of our simulation study, a sample size within the range is large enough to ensure reliable results, if missing values are only a small proportion as mentioned in Section 2.4. To better present the data availability, we plot Figure 2. In what follows, we consider two cases[4] below using 1950 and 1955 as the initial year respectively. Therefore, $T = 859$ and 799.

Once we set the starting year, we standardize each time series (for the period with observations). Notably, the standardization does not alter the properties of the time series fundamentally, and also ensures the numbers used for regression are unit free. Such a procedure is commonly adopted in the literature (e.g., Stock & Watson 2005, Fan et al. 2013). For the purpose of demonstration, we plot the original data (using 1950 as the initial year) and the standardized data in Figure 3 to show the range of observations before and after standardization.

Note that the U.K. Meteorological Office defines the four seasons as follows: Spring includes March, April, May; Summer includes June, July, August; Autumn includes Septem-

[4]The choices of 1950 and 1955 are arbitrary. Certainly, other options can be further explored to show the robustness of the results. Due to space limit, we no longer explore these choices further.

Figure 2: Data availability. The left panel plots the years with observations for each station, and the right panel plots the percentages of missing values using different initial year.



ber, October, November; and Winter includes December, January, February. Therefore, we let $J = 3$ below in our analysis. By doing so, we are not only able to investigate the quarterly effects, but also can look into the monthly effects[5].

4.2 Empirical Results

The first result that we note is the number of factors is always one, although we explore two different initial years and three different datasets. We will not mention it again below.

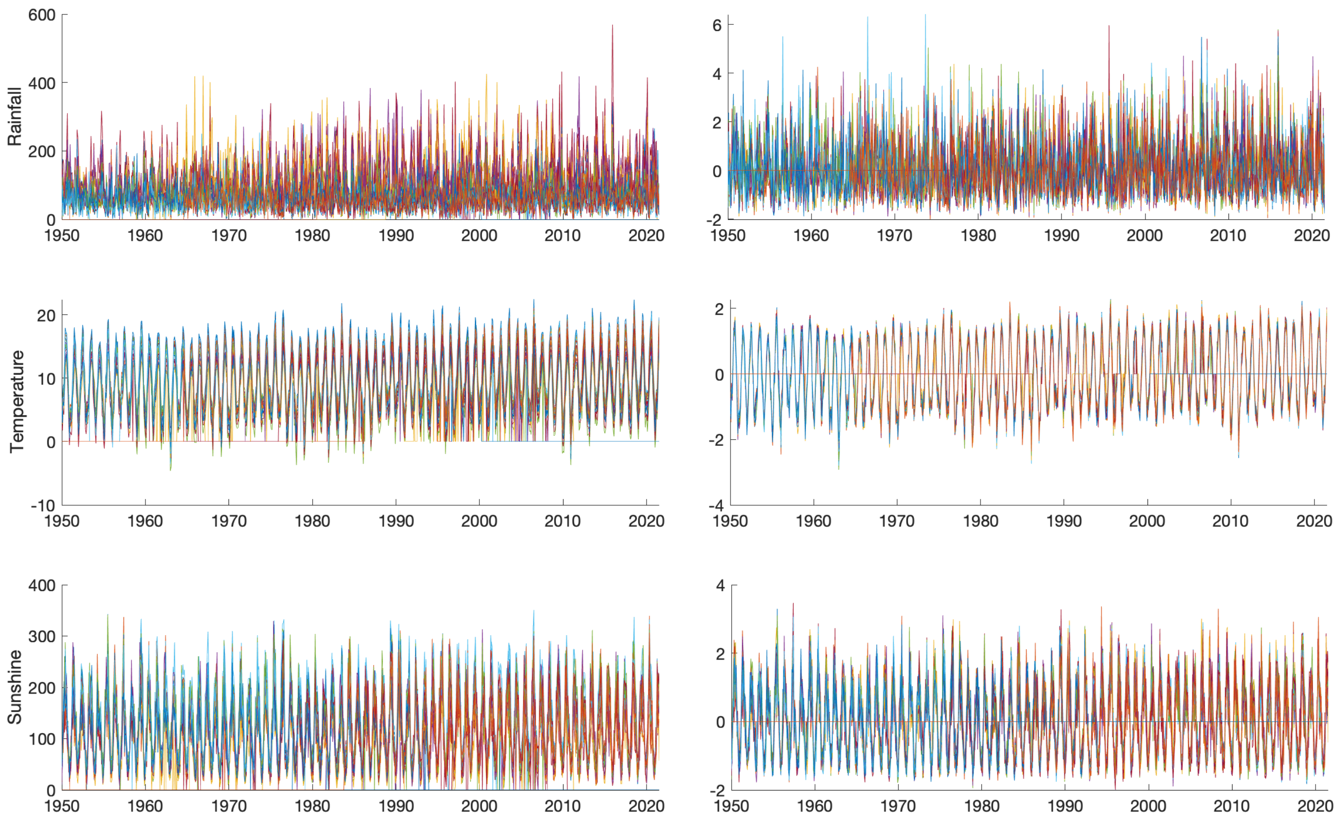
To quantify the changes over the past half a century in each climate dataset, we specifically test the differences[6] between Mar-1970 to Feb-1971 and Mar-2020 to Feb-2021 for each station at each season (or month) using Corollary 2.1, and also test the averaged change using Corollary 2.3. The results are summarized in Table 3-5 below.

Overall, we do find that the temperature has gone up as indicated by majority of

[5]One may also let $J = 11$ in order to simply investigate the monthly effects. Due to space limit, we do not explore this option. Also, by increasing the value of J, the number of observations included in each period will decrease. As a result, the results may become less significant.

[6]We use the estimates of recent year to minus the estimates of early year. The tests are all one-tail tests and carried on at the 5% significance level.

Figure 3: Original and standardized data. The left panel plots the original data, and the right panel plots the standardised data.



previous studies. The only difference is that the temperature of September and November does not have significant change over the past fifty years. For the rainfall data, Winter indeed becomes wetter, however, there is no significant change in the other seasons. When looking at the sunshine data, U.K. actually gets more sunshine during spring and winter, but receives less sunshine over the summer. To sum up, it seems that winter has changed significantly over the past half a century. For the other seasons, changes may vary with respect to locations. For the sake of space, we plot the estimated trends in the online supplementary appendix.

5 Conclusion

In this paper, we have considered a panel data model which allows for heterogeneous time trends at different locations. The framework suits the climate data well. Accordingly, we have established the asymptotic theory of the proposed estimation method. We have also

Table 3: Differences from Rainfall data. The third, fourth, seventh, and eighth columns report the percentages of stations getting drier and wetter over the past fifty years.

Initial Year	Season	Drier	Wetter	Overall		Drier	Wetter	Overall
1950	Spring	0.2432	0.1892		Mar	0.1081	0.0811	
					Apr	0.1892	0.1081	
					May	0.1081	0.0811	
	Summer	0.0541	0.2703		Jun	0.0270	0.0541	
					Jul	0.0000	0.0541	
					Aug	0.0270	0.0541	
	Autumn	0.1081	0.2162		Sep	0.1081	0.0541	
					Oct	0.1081	0.0541	
					Nov	0.1081	0.0000	
	Winter	0.0811	0.7027	Wetter	Dec	0.0811	0.4324	Wetter
					Jan	0.0541	0.4324	Wetter
					Feb	0.0811	0.4324	Wetter
1955	Spring	0.2432	0.1892		Mar	0.1351	0.0811	
					Apr	0.1351	0.0811	
					May	0.1081	0.0811	
	Summer	0.0541	0.2703		Jun	0.0000	0.0541	
					Jul	0.0000	0.0541	
					Aug	0.0000	0.0541	
	Autumn	0.1081	0.2432		Sep	0.1081	0.0541	
					Oct	0.1081	0.0541	
					Nov	0.1081	0.0000	
	Winter	0.0811	0.7297	Wetter	Dec	0.0811	0.3784	Wetter
					Jan	0.0811	0.4324	Wetter
					Feb	0.0811	0.4865	Wetter

developed the corresponding theory for the proposed nonparametric DWB method for valid inference in the case where weak correlation presents in both dimensions of the error terms. We have examined the finite–sample properties of the proposed methods through extensive simulated studies. Finally, we have examined the applicability of the model and methods in the investigation of rainfall, temperature and sunshine data of U.K. Overall, we have found the weather of winter has changed dramatically over the past fifty years. For the other seasons, changes may vary with respect to locations.

References

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Table 4: Differences from Temperature data. The third, fourth, seventh, and eighth columns report the percentages of stations getting colder and warmer over the past fifty years.

Initial Year	Season	Colder	Warmer	Overall		Colder	Warmer	Overall
1950	Spring	0.0541	0.8649	Warmer	Mar	0.0541	0.8378	Warmer
					Apr	0.0541	0.8378	Warmer
					May	0.0541	0.7568	Warmer
	Summer	0.0811	0.9189	Warmer	Jun	0.0811	0.9189	Warmer
					Jul	0.0811	0.9189	Warmer
					Aug	0.0811	0.9189	Warmer
	Autumn	0.0811	0.7568	Warmer	Sep	0.0811	0.4595	
					Oct	0.0811	0.5676	Warmer
					Nov	0.0811	0.4324	
	Winter	0.1351	0.8649	Warmer	Dec	0.1351	0.8108	Warmer
					Jan	0.1351	0.8378	Warmer
					Feb	0.1351	0.7838	Warmer
1955	Spring	0.0541	0.8649	Warmer	Mar	0.0541	0.8649	Warmer
					Apr	0.0541	0.8378	Warmer
					May	0.0541	0.8649	Warmer
	Summer	0.0811	0.9189	Warmer	Jun	0.0811	0.9189	Warmer
					Jul	0.0811	0.9189	Warmer
					Aug	0.0811	0.9189	Warmer
	Autumn	0.0811	0.7297	Warmer	Sep	0.0811	0.4595	
					Oct	0.0811	0.5135	Warmer
					Nov	0.0811	0.4324	
	Winter	0.1351	0.8649	Warmer	Dec	0.1351	0.8378	Warmer
					Jan	0.1351	0.8108	Warmer
					Feb	0.1351	0.7297	Warmer

Bai, J., Choi, S. H. & Liao, Y. (2020), 'Standard errors for panel data models with unknown clusters', *Journal of Econometrics* p. forthcoming.

Bai, J. & Li, K. (2014), 'Theory and methods of panel data models with interactive effects', *Annals of Statistics* **42**(1), 142–170.

Bai, J. & Ng, S. (2021), 'Matrix completion, counterfactuals, and factor analysis of missing data', *Journal of the American Statistical Association* **0**(0), 1–18.

Boneva, L. & Linton, O. (2017), 'A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance', *Journal of Applied Econometrics* **32**(7), 1226–1243.

Chen, B. & Hong, Y. (2012), 'Testing for smooth structural changes in time series models via nonparametric regression', *Econometrica* **80**(3), 1157–1183.

Chen, J., Gao, J. & Li, D. (2012*a*), 'A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models', *Econometric Theory* **28**(5), 1144–1163.

Table 5: Differences from Sunshine data. The third, fourth, seventh, and eighth columns report the percentages of stations getting less and more sunshine over the past fifty years.

Initial Year	Season	Less	More	Overall		Less	More	Overall
1950	Spring	0.3243	0.5676	More	Mar	0.3243	0.3784	
					Apr	0.3243	0.3784	
					May	0.3243	0.3784	
	Summer	0.5135	0.2162	Less	Jun	0.3784	0.1892	Less
					Jul	0.4054	0.1892	Less
					Aug	0.3784	0.1622	Less
	Autumn	0.2703	0.4595		Sep	0.1892	0.3784	
					Oct	0.1892	0.3514	
					Nov	0.1892	0.3784	
	Winter	0.2162	0.6757	More	Dec	0.1622	0.5405	More
					Jan	0.1622	0.5405	More
					Feb	0.1622	0.5405	More
1955	Spring	0.3243	0.5405	More	Mar	0.3243	0.3784	
					Apr	0.3243	0.3784	
					May	0.3243	0.3784	
	Summer	0.5135	0.2162	Less	Jun	0.3784	0.1622	Less
					Jul	0.4054	0.1892	Less
					Aug	0.3784	0.1892	Less
	Autumn	0.2703	0.4324		Sep	0.1892	0.3514	
					Oct	0.1892	0.3243	
					Nov	0.1892	0.3243	
	Winter	0.1892	0.6216	More	Dec	0.1622	0.5405	More
					Jan	0.1622	0.5405	More
					Feb	0.1622	0.5405	More

Chen, J., Gao, J. & Li, D. (2012*b*), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Connor, G., Hagmann, M. & Linton, O. (2012), 'Efficient semiparametric estimation of the fama-french model and extensions', *Econometrica* **80**(2), 713–754.

Dong, C., Gao, J. & Peng, B. (2021), 'Varying-coefficient panel data models with nonstationarity and partially observed factor structure', *Journal of Business & Economic Statistics* **39**(3), 700–711.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society: Series B* **75**(4), 603–680.

Fan, J., Liao, Y. & Wang, W. (2016), 'Projected principal component analysis in factor models', *Annals of Statistics* **44**(1), 219–254.

Gao, J. & Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *The Econometrics Journal* **9**(2), 332–355.

Gao, J., Peng, B. & Yan, Y. (2022), A simple bootstrap method for panel data inferences. https://ssrn.com/abstract=4098007.

Golub, G. H. & Van Loan, C. F. (2013), *Matrix Computations (4th Edition)*, The Johns Hopkins University Press.

Gonçalves, S. (2011), 'The moving blocks bootstrap for panel linear regression models with individual fixed effects', *Econometric Theory* **27**(5), 1048–1082.

Hansen, B. E. (1991), 'Strong laws for dependent heterogeneous processes', *Econometric Theory* **7**(2), 213–221.

Hillebrand, E., Pretis, F. & Proietti, T. (2020), 'Econometric models of climate change: Introduction by the guest editors', *Journal of Econometrics* **214**(1), 1–5. Annals Issue: Econometric Models of Climate Change.

Hong, Y. & Li, H. (2005), 'Nonparametric specification testing for continuous-time models with applications to term structure of interest rates', *The Review of Financial Studies* **18**(1), 37–84.

Huang, W., Jin, S., Phillips, P. C. & Su, L. (2021), 'Nonstationary panel models with latent group structures and cross-section dependence', *Journal of Econometrics* **221**(1), 198–222.

Jin, S., Miao, K. & Su, L. (2021), 'On factor models with random missing: Em estimation, inference, and cross validation', *Journal of Econometrics* **222**(1), 745–777.

Jolliffe, I. T. (2012), *Principal Component Analysis*, second edn, Springer.

Kaufmann, R. K., Mann, M. L., Gopal, S., Liederman, J. A., Howe, P. D., Pretis, F., Tang, X. & Gilmore, M. (2017), 'Spatial heterogeneity of climate change as an experiential basis for skepticism', *Proceedings of the National Academy of Sciences* **114**(1), 67–71.

Lam, C. & Yao, Q. (2012), 'Factor modeling for high-dimensional time series: Inference for the number of factors', *Annals of Statistics* **40**(2), 694–726.

Li, K., Cui, G. & Lu, L. (2020), 'Efficient estimation of heterogeneous coefficients in panel data models with common shocks', *Journal of Econometrics* **216**(2), 327–353.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Linton, O. B. (1997), 'Efficient estimation of additive nonparametric regression models', *Biometrika* **84**(2), 469–473.

Liu, F. (2020), Nonparametric time-varying panel data models with heterogeneity. `https://ssrn.com/abstract=3743529`.

Magnus, J. R. & Neudecker, H. (2007), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, third edn, John Wiley & Sons Ltd.

Palm, F., Smeekes, S. & Urbain, J.-P. (2011), 'Cross-sectional dependence robust block bootstrap panel unit root tests', *Journal of Econometrics* **163**(1), 85–104.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Pesaran, M. H. & Tosetti, E. (2011), 'Large panels with common factors and spatial correlation', *Journal of Econometrics* **161**(2), 182–202.

Phillips, P. C., Leirvik, T. & Storelvmo, T. (2020), 'Econometric estimates of earth's transient climate sensitivity', *Journal of Econometrics* **214**(1), 6–32. Annals Issue: Econometric Models of Climate Change.

Shao, X. (2010), 'The dependent wild bootstrap', *Journal of the American Statistical Association* **105**(489), 218–235.

Sperlich, S., Tjøstheim, D. & Yang, L. (2002), 'Nonparametric estimation and testing of interaction in additive models', *Econometric Theory* **18**(2), 197–251.

Stock, J. H. & Watson, M. W. (2005), Implications of dynamic factor models for VAR analysis. Working Paper 11467. National Bureau of Economic Research, Cambridge.

Su, L. & Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* **169**(1), 34–47.

Su, L. & Wang, X. (2017), 'On time-varying factor models: Estimation and testing', *Journal of Econometrics* **198**(1), 84–101.

UKMet (2021), Uk extreme events – heavy rainfall and floods. Available at `https://www.metoffice.gov.uk/research/climate/understanding-climate/uk-extreme-events-\_heavy-rainfall-and-floods`.

Vogelsang, T. J. & Franses, P. H. (2005), 'Testing for common deterministic trend slopes', *Journal of Econometrics* **126**(1), 1–24.

Zhang, T. & Wu, W. B. (2012), 'Inference of time-varying regression models', *Annals of Statistics* **40**(3), 1376–1402.

Online Supplementary Appendix to "A Nonparametric Panel Model for Climate Data with Seasonal and Spatial Variation"

In this appendix, Appendix A.1 presents the omitted plots of the main text; Appendix A.2 states some preliminary lemmas; Appendix A.3 includes the proofs of the theoretical results. In what follows, $O(1)$ always stands for a constant, and may be different at each appearance.

A.1 Omitted Plots

In this section, we plot the trends[1] using 1950 as the initial year for our analysis. Specifically, Figures A.1-A.12 plot the trends associated with each station for four seasons using three datasets (i.e., rainfall, temperature and sunshine), and also present the averaged trend[2], which is presented at the bottom right in each figure.

[1]When conducting estimation, we use Winter as the reference group. After estimation, we sum up the global trend with each seasonal trend to recover the real trend of each season.

[2]The average is taken across stations.

Figure A.1: The trends of spring for rainfall data

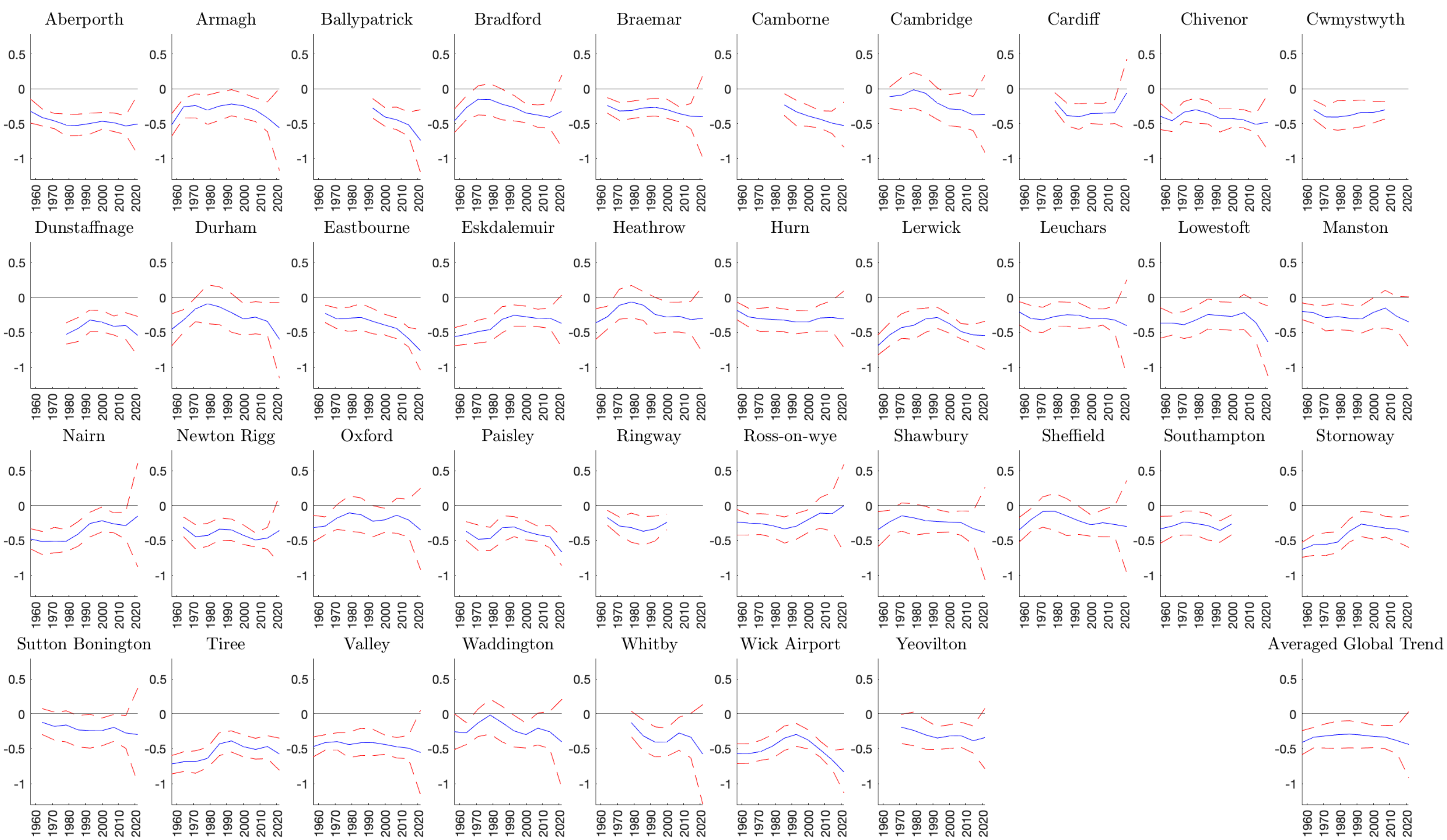


Figure A.2: The trends of summer for rainfall data

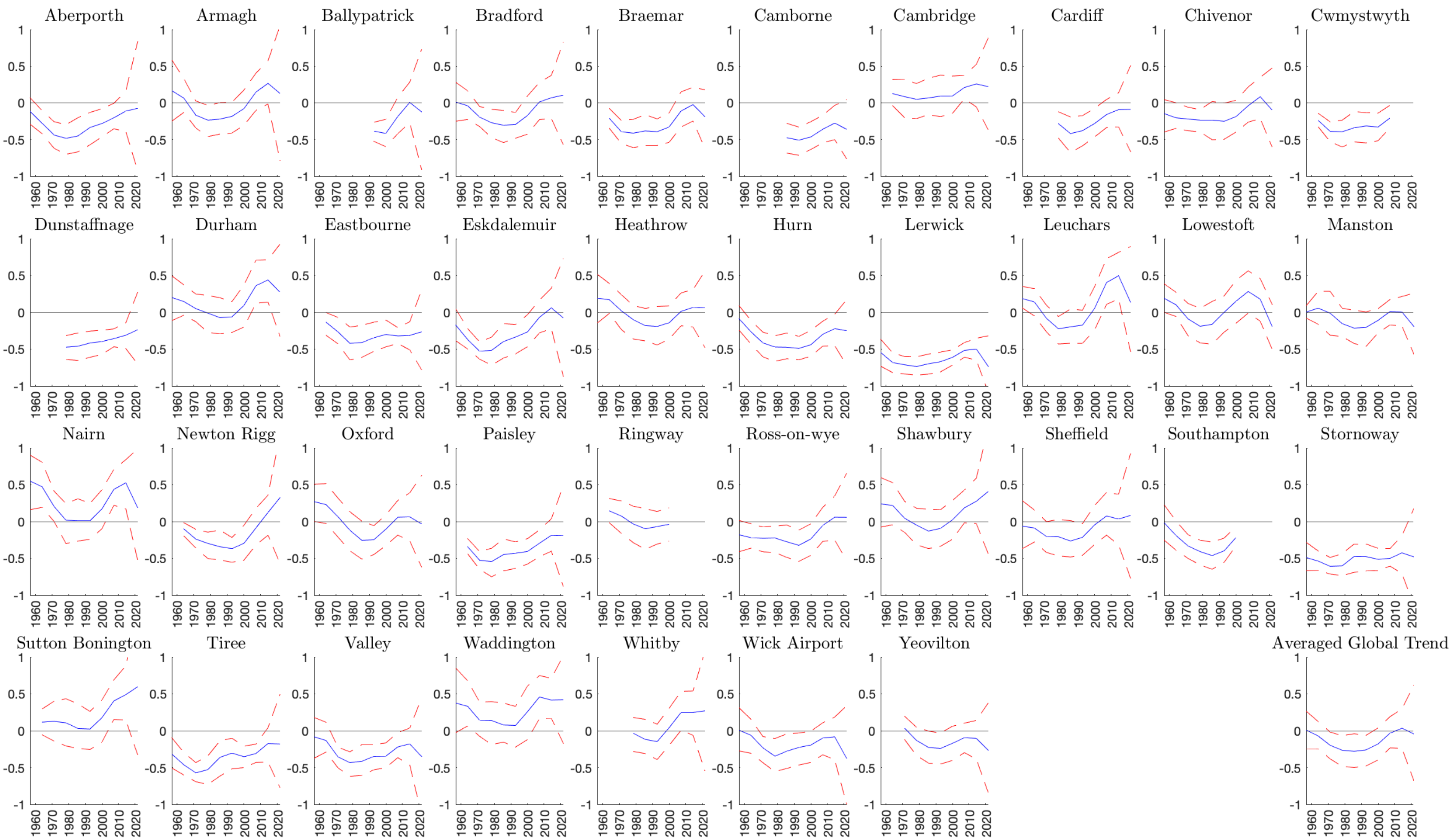


Figure A.3: The trends of autumn for rainfall data

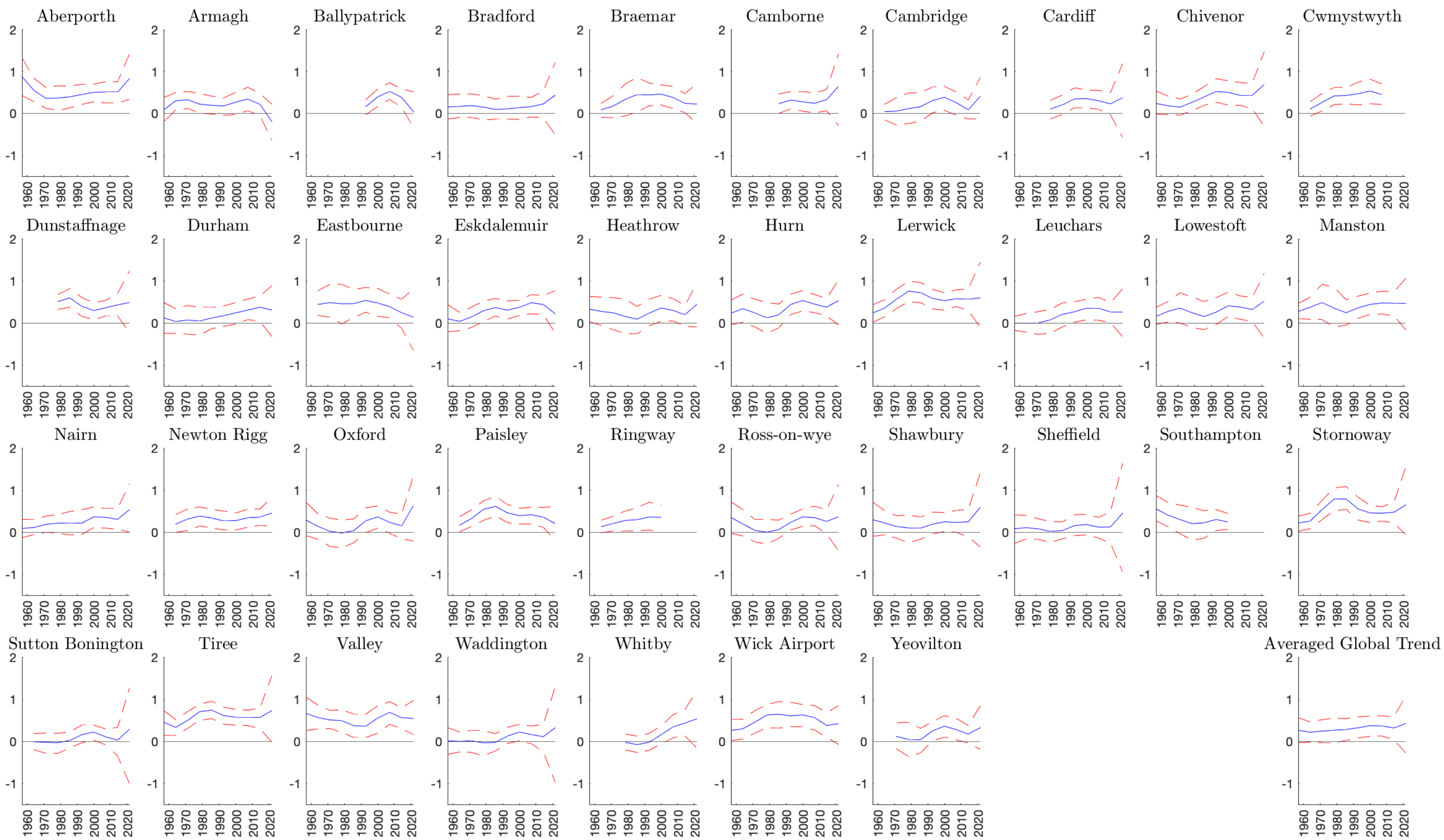


Figure A.4: The trends of winter for rainfall data

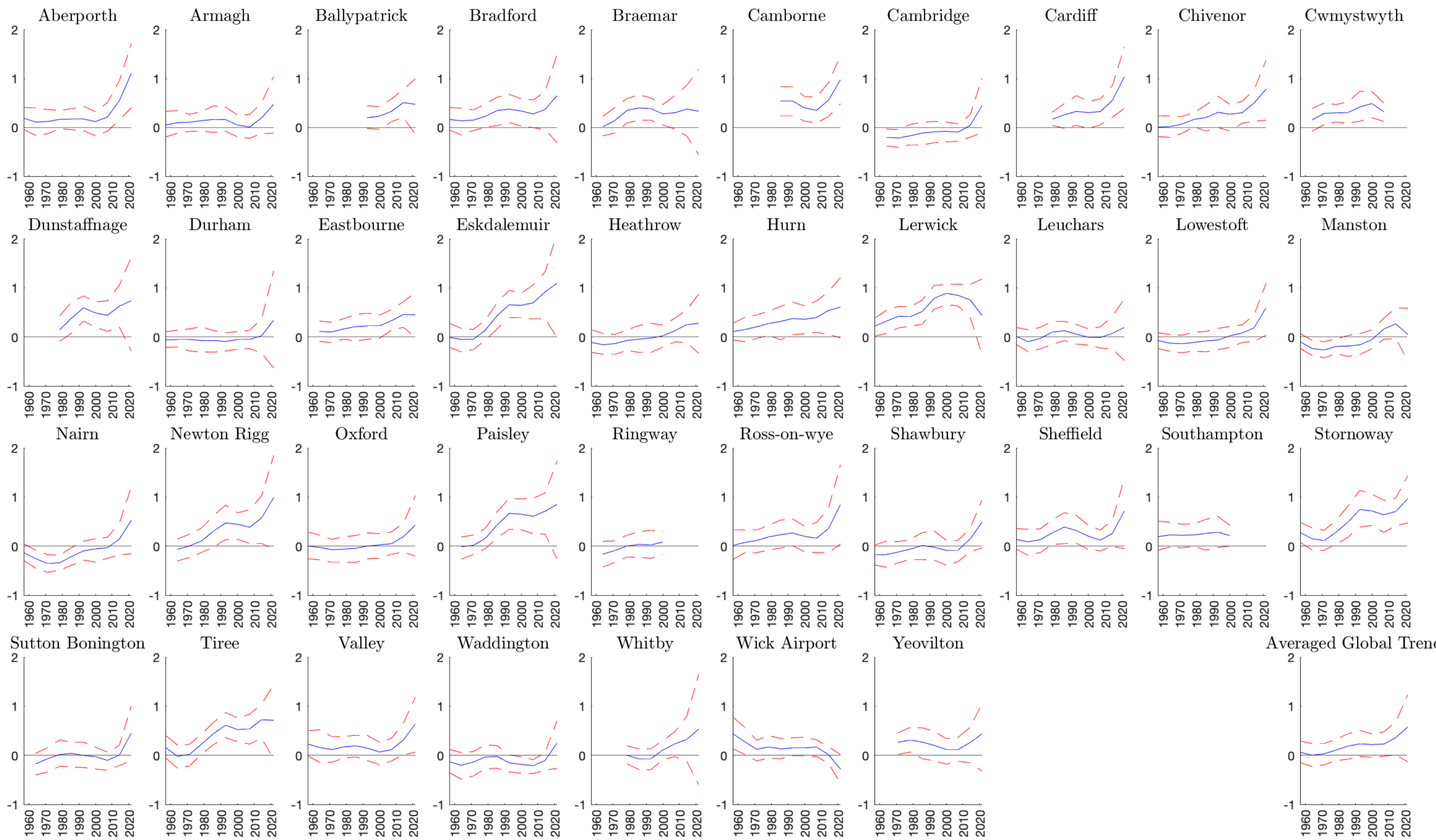


Figure A.5: The trends of spring for temperature data

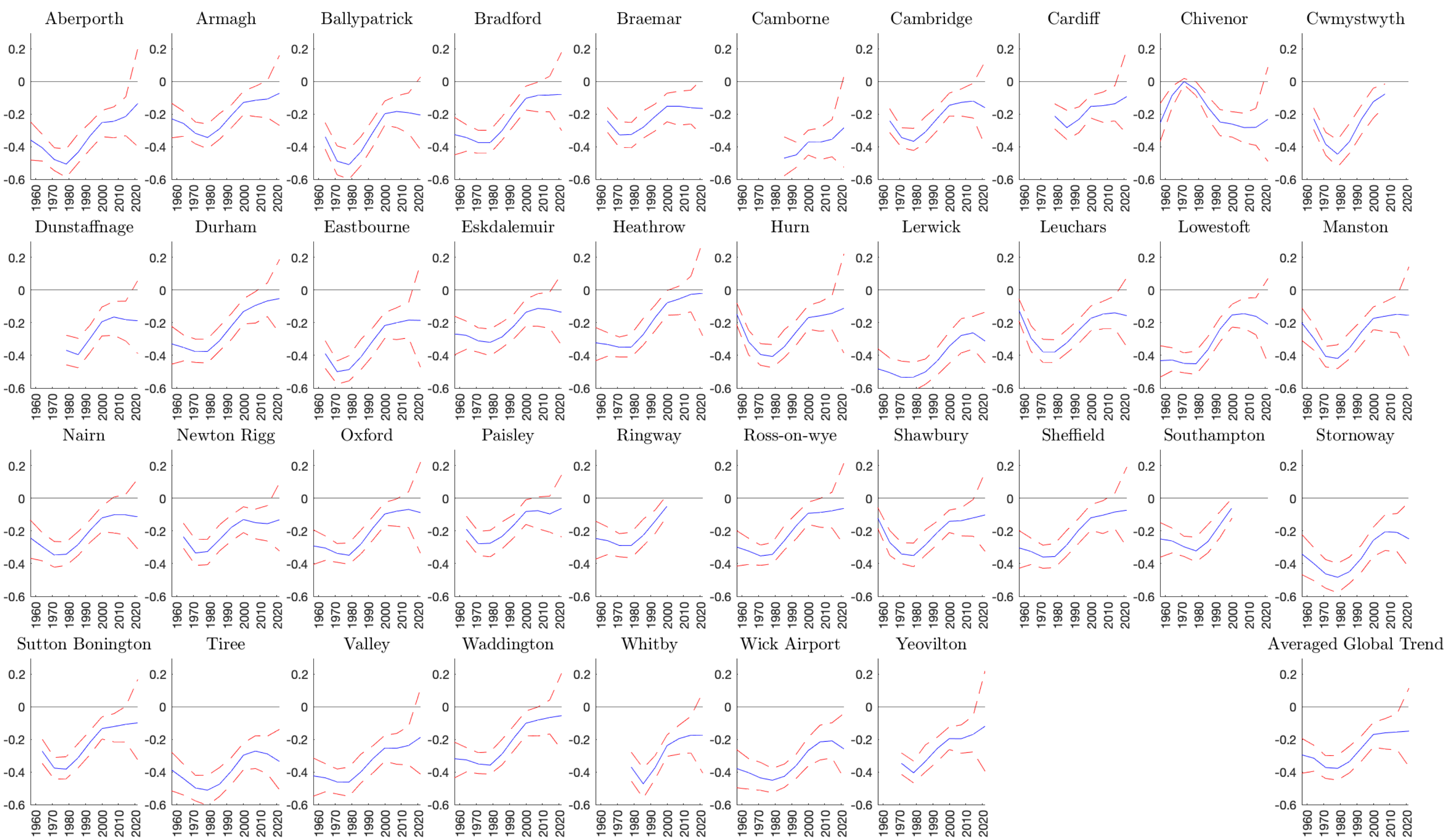


Figure A.6: The trends of summer for temperature data

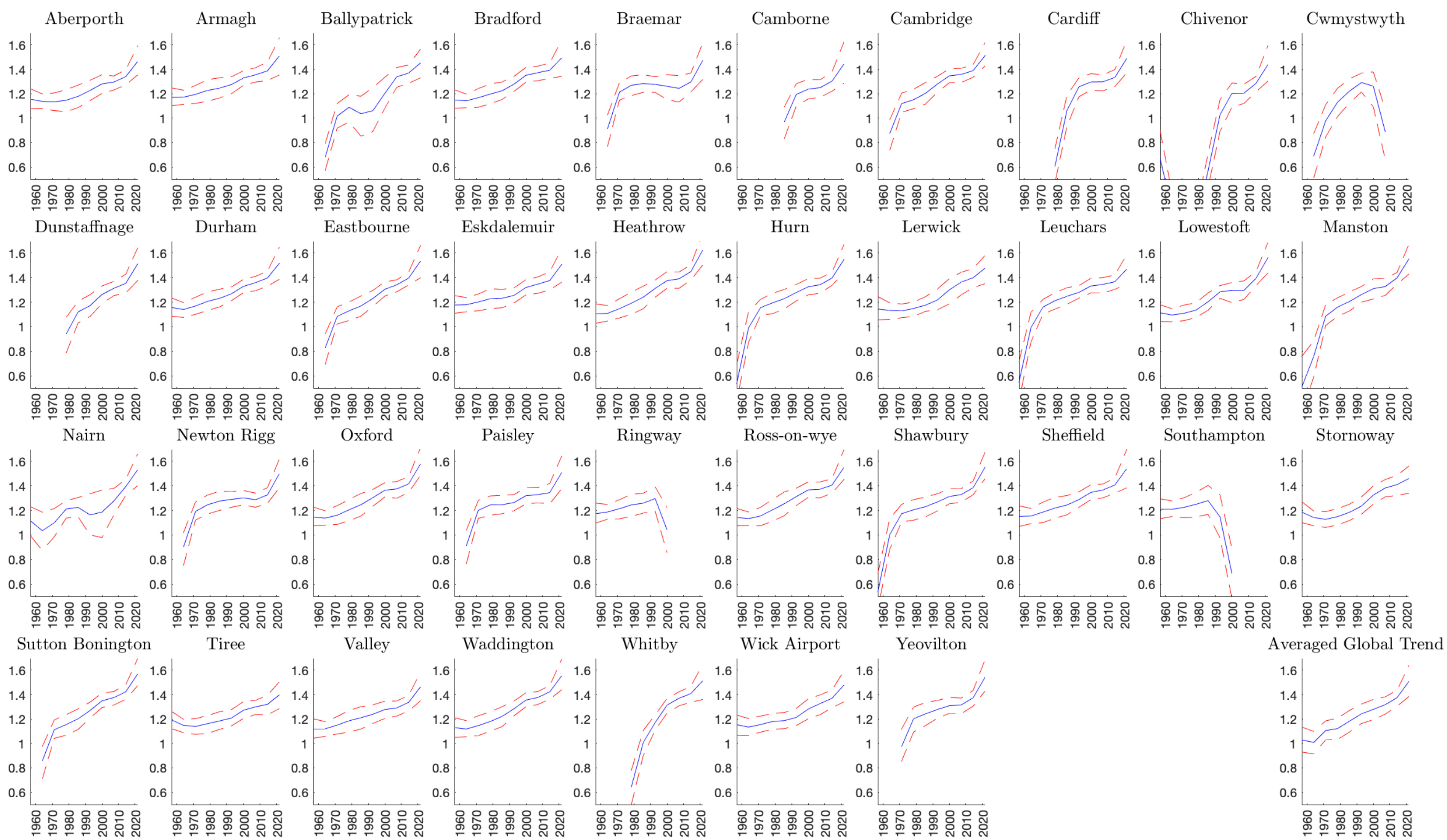


Figure A.7: The trends of autumn for temperature data

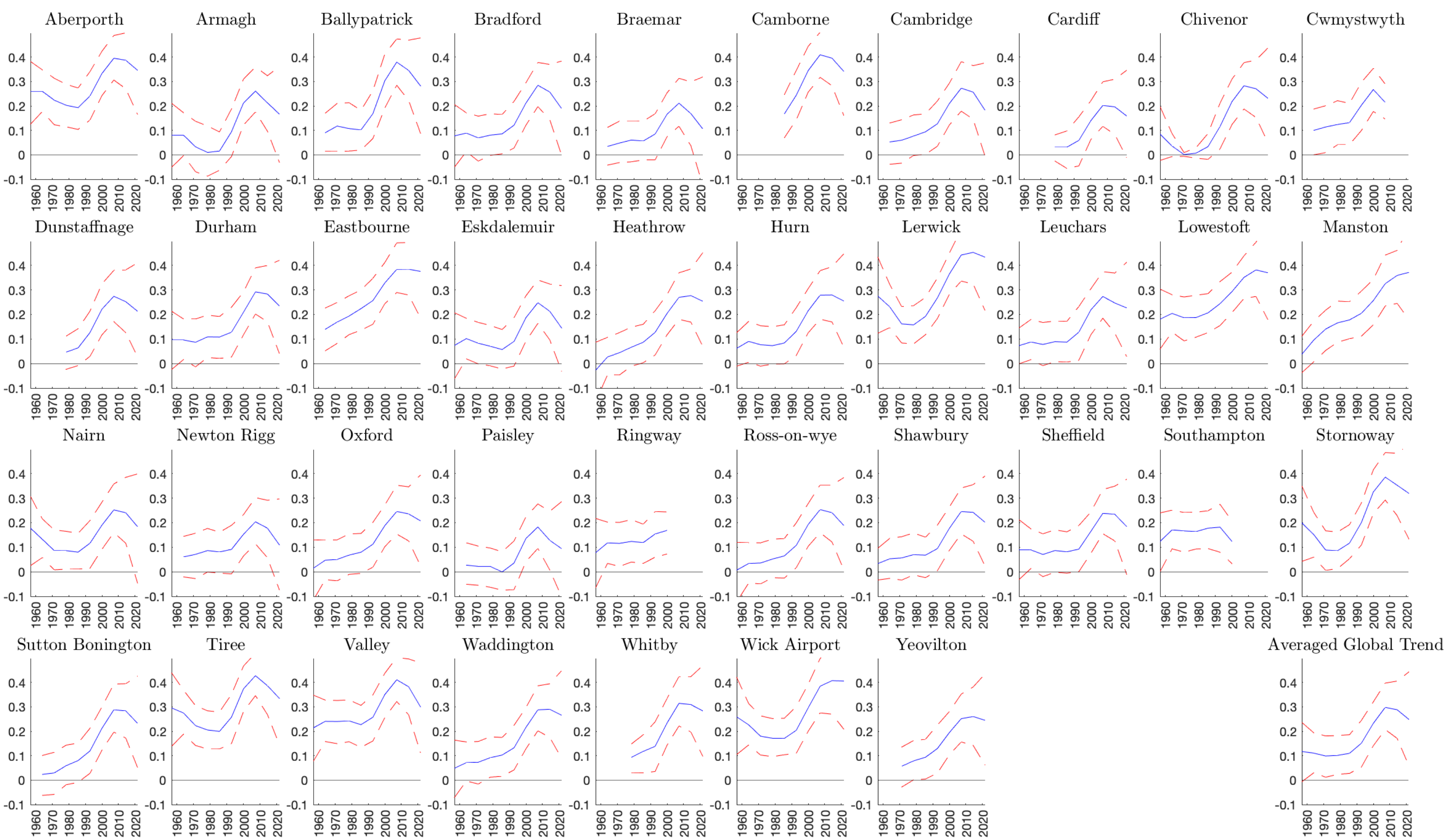
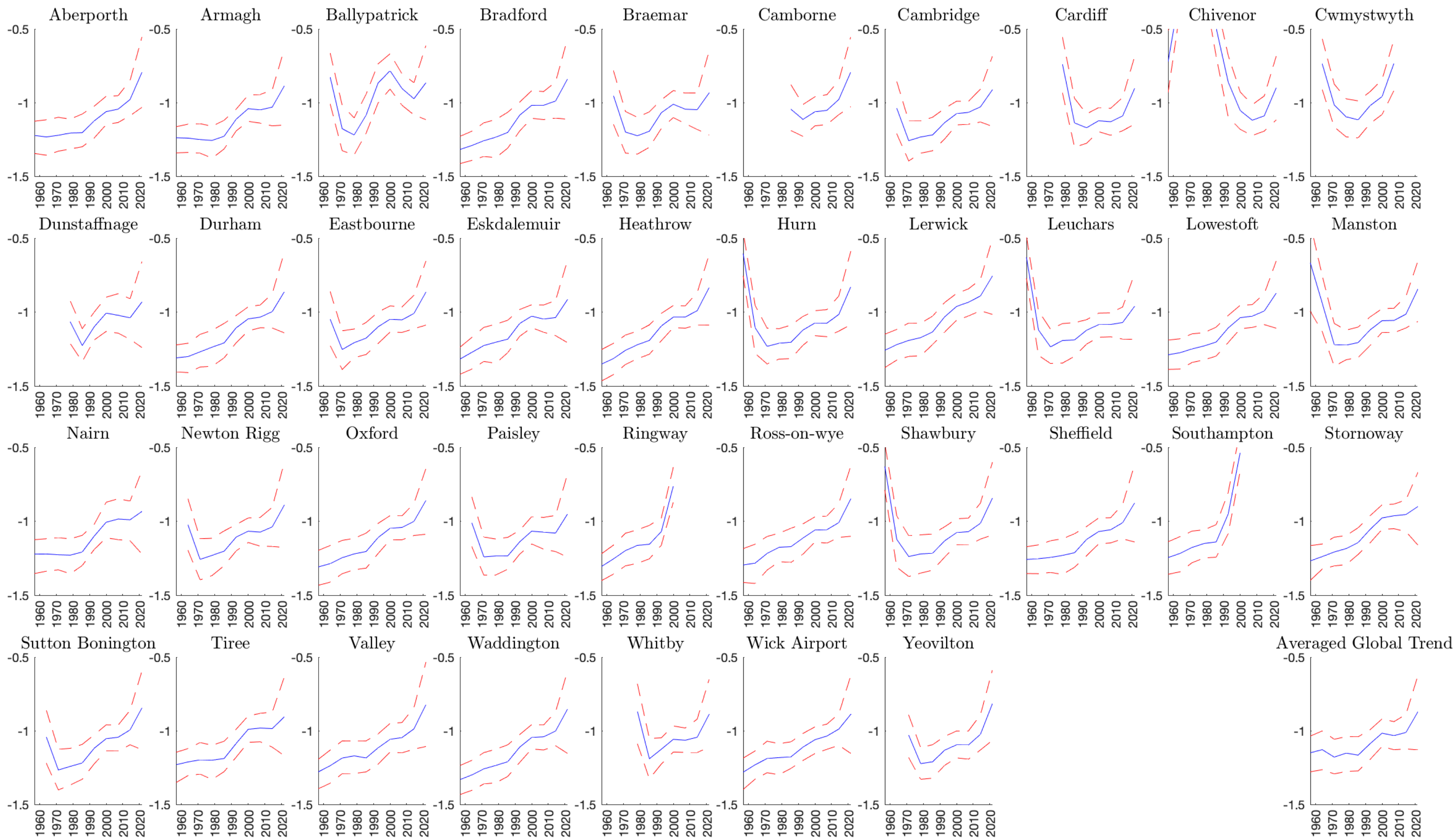


Figure A.8: The trends of winter for temperature data



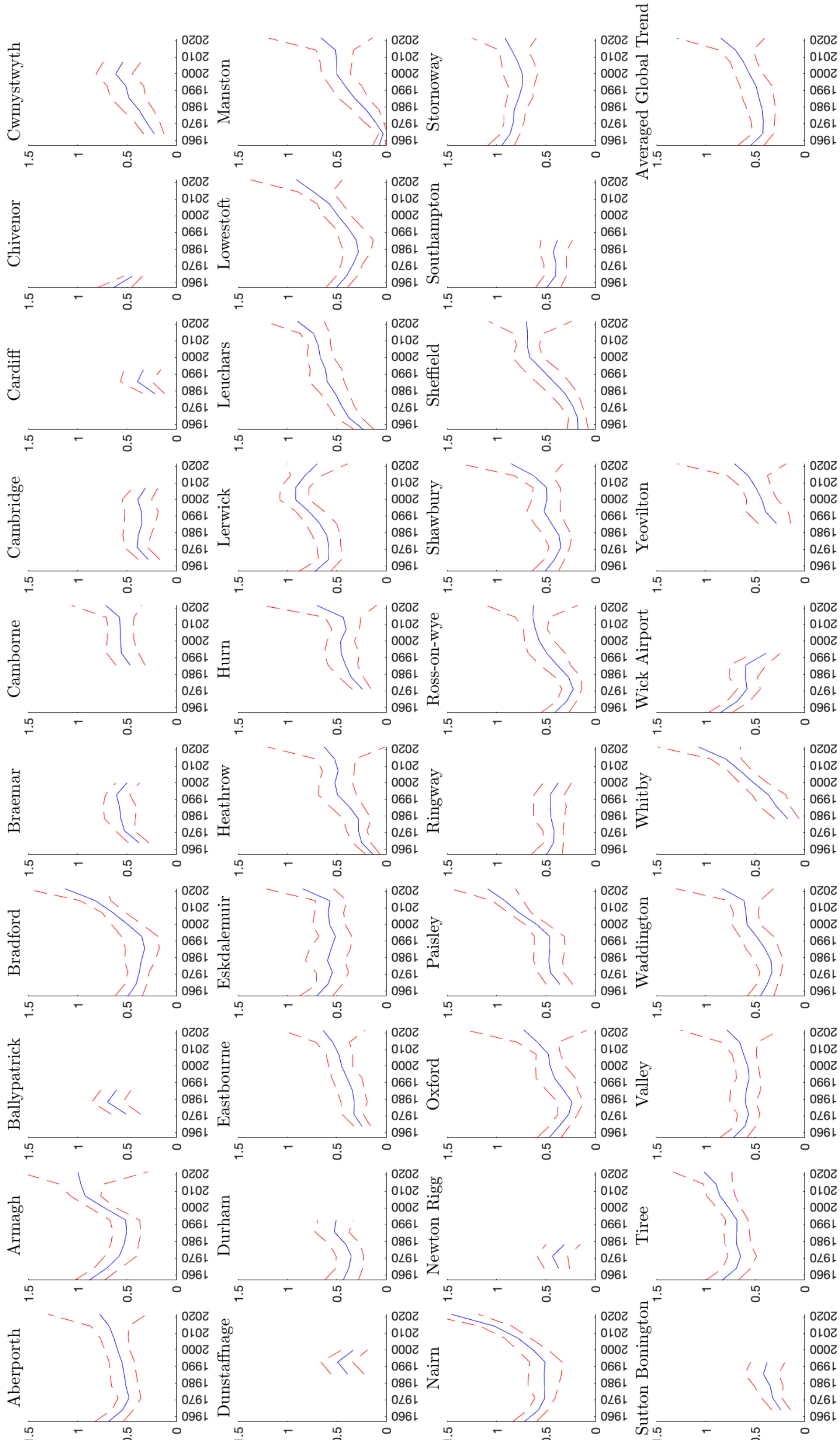


Figure A.9: The trends of spring for sunshine data

Figure A.10: The trends of summer for sunshine data

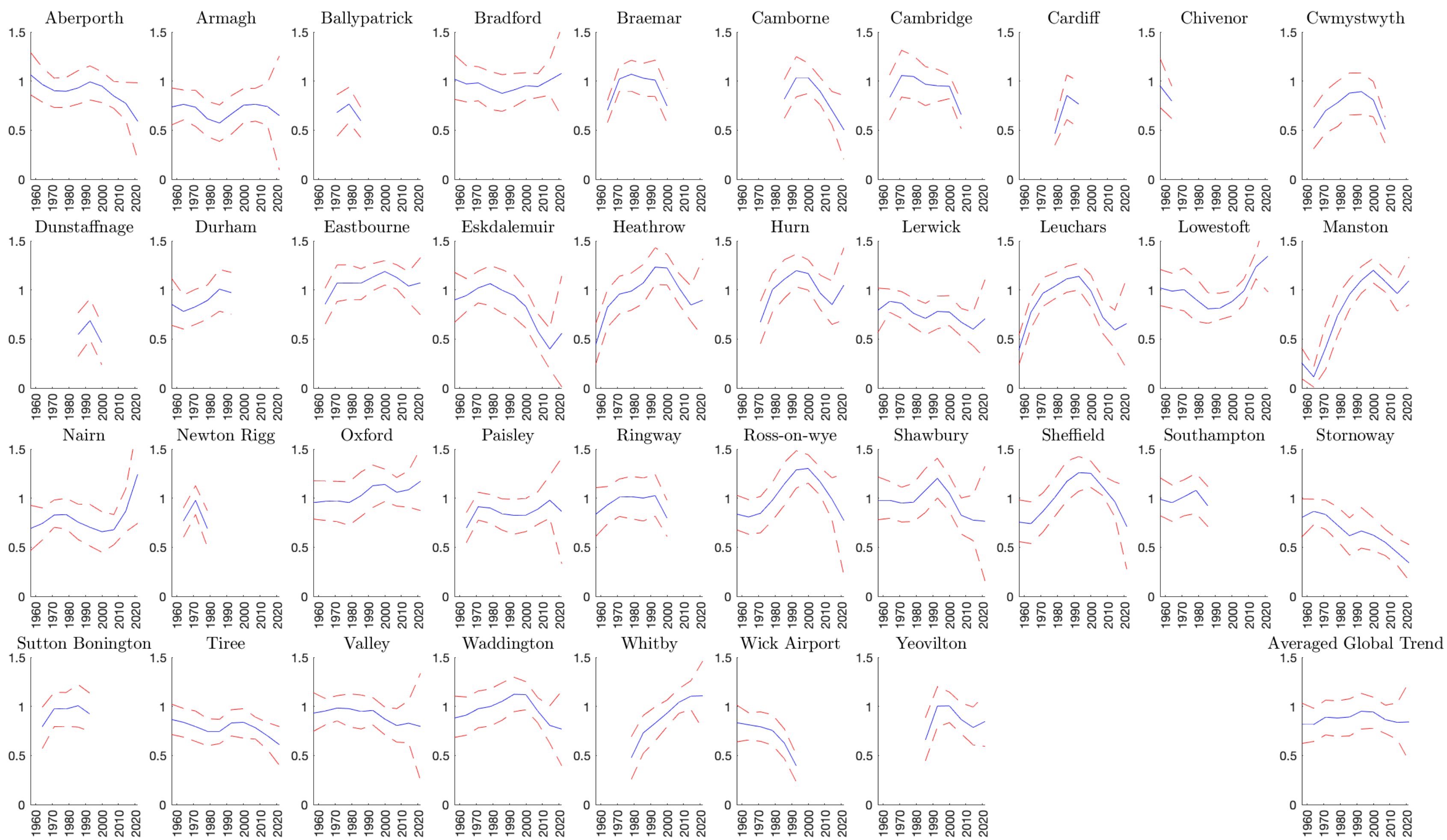


Figure A.11: The trends of autumn for sunshine data

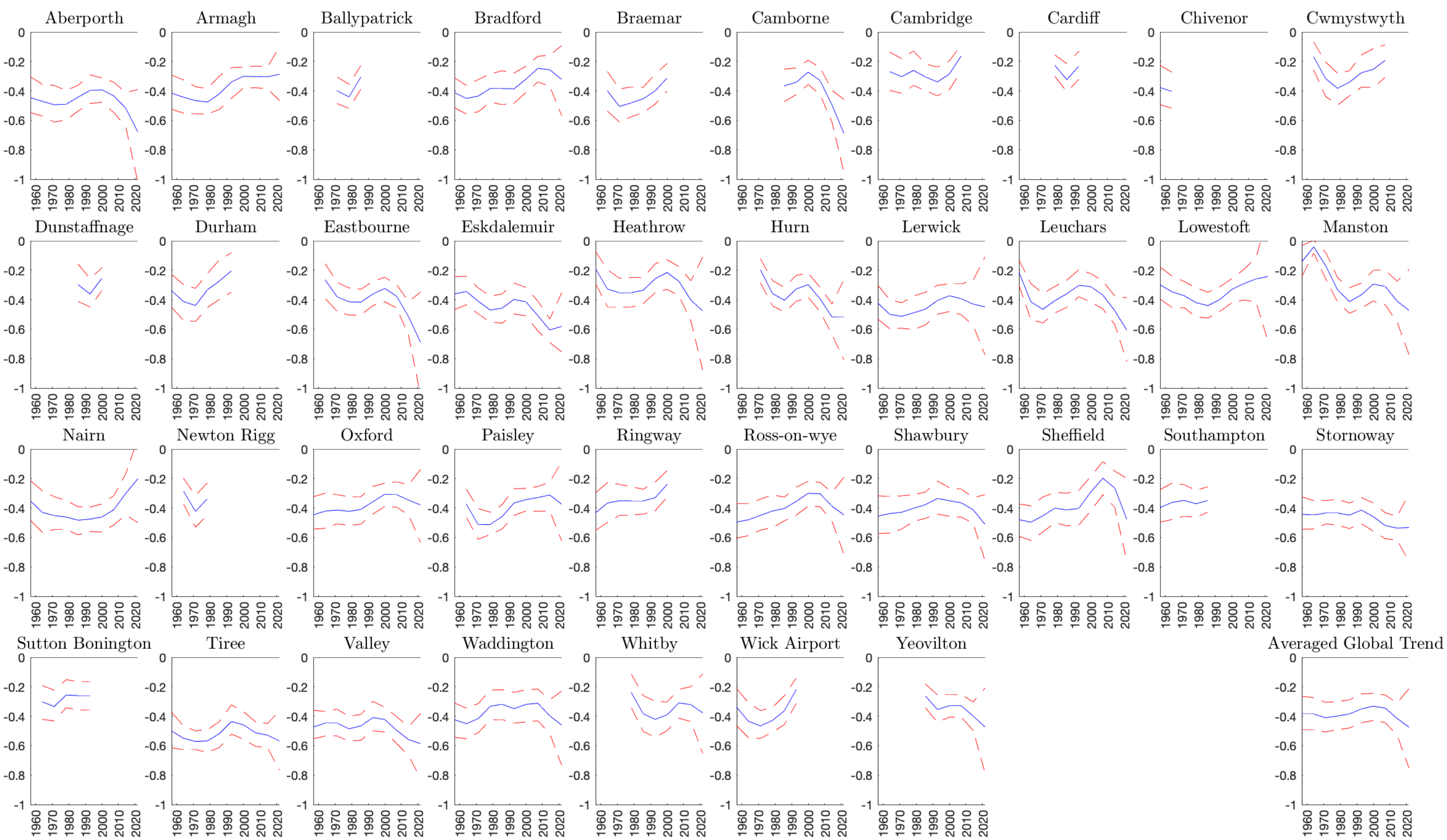
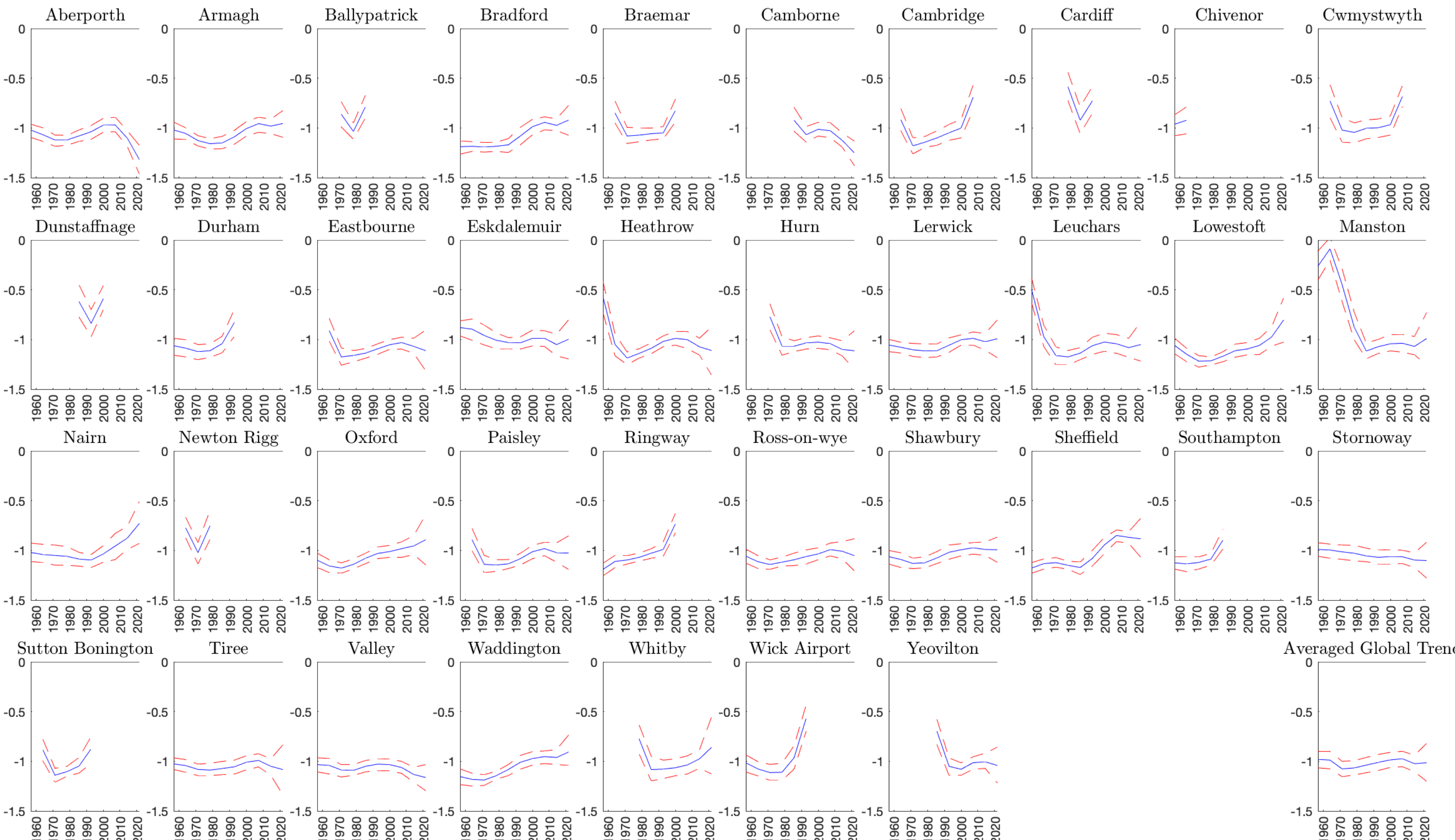


Figure A.12: The trends of winter for sunshine data



A.2 Preliminary Lemmas

Lemma A.1. *Suppose that A and $A+E$ are $n \times n$ symmetric matrices and that $Q = (Q_1, Q_2)$, where Q_1 is $n \times r$ and Q_2 is $n \times (n-r)$, is an orthogonal matrix such that* $\text{span}(Q_1)$ *is an invariant subspace for A. Decompose $Q^\top AQ$ and $Q^\top EQ$ as $Q^\top AQ =$* $\text{diag}(D_1, D_2)$ *and $Q^\top EQ = \{E_{ij}\}_{2\times 2}$. Let $\text{sep}(D_1, D_2) = \min_{\lambda_1 \in \lambda(D_1),\ \lambda_2 \in \lambda(D_2)} |\lambda_1 - \lambda_2|$. If* $\text{sep}(D_1, D_2) > 0$ *and $\|E\|_2 \leq \text{sep}(D_1, D_2)/5$, then there exists a $(n-r) \times r$ matrix P with $\|P\|_2 \leq 4\|E_{21}\|_2/\text{sep}(D_1, D_2)$, such that the columns of $Q_1^0 = (Q_1 + Q_2 P)(I_r + P^\top P)^{-1/2}$ define an orthonormal basis for a subspace that is invariant for $A + E$.*

Lemma A.2. *Suppose that $\{x_i, \mathcal{F}_i\}$ is an L^r mixingale for $r > 1$ such that there exist nonnegative constants $\{c_i : i \geq 1\}$ and $\{\psi_m : m \geq 0\}$ satisfying that (a). $\psi_m \to 0$ as $m \to \infty$, (b). $\sum_{k=1}^{\infty} \psi_k < \infty$, (c). for all $i \geq 0$ and $m \geq 0$, $\|E[x_i \mid \mathcal{F}_{i-m}]\|_r \leq c_i \psi_m$, and $\|x_i - E[x_i \mid \mathcal{F}_{i+m}]\|_r \leq c_i \psi_{m+1}$. Then there exists some $K < \infty$ such that for all $n \geq 1$, $\| \max_{j \leq n} |\sum_{i=1}^{j} x_i| \|_r \leq K \sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{n} c_i^2\right)^{1/2}$.*

Lemma A.3. *Under Assumptions 1 and 2, as $(N,T) \to (\infty, \infty)$*

1. $\sup_{\tau \in [0,1]} \left\| \frac{1}{N} \sum_{i=1}^{N} (\widehat{\beta}_{\tau,i} - \widetilde{m}_i(\tau)) \right\|_2 = O_P\left(h^2 + \frac{\sqrt{\log(NT)}}{\sqrt{NTh}}\right)$,

2. $\sup_{\tau \in [0,1]} \|\widehat{\beta}_{\tau,i} - \widetilde{m}_i(\tau) - \Delta_{\tau,i}\| = O_P\left(h^2 + \frac{\sqrt{\log(NT)}}{\sqrt{NTh}}\right)$,

where $\Delta_{\tau,i} := \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathcal{E}_i$.

A.3 Proofs

Proof of Lemma A.1:

The proof is given in Theorem 8.1.10 of Golub & Van Loan (2013), and is therefore omitted. ■

Proof of Lemma A.1:

The proof is given in Lemma 2 of Hansen (1991), and is therefore omitted. ■

As the proof of Lemma 2.1 is quite lengthy, we divide it into two parts. First, we show the first two results of Lemma 2.1 hold.

Proof of Lemma 2.1.1 and 2.1.2:

Observe that

$$Q_\tau(\mathbb{B}, \mathbb{F}) = \sum_{i=1}^{N} (DM_i - \mathbb{Z}\beta_i)^\top K_\tau M_{\mathbb{F}} K_\tau (DM_i - \mathbb{Z}\beta_i)$$

$$
\begin{aligned}
&+\sum_{i=1}^{N}\gamma_i^\top F^\top K_\tau M_{\mathbb{F}} K_\tau F\gamma_i+\sum_{i=1}^{N}\mathcal{E}_i^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i\\
&+2\sum_{i=1}^{N}(DM_i-\mathbb{Z}\beta_i)^\top K_\tau M_{\mathbb{F}} K_\tau F\gamma_i+2\sum_{i=1}^{N}(DM_i-\mathbb{Z}\beta_i)^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i\\
&+2\sum_{i=1}^{N}\gamma_i^\top F^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i\\
:=\ & A_1+A_2+A_3+2A_4+2A_5+2A_6, \qquad\qquad (\text{A.1})
\end{aligned}
$$

where the definitions of A_1 to A_6 are obvious. Below, we consider the terms on the right hand side of (A.1) one by one.

Start with A_1.

$$
\begin{aligned}
\frac{1}{NT}A_1 =\ & \frac{1}{NT}\sum_{i=1}^{N}(DM_i-\mathbb{Z}\widetilde{m}_i(\tau))^\top K_\tau M_{\mathbb{F}} K_\tau (DM_i-\mathbb{Z}\widetilde{m}_i(\tau))\\
&+\frac{1}{NT}\sum_{i=1}^{N}(\widetilde{m}_i(\tau)-\beta_i)^\top \mathbb{Z}^\top K_\tau M_{\mathbb{F}} K_\tau \mathbb{Z}(\widetilde{m}_i(\tau)-\beta_i)\\
&+\frac{2}{NT}\sum_{i=1}^{N}(DM_i-\mathbb{Z}\widetilde{m}_i(\tau))^\top K_\tau M_{\mathbb{F}} K_\tau \mathbb{Z}(\widetilde{m}_i(\tau)-\beta_i)\\
:=\ & \frac{1}{NT}A_{11}+\frac{1}{NT}A_{12}+\frac{2}{NT}A_{13},
\end{aligned}
$$

where the definitions of A_{11}, A_{12} and A_{13} are obvious.

For A_{11}, we write

$$
\begin{aligned}
&\sup_{\tau\in[0,1]}\frac{1}{NT}A_{11}\le \sup_{\tau\in[0,1]}\frac{1}{NT}\sum_{i=1}^{N}(DM_i-\mathbb{Z}\widetilde{m}_i(\tau))^\top K_\tau^2(DM_i-\mathbb{Z}\widetilde{m}_i(\tau))\\
&=\sup_{\tau\in[0,1]}\frac{1}{NTh}\sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{t=1}^{T}(m_{ij}(\tau_t)-z_t^\top\widetilde{m}_i(\tau))^2K\left(\frac{\tau_t-\tau}{h}\right)\\
&=\sup_{\tau\in[0,1]}\frac{(1+o(1))}{Nh}\sum_{j=1}^{J}\sum_{i=1}^{N}\int_0^1\left(m_{ij}(w)-m_{ij}(\tau)-m_i^{(1)}(\tau)(w-\tau)\right)^2K\left(\frac{w-\tau}{h}\right)dw\\
&=\sup_{\tau\in[0,1]}\frac{(1+o(1))}{N}\sum_{j=1}^{J}\sum_{i=1}^{N}\int_{-\tau/h}^{(1-\tau)/h}\Big(m_{ij}(\tau+wh)\\
&\quad -m_{ij}(\tau)-m_{ij}^{(1)}(\tau)(\tau+wh-\tau)\Big)^2K(w)dw\\
&=\sup_{\tau\in[0,1]}\frac{(1+o(1))}{N}\sum_{j=1}^{J}\sum_{i=1}^{N}\int_{-\tau/h}^{(1-\tau)/h}\left(\frac{1}{2}m_{ij}^{(2)}(\widetilde{\tau})(wh)^2\right)^2K(w)dw\cdot(1+o(1))\\
&=O(h^4), \qquad\qquad (\text{A.2})
\end{aligned}
$$

where $\widetilde{\tau}$ lies between τ and $\tau + wh$, the second equality follows from the definition of Riemann integral, and the fourth equality follows from the fact that $m_{ij}(\cdot)$'s are twice continuously differentiable on $[0,1]$ of Assumption 1.

For A_{13}, write

$$\begin{aligned}
\sup_{\tau\in[0,1]} \frac{1}{NT}|A_{13}| &= \sup_{\tau\in[0,1]} \frac{1}{NT}\left|\sum_{i=1}^{N}(DM_i - \mathbb{Z}\widetilde{m}_i(\tau))^\top K_\tau M_{\mathbb{F}} K_\tau \mathbb{Z}(\widetilde{m}_i(\tau) - \beta_i)\right| \\
&\le \sup_{\tau\in[0,1]} \left\{\frac{1}{NT}\sum_{i=1}^{N}(DM_i - \mathbb{Z}\widetilde{m}_i(\tau))^\top K_\tau^2 (DM_i - \mathbb{Z}\widetilde{m}_i(\tau))\right\}^{1/2} \\
&\quad \cdot \left\{\frac{1}{NT}\sum_{i=1}^{N}(\widetilde{m}_i(\tau) - \beta_i)^\top \mathbb{Z}^\top K_\tau^2 \mathbb{Z}(\widetilde{m}_i(\tau) - \beta_i)\right\}^{1/2} \\
&\le O_P(h^2)\cdot \sup_{\tau\in[0,1]} \left\{\frac{1}{N}\sum_{i=1}^{N}\|\widetilde{m}_i(\tau) - \beta_i\|^2 \cdot \frac{1}{T}\|\mathbb{Z}^\top K_\tau^2 \mathbb{Z}\|\right\}^{1/2} \\
&= O_P(h^2), \qquad \text{(A.3)}
\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz inequality and $\|M_{\mathbb{F}}\|_2 = 1$, the second inequality follows from (A.2), and the last line follows from $\max_i \|\beta_i\| < \infty$.

By (A.2) and (A.3), we can conclude that

$$\sup_{\tau\in[0,1]} \frac{1}{NT}|A_1 - A_{12}| = O(h^2).$$

For A_3, we write

$$\begin{aligned}
A_3 &= \sum_{i=1}^{N} \mathcal{E}_i^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i = \sum_{i=1}^{N} \mathcal{E}_i^\top K_\tau^2 \mathcal{E}_i + \sum_{i=1}^{N} \mathcal{E}_i^\top K_\tau P_{\mathbb{F}} K_\tau \mathcal{E}_i \\
&:= A_{31} + A_{32},
\end{aligned}$$

where the definitions of A_{31} and A_{32} are obvious. We skip the term A_{31} for now, as it will be cancelled automatically in the following development. Thus, consider A_{32} and write

$$\begin{aligned}
\sup_{\tau\in[0,1]} \frac{1}{NT}|A_{32}| &= \sup_{\tau\in[0,1]} \frac{1}{NT}\sum_{i=1}^{N} \mathcal{E}_i^\top K_\tau P_{\mathbb{F}} K_\tau \mathcal{E}_i = \sup_{\tau\in[0,1]} \frac{1}{NT} \operatorname{Tr}\left\{K_\tau P_{\mathbb{F}} K_\tau \mathcal{E}^\top \mathcal{E}\right\} \\
&\le \sup_{\tau\in[0,1]} O(1)\frac{1}{NTh}\|\mathcal{E}\|_2^2 = O_P\left(\frac{1}{(N\wedge T)h}\right),
\end{aligned}$$

where the inequality follows from $|\operatorname{Tr}\{A\}| \le \operatorname{rank}(A)\|A\|_2$ and the facts that $\mathcal{K}(\cdot)$ is uniformly bounded on $[-1,1]$ and $\|P_{\mathbb{F}}\|_2 = 1$, and the last equality follows from $\|\mathcal{E}\|_2 = O_P(\sqrt{N} \vee \sqrt{T})$ of Assumption 1.

For A_5, we write

$$\begin{aligned} A_5 &= \sum_{i=1}^N (DM_i - \mathbb{Z}\beta_i)^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i \\ &= \sum_{i=1}^N (DM_i - \mathbb{Z}\widetilde{m}_i(\tau))^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i + \sum_{i=1}^N (\widetilde{m}_i(\tau) - \beta_i)^\top \mathbb{Z}^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i \\ &:= A_{51} + A_{52}, \end{aligned}$$

where the definitions of A_{51} and A_{52} are obvious.

Using a development similar to (A.3), we obtain

$$\sup_{\tau\in[0,1]} \frac{1}{NT} |A_{51}| = O_P(h^2).$$

For A_{52}, we can write

$$\begin{aligned} &\sup_{\tau\in[0,1]} \frac{1}{NT} \left| \sum_{i=1}^N (\widetilde{m}_i(\tau) - \beta_i)^\top \mathbb{Z}^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}_i \right| \\ &= \sup_{\tau\in[0,1]} \frac{1}{NT} |\operatorname{Tr}\{\mathbb{Z}^\top K_\tau M_{\mathbb{F}} K_\tau \mathcal{E}^\top (\mathbb{B} - \mathbb{B}_{0,\tau})\}| \\ &\le O_P(1) \sup_{\tau\in[0,1]} \frac{1}{\sqrt{NT}} \|\mathbb{Z}^\top K_\tau\|_2 \cdot \|K_\tau\|_2 \cdot \|\mathcal{E}\|_2 \\ &= O_P\left(\frac{1}{\sqrt{(N\wedge T)h}} \right), \end{aligned} \tag{A.4}$$

where $\mathbb{B}_{0,\tau} = (\widetilde{m}_1(\tau), \ldots, \widetilde{m}_N(\tau))^\top$, the first inequality follows from $|\operatorname{Tr}\{A\}| \le \operatorname{rank}(A)\|A\|_2$ and the fact that $\max_i \|\beta_i\| < \infty$, and the last line follows from $\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}} \|\mathbb{Z}K_\tau\|_2 = O_P(1)$, $\|K_r\|_2 = O(\frac{1}{\sqrt{h}})$, and $\|\mathcal{E}\|_2 = O_P(\sqrt{N} \vee \sqrt{T})$ of Assumption 1.

By the results associated with A_{51} and A_{52}, we can conclude that

$$\sup_{\tau\in[0,1]} \frac{1}{NT} |A_5| = O_P\left(h^2 + \frac{1}{\sqrt{(N\wedge T)h}} \right).$$

Similar to (A.4), we can also conclude that

$$\sup_{\tau\in[0,1]} \frac{1}{NT} |A_6| = O_P\left(\frac{1}{\sqrt{(N\wedge T)h}} \right).$$

Up to this point, we know that to investigate $Q_\tau(\mathbb{B}, \mathbb{F})$, we need only to focus on A_{12}, A_2, A_{31} and A_4 in what follows. We consider $Q_\tau(\mathbb{B}, \mathbb{F}) - Q_\tau(\mathbb{B}_{0,\tau}, K_\tau F)$, in which taking the difference further eliminates A_{31}, and $\mathbb{B}_{0,\tau}$ has been defined in (A.4).

$$\begin{aligned} &\frac{1}{NT} \left(Q_\tau(\mathbb{B}, \mathbb{F}) - Q_\tau(\mathbb{B}_{0,\tau}, K_\tau F) \right) \\ &= \frac{1}{NT} \sum_{i=1}^N (\mathbb{Z}(\widetilde{m}_i(\tau) - \beta_i) + F\gamma_i)^\top K_\tau M_{\mathbb{F}} K_\tau (\mathbb{Z}(\widetilde{m}_i(\tau) - \beta_i) + F\gamma_i) \end{aligned}$$

$$+O_P\left(h^2+\frac{1}{\sqrt{Nh}}\vee\frac{1}{\sqrt{Th}}\right)$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\left[(\widetilde{m}_i(\tau)-\beta_i)^\top\mathbb{A}_{\tau,\mathbb{F}}(\widetilde{m}_i(\tau)-\beta_i)+\eta_{\mathbb{F}}^\top\mathbb{B}\eta_{\mathbb{F}}+(m_i^*(\tau)-\beta_i)^\top\mathbb{C}_{i,\tau,\mathbb{F}}^\top\eta_{\mathbb{F}}\right]$$

$$+O_P\left(h^2+\frac{1}{\sqrt{Nh}}\vee\frac{1}{\sqrt{Th}}\right), \tag{A.5}$$

where the second equality follows from Assumption 1.3.b, $\mathbb{A}_{\tau,\mathbb{F}}=\mathbb{Z}^\top K_\tau M_{\mathbb{F}}K_\tau\mathbb{Z}$, $\mathbb{B}=\Sigma_\gamma\otimes I_T$, $\mathbb{C}_{i,\tau,\mathbb{F}}=\gamma_i\otimes(M_{\mathbb{F}}K_\tau\mathbb{Z})$, and $\eta_{\mathbb{F}}=\mathrm{vec}(M_{\mathbb{F}}K_\tau F)$.

Note that if $\sup_{\tau\in[0,1]}\frac{1}{\sqrt{T}}\|\eta_{\widehat{\mathbb{F}}_\tau}\|=o_P(1)$, we can obtain that

$$o_P(1)=\sup_{\tau\in[0,1]}\frac{1}{T}\|F^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau F\|=\sup_{\tau\in[0,1]}\left\|\frac{1}{T}F^\top K_\tau^2F-\frac{1}{T}F^\top K_\tau\frac{1}{T}\widehat{\mathbb{F}}_\tau\widehat{\mathbb{F}}_\tau^\top K_\tau F\right\|,$$

which further yields that

$$o_P(1)=\sup_{\tau\in[0,1]}\left\|2\,\mathrm{Tr}\left(I-\frac{\widehat{\mathbb{F}}_\tau^\top P_{K_\tau F}\widehat{\mathbb{F}}_\tau}{T}\right)\right\|=\sup_{\tau\in[0,1]}\|2\,\mathrm{Tr}(I_{d_f})-2\,\mathrm{Tr}(P_{K_\tau F}P_{\widehat{\mathbb{F}}_\tau})\|$$

$$=\sup_{\tau\in[0,1]}\mathrm{Tr}((P_{K_\tau F}-P_{\widehat{\mathbb{F}}_\tau})^2)=\sup_{\tau\in[0,1]}\|P_{K_\tau F}-P_{\widehat{\mathbb{F}}_\tau}\|^2.$$

Then the first result of this lemma is proved.

That said, we now show that if $\frac{1}{\sqrt{T}}\|\eta_{\widehat{\mathbb{F}}_\tau}\|\neq o_P(1)$ for some given τ, then

$$0\geq Q_\tau(\widehat{\mathbb{B}},\widehat{\mathbb{F}}_\tau)-Q_\tau(\mathbb{B}_{0,\tau},K_\tau F) \tag{A.6}$$

can not be fulfilled with probability one. This can be done by considering the cases: (i). $\sup_{\tau\in[0,1]}\frac{1}{N}\|\widehat{\mathbb{B}}-\mathbb{B}_{0,\tau}\|^2=o_P(1)$, and (ii). $\sup_{\tau\in[0,1]}\frac{1}{N}\|\widehat{\mathbb{B}}-\mathbb{B}_{0,\tau}\|^2\neq o_P(1)$ respectively as follows.

Consider Case (i) first. In this case, it is easy to show that (A.5) can be simplified as

$$\frac{1}{NT}\left(Q_\tau(\mathbb{B},\mathbb{F})-Q_\tau(\mathbb{B}_{0,\tau},K_\tau F)\right)=\frac{1}{T}\eta_{\mathbb{F}}^\top\mathbb{B}\eta_{\mathbb{F}}+o_P(1),$$

which is contradictory to (A.6).

Consider Case (ii). Without loss of generality, we suppose that $E[\gamma_i]\equiv\gamma$ below. Denote $\gamma_i^*=\gamma_i-\gamma$. Using (A.5), we have

$$\frac{1}{NT}\left(Q_\tau(\mathbb{B},\mathbb{F})-Q_\tau(\mathbb{B}_{0,\tau},K_\tau F)\right)$$

$$=\frac{1}{NT}\sum_{i=1}^{N}(\mathbb{Z}(\widetilde{m}_i(\tau)-\beta_i)+F\gamma)^\top K_\tau M_{\mathbb{F}}K_\tau(\mathbb{Z}(\widetilde{m}_i(\tau)-\beta_i)+F\gamma)$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\gamma_i^{*\top}F^\top K_\tau M_{\mathbb{F}}K_\tau F\gamma_i^*+O_P\left(h^2+\frac{1}{\sqrt{(N\wedge T)h}}\right),$$

where the equality follows from the fact that $\max_i \|\beta_i\| < \infty$. Then it is obvious that the right hand side is contradictory to (A.6).

Thus, we conclude that $\sup_{\tau\in[0,1]} \|P_{K_\tau F} - P_{\widehat{\mathbb{F}}_\tau}\| = o_P(1)$, which in connection with (A.5) yields that $\sup_{\tau\in[0,1]} \frac{1}{N}\|\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau}\|^2 = o_P(1)$. The proof of the first two results of Lemma 2.1 is now completed.

In fact, (A.5) further indicates that

$$\sup_{\tau\in[0,1]} \frac{1}{\sqrt{N}}\|\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau}\| = O_P\left(h^2 + \frac{1}{\sqrt{(N\wedge T)h}}\right), \tag{A.7}$$

which will be used later. ■

Proof of Lemma 2.1.3:

Write

$$\begin{aligned}
\widehat{\mathbb{F}}_\tau \widehat{V}_\tau =& \frac{1}{NT}\sum_{i=1}^N K_\tau(DM_i - \mathbb{Z}\widehat{\beta}_i)(DM_i - \mathbb{Z}\widehat{\beta}_i)^\top K_\tau \widehat{\mathbb{F}}_\tau \\
&+\frac{1}{NT}\sum_{i=1}^N K_\tau(DM_i - \mathbb{Z}\widehat{\beta}_i)\gamma_i^\top F^\top K_\tau\widehat{\mathbb{F}}_\tau + \frac{1}{NT}\sum_{i=1}^N K_\tau F\gamma_i(DM_i - \mathbb{Z}\widehat{\beta}_i)^\top K_\tau\widehat{\mathbb{F}}_\tau \\
&+\frac{1}{NT}\sum_{i=1}^N K_\tau(DM_i - \mathbb{Z}\widehat{\beta}_i)\mathcal{E}_i^\top K_\tau\widehat{\mathbb{F}}_\tau + \frac{1}{NT}\sum_{i=1}^N K_\tau\mathcal{E}_i(DM_i - \mathbb{Z}\widehat{\beta}_i)^\top K_\tau\widehat{\mathbb{F}}_\tau \\
&+\frac{1}{NT}\sum_{i=1}^N K_\tau F\gamma_i\gamma_i^\top F^\top K_\tau\widehat{\mathbb{F}}_\tau + \frac{1}{NT}\sum_{i=1}^N K_\tau F\gamma_i\mathcal{E}_i^\top K_\tau\widehat{\mathbb{F}}_\tau \\
&+\frac{1}{NT}\sum_{i=1}^N K_\tau\mathcal{E}_i\gamma_i^\top F^\top K_\tau\widehat{\mathbb{F}}_\tau + \frac{1}{NT}\sum_{i=1}^N K_\tau\mathcal{E}_i\mathcal{E}_i^\top K_\tau\widehat{\mathbb{F}}_\tau \\
:=& A_1(\tau) + \cdots + A_9(\tau),
\end{aligned} \tag{A.8}$$

where the definitions of $A_1(\tau)$ to $A_9(\tau)$ are obvious. Below, we investigate the terms on the right hand side of (A.8) one by one.

For $A_1(\tau)$, write

$$\begin{aligned}
\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\|A_1(\tau)\|_2 &= \frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\sum_{i=1}^N K_\tau(DM_i - \mathbb{Z}\widehat{\beta}_i)(DM_i - \mathbb{Z}\widehat{\beta}_i)^\top K_\tau\widehat{\mathbb{F}}_\tau\right\|_2 \\
&\le O(1)\frac{1}{NT}\sum_{i=1}^N (DM_i - \mathbb{Z}\widehat{\beta}_i)^\top K_\tau^2 (DM_i - \mathbb{Z}\widehat{\beta}_i) \\
&= O_P\left(\frac{1}{N}\|\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau}\|^2 + h^4\right),
\end{aligned} \tag{A.9}$$

where the first inequality follows from the fact that $\frac{1}{T}\widehat{\mathbb{F}}_\tau^\top\widehat{\mathbb{F}}_\tau = I_{d_f}$, and the second equality follows from (A.2).

Using the Cauchy-Schwarz inequality and (A.9), it is easy to show that

$$\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\|A_\ell(\tau)\|_2 = O_P\left(\frac{1}{\sqrt{N}}\|\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau}\| + h^2\right)$$

for $\ell = 2, 3, 4, 5$.

For $A_7(\tau)$, write

$$\begin{aligned}\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\|A_7(\tau)\|_2 &= \sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\sum_{i=1}^{N} K_\tau F\gamma_i\mathcal{E}_i^\top K_\tau\widehat{\mathbb{F}}_\tau\right\|_2 \\ &\le O(1) \sup_{\tau\in[0,1]} \frac{1}{NT}\|K_\tau F\Gamma^\top\mathcal{E}K_\tau\|_2 \\ &\le O \sup_{\tau\in[0,1]} \frac{1}{NT}\|K_\tau F\|_2\|\Gamma\|_2\|K_\tau\|_2\|\mathcal{E}\|_2 \\ &\le O_P(1) \sup_{\tau\in[0,1]} \frac{1}{NT}\sqrt{T}\sqrt{N}\frac{1}{\sqrt{h}}(\sqrt{N}\vee\sqrt{T}) \\ &\le O_P\left(\frac{1}{\sqrt{(N\wedge T)h}}\right),\end{aligned}$$

where the first inequality follows from the fact that $\frac{1}{T}\widehat{\mathbb{F}}_\tau^\top\widehat{\mathbb{F}}_\tau = I_{d_f}$, the third inequality follows from $\sup_{\tau\in[0,1]}\frac{1}{\sqrt{T}}\|K_\tau F\|_2 = O_P(1)$, $\|\Gamma\|_2 = O_P(\sqrt{N})$ and $\|\mathcal{E}\|_2 = O_P(\sqrt{N}\vee\sqrt{T})$ of Assumption 1. Similarly, we obtain that

$$\begin{aligned}&\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\|A_8(\tau)\|_2 = O_P\left(\frac{1}{\sqrt{(N\wedge T)h}}\right), \\ &\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\|A_9(\tau)\|_2 \le O(1)\frac{1}{NTh}\|\mathcal{E}\|_2^2 = O_P\left(\frac{1}{(N\wedge T)h}\right).\end{aligned}$$

Collecting the above results, we immediately obtain that

$$\begin{aligned}&\sup_{\tau\in[0,1]}\left\|\widehat{V}_\tau - \frac{\widehat{\mathbb{F}}_\tau^\top K_\tau F}{T}\cdot\frac{\Gamma^\top\Gamma}{N}\cdot\frac{FK_\tau\widehat{\mathbb{F}}_\tau}{T}\right\|_2 \\ &= O_P\left(\frac{1}{\sqrt{N}}\|\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau}\| + h^2 + \frac{1}{\sqrt{(N\wedge T)h}}\right). \qquad (A.10)\end{aligned}$$

Note that by Assumption 1,

$$\frac{1}{N}\Gamma^\top\Gamma = \Sigma_\gamma + o_P(1),$$

and

$$\sup_{\tau\in[0,1]} \frac{1}{T}\|\widehat{\mathbb{F}}_\tau^\top K_\tau F\|_2 \le \sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}}\|\widehat{\mathbb{F}}_\tau\|\cdot\frac{1}{\sqrt{T}}\|K_\tau F\|_2 = O_P(1),$$

which in connection with (A.10) immediately yields that $\sup_{\tau\in[0,1]}\|\widehat{V}_\tau\| = O_P(1)$.

Note further that

$$\begin{aligned}\lambda_{\min}\left(\frac{\widehat{\mathbb{F}}_\tau^\top K_\tau F}{T}\cdot\frac{\Gamma^\top\Gamma}{N}\cdot\frac{FK_\tau\widehat{\mathbb{F}}_\tau}{T}\right) &\asymp \lambda_{\min}\left(\frac{\widehat{\mathbb{F}}_\tau^\top K_\tau F}{T}\cdot\frac{FK_\tau\widehat{\mathbb{F}}_\tau}{T}\right)\\ &\asymp \lambda_{\min}\left(\frac{FK_\tau\widehat{\mathbb{F}}_\tau}{T}\cdot\frac{\widehat{\mathbb{F}}_\tau^\top K_\tau F}{T}\right)\\ &\asymp \lambda_{\min}\left(\frac{F^\top K_\tau^2 F}{T}\right)\asymp\lambda_{\min}(\Sigma_f).\end{aligned}$$

Similarly, we have

$$\lambda_{\max}\left(\frac{\widehat{\mathbb{F}}_\tau^\top K_\tau F}{T}\cdot\frac{\Gamma^\top\Gamma}{N}\cdot\frac{FK_\tau\widehat{\mathbb{F}}_\tau}{T}\right)\asymp\lambda_{\max}\left(\frac{F^\top K_\tau^2 F}{T}\right)\asymp\lambda_{\max}(\Sigma_f).$$

Thus, using (A.10), we conclude that $\sup_{\tau\in[0,1]}\|\widehat{V}_\tau^{-1}\|_2 = O_P(1)$. The proof of Lemma 2.1.3 is now complete. ■

Proof of Lemma A.3:

(1). We now take a look at the term $\frac{1}{N}\sum_{i=1}^N(\widehat{\beta}_{\tau,i}-\widetilde{m}_i(\tau))$, which is an important part when analysing $\widehat{\beta}_{\tau,i}$:

$$\begin{aligned}\widehat{\beta}_{\tau,i} &= \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau Y_i\\ &= \widetilde{m}_i(\tau)+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau(DM_i-\mathbb{Z}\widetilde{m}_i(\tau))\\ &\quad+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}(K_\tau F-\widehat{\mathbb{F}}_\tau\Pi_\tau^{-1})\gamma_i\\ &\quad+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathcal{E}_i\\ &:= \widetilde{m}_i(\tau)+B_{i,1}+B_{i,2}+B_{i,3},\end{aligned}\tag{A.11}$$

where the definitions of $B_{i,1}$, $B_{i,2}$ and $B_{i,3}$ are obvious, $\Pi_\tau = G_N G_\tau\widehat{V}_\tau^{-1}$, $G_N=\frac{\Gamma^\top\Gamma}{N}$, and $G_\tau=\frac{F^\top K_\tau\widehat{\mathbb{F}}_\tau}{T}$. Taking average across i, we obtain

$$\frac{1}{N}\sum_{i=1}^N(\widehat{\beta}_{\tau,i}-\widetilde{m}_i(\tau))=\frac{1}{N}\sum_{i=1}^N B_{i,1}+\frac{1}{N}\sum_{i=1}^N B_{i,2}+\frac{1}{N}\sum_{i=1}^N B_{i,3}.\tag{A.12}$$

Similarly to (A.2), it is easy to know that

$$\sup_{\tau\in[0,1]}\left\|\frac{1}{N}\sum_{i=1}^N B_{i,1}\right\|=O(h^2).$$

Below, we focus on $\frac{1}{N}\sum_{i=1}^N B_{i,2}$. Note that

$$\frac{1}{N}\sum_{i=1}^N B_{i,2} = -\frac{1}{N}\sum_{i=1}^N\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}(\widehat{\mathbb{F}}_\tau\Pi_\tau^{-1}-K_\tau F)\gamma_i$$

$$
\begin{aligned}
&= -\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_1(\tau)\Xi_\tau \overline{\gamma} \\
&\quad - \cdots - \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_5(\tau)\Xi_\tau \overline{\gamma} \\
&\quad - \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_7(\tau)\Xi_\tau \overline{\gamma} \\
&\quad - \cdots - \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_9(\tau)\Xi_\tau \overline{\gamma} \\
&:= -B_{21} \cdots - B_{28},
\end{aligned} \tag{A.13}
$$

where $\overline{\gamma} = \frac{1}{N}\sum_{i=1}^N \gamma_i$, $\Xi_\tau = G_\tau^{-1} G_N^{-1}$, $A_j(\tau)$'s have been defined in the proof of Lemma 2.1.3, and the definitions of B_{2j}'s are obvious. In addition, by the development of Lemma 2.1.3, it is easy to know that

$$
\sup_{\tau \in [0,1]} (\|\Pi_\tau\|_2 + \|\Pi_\tau^{-1}\|_2) = O_P(1).
$$

We now proceed and examine the terms on the right hand side of (A.13). In view of the development of B_{22} below, it is easy to know that B_{21} is negligible. Thus, we start with B_{22}, and write

$$
\begin{aligned}
\frac{1}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_2(\tau)\Pi_\tau^{-1}\overline{\gamma} &= \frac{1}{NT}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \sum_{\ell=1}^N K_\tau (DM_\ell - \mathbb{Z}\widetilde{m}_\ell(\tau))\gamma_\ell^\top G_N^{-1}\overline{\gamma} \\
&\quad + \frac{1}{NT}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \sum_{\ell=1}^N K_\tau \mathbb{Z}(\widetilde{m}_\ell(\tau) - \widehat{\beta}_\ell)\gamma_\ell^\top G_N^{-1}\overline{\gamma} \\
&:= B_{221} + B_{222},
\end{aligned}
$$

where the first equality follows from the definition of Π_τ^{-1}. Note that

$$
\sup_{\tau\in[0,1]} \|B_{221}\|_2 \le O_P(1) \sup_{\tau\in[0,1]} \left\| \frac{1}{NT}\sum_{\ell=1}^N \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau (DM_\ell - \mathbb{Z}\widetilde{m}_\ell(\tau))\gamma_\ell^\top \right\|_2 = O_P(h^2),
$$

where the first inequality follows from $G_N^{-1} = O_P(1)$ and $\overline{\gamma} = O_P(1)$, and the last step follows from the Cauchy-Schwarz inequality and the development similar to (A.3).

The term B_{222} can be rearranged as follows.

$$
\begin{aligned}
&-\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} T B_{222} \\
&= -\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z} \cdot \frac{1}{N}\sum_{\ell=1}^N (\widetilde{m}_\ell(\tau) - \widehat{\beta}_\ell)\gamma_\ell^\top G_N^{-1}\overline{\gamma} \\
&= \frac{1}{N}\sum_{\ell=1}^N (\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau))\gamma_\ell^\top G_N^{-1}\overline{\gamma}.
\end{aligned} \tag{A.14}
$$

For notational simplicity, we suppose that $E[\gamma_\ell] = \gamma$ for all ℓ, so moving (A.14) to the left hand side of (A.12) yields that

$$
\begin{aligned}
&\frac{1}{N}\sum_{\ell=1}^{N}(\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau)) - \frac{1}{N}\sum_{\ell=1}^{N}(\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau))\gamma_\ell^\top G_N^{-1}\overline{\gamma} \\
&= \frac{1}{N}\sum_{\ell=1}^{N}(\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau))(1 - \gamma_\ell^\top G_N^{-1}\overline{\gamma}) \\
&= \frac{1}{N}\sum_{\ell=1}^{N}(\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau))(1 - \gamma^\top G_N^{-1}\gamma) \cdot (1 + o_P(1)),
\end{aligned}
$$

where the second equality follows in an obvious manner.

For B_{23} write,

$$
\begin{aligned}
&\sup_{\tau\in[0,1]} \|B_{23}\|_2 = \sup_{\tau\in[0,1]} \| \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_3(\tau)\Xi_\tau\overline{\gamma}\|_2 \\
&= \sup_{\tau\in[0,1]} \left\| \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} (K_\tau F - \widehat{\mathbb{F}}_\tau \Pi_\tau^{-1}) \frac{1}{NT}\sum_{i=1}^{N} \gamma_i (DM_i - \mathbb{Z}\widehat{\beta}_i)^\top K_\tau \widehat{\mathbb{F}}_\tau \Xi_\tau \overline{\gamma} \right\|_2 \\
&= o_P\left(\frac{1}{\sqrt{T}} \|M_{\widehat{\mathbb{F}}_\tau}(K_\tau F - \widehat{\mathbb{F}}_\tau \Pi_\tau^{-1})\|\right),
\end{aligned}
$$

where the last line follows from Lemma 2.1.2, and the development similar to (A.3).

For B_{24}, write

$$
\begin{aligned}
&\|B_{24}\|_2 = \| \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_4(\tau)\Xi_\tau\overline{\gamma}\|_2 \\
&= \left\| \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \frac{1}{NT}\sum_{i=1}^{N} K_\tau (DM_i - \mathbb{Z}\widehat{\beta}_i)\mathcal{E}_i^\top K_\tau \widehat{\mathbb{F}}_\tau \Xi_\tau \overline{\gamma} \right\|_2 \\
&\leq \left\| \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \frac{1}{NT}\sum_{i=1}^{N} K_\tau (DM_i - \mathbb{Z}\widehat{\beta}_i)\mathcal{E}_i^\top K_\tau^2 F \Pi_\tau \Xi_\tau \overline{\gamma} \right\|_2 \\
&\quad + \left\| \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1} \mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \frac{1}{NT}\sum_{i=1}^{N} K_\tau (DM_i - \mathbb{Z}\widehat{\beta}_i)\mathcal{E}_i^\top K_\tau (\widehat{\mathbb{F}}_\tau - K_\tau F \Pi_\tau) \Xi_\tau \overline{\gamma} \right\|_2 \\
&:= B_{241} + B_{242},
\end{aligned}
$$

where the definitions of B_{241} and B_{242} are obvious. For B_{241}, we can write

$$
\begin{aligned}
&\sup_{\tau\in[0,1]} \left\| \frac{1}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \frac{1}{NT}\sum_{i=1}^{N} K_\tau (DM_i - \mathbb{Z}\widehat{\beta}_i)\mathcal{E}_i^\top K_\tau^2 F \Pi_\tau \Xi_\tau \overline{\gamma} \right\|_2 \\
&\leq O_P(1) \sup_{\tau\in[0,1]} \left\| \frac{1}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} \frac{1}{NT}\sum_{i=1}^{N} K_\tau (DM_i - \mathbb{Z}\widetilde{m}_i(\tau))\mathcal{E}_i^\top K_\tau^2 F \right\|_2 \\
&\quad + O_P(1) \sup_{\tau\in[0,1]} \left\| \frac{1}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z} \frac{1}{NT}\sum_{i=1}^{N} (\widetilde{m}_i(\tau) - \widehat{\beta}_i)\mathcal{E}_i^\top K_\tau^2 F \right\|_2 \\
&\leq o_P(h^2) + O_P(1) \sup_{\tau\in[0,1]} \frac{1}{NT} \|(\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau})^\top \mathcal{E} K_\tau^2 F\|_2
\end{aligned}
$$

$$\leq o_P(h^2) + O_P(1) \sup_{\tau\in[0,1]} \frac{1}{\sqrt{N}}\|\widehat{\mathbb{B}} - \mathbb{B}_{0,\tau}\| \frac{\sqrt{N}\vee\sqrt{T}}{\sqrt{NTh}}$$
$$= o_P\left(h^2\right) + O_P\left(\frac{1}{(N\wedge T)h}\right),$$

where the term $o_P(h^2)$ follows from a development similar to (A.3) and Assumption 2.1, the third inequality follows from $\|\mathcal{E}\|_2 = O_P(\sqrt{N}\vee\sqrt{T})$ and $\sup_{\tau\in[0,1]}\frac{1}{\sqrt{T}}\|K_\tau F\|_2 = O_P(1)$ of Assumption 1, and the last equality follows from (A.7). Thus, we conclude that

$$\sup_{\tau\in[0,1]} B_{241} = o_P\left(h^2\right) + O_P\left(\frac{1}{(N\wedge T)h}\right).$$

Similarly, we can show that $\sup_{\tau\in[0,1]} B_{242} = o_P(h^2)$ in view of the fact that $\frac{1}{\sqrt{T}}\|\widehat{\mathbb{F}}_\tau - K_\tau F\Pi_\tau\|_2 = o_P(1)$. Thus,

$$\sup_{\tau\in[0,1]} \|B_{24}\|_2 = o_P\left(h^2\right) + O_P\left(\frac{1}{(N\wedge T)h}\right).$$

Following the development similar to B_{24} and B_{23} respectively, we can show that

$$\sup_{\tau\in[0,1]} \|B_{25}\|_2 = o_P\left(h^2\right) + O_P\left(\frac{1}{(N\wedge T)h}\right),$$
$$\sup_{\tau\in[0,1]} \|B_{26}\|_2 = o_P\left(\frac{1}{\sqrt{T}}\|M_{\widehat{\mathbb{F}}_\tau}(K_\tau F - \widehat{\mathbb{F}}_\tau\Pi_\tau^{-1})\|\right).$$

We will consider B_{27} later together with $\frac{1}{N}\sum_{i=1}^N B_{i,3}$. Thus, we now move on to B_{28}.

$$\begin{aligned}
&\sup_{\tau\in[0,1]} \frac{1}{T}\|\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_9(\tau)\Xi_\tau\overline{\gamma}\|_2 \\
&\leq O_P(1) \sup_{\tau\in[0,1]} \frac{1}{T}\left\|\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\frac{1}{NT}\sum_{i=1}^N K_\tau\mathcal{E}_i\mathcal{E}_i^\top K_\tau\widehat{\mathbb{F}}_\tau\right\|_2 \\
&\leq O_P(1)\frac{1}{T\sqrt{h}}\cdot\frac{1}{N}\|\mathcal{E}\|_2^2\cdot\sup_{\tau\in[0,1]}\sqrt{\frac{h}{T}\lambda_{\max}\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau^2 M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)}\sqrt{\frac{1}{T}\lambda_{\max}(\widehat{\mathbb{F}}_\tau^\top K_\tau^2\widehat{\mathbb{F}}_\tau)} \\
&\leq O_P(1)\frac{1}{Th}\cdot\frac{1}{N}\|\mathcal{E}\|_2^2\cdot\sup_{\tau\in[0,1]}\sqrt{\frac{1}{T}\lambda_{\max}(\mathbb{Z}^\top K_\tau^2\mathbb{Z})}\sqrt{\frac{1}{T}\lambda_{\max}(\widehat{\mathbb{F}}_\tau^\top\widehat{\mathbb{F}}_\tau)} \\
&= O_P\left(\frac{1}{(N\wedge T)h}\right), \qquad\qquad (A.15)
\end{aligned}$$

where the last line follows from $\|\mathcal{E}\|_2 = O_P(\sqrt{N}\vee\sqrt{T})$ of Assumption 1. Thus, we have

$$\sup_{\tau\in[0,1]} \|B_{28}\|_2 = O_P\left(\frac{1}{(N\wedge T)h}\right).$$

Collecting the above results, we obtain

$$
\begin{aligned}
&\frac{1}{N}\sum_{\ell=1}^{N}(\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau))(1-\gamma^\top G_N^{-1}\gamma)\\
&= \frac{1}{N}\sum_{i=1}^{N} B_{i,3} + O_P\left(h^2 + \frac{1}{(N\wedge T)h}\right) - B_{27},
\end{aligned}
$$

where

$$
\begin{aligned}
B_{27} &= \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\frac{1}{NT}\sum_{i=1}^{N} K_\tau \mathcal{E}_i \gamma_i^\top F^\top K_\tau \widehat{\mathbb{F}}_\tau \Xi_\tau \overline{\gamma}\\
&= \left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\frac{1}{N}\sum_{i=1}^{N} K_\tau \mathcal{E}_i \gamma_i^\top G_N^{-1}\overline{\gamma}.
\end{aligned}
$$

We now invoke Assumption 2.2 to obtain the following result

$$
\sup_{\tau\in[0,1]}\left\|\frac{1}{N}\sum_{\ell=1}^{N}(\widehat{\beta}_\ell - \widetilde{m}_\ell(\tau))\right\|_2 = O_P\left(h^2 + \frac{\sqrt{\log(NT)}}{\sqrt{NTh}}\right).
$$

The proof of the first result is now completed.

(2). Consider $\widehat{\beta}_i$, and recall that we have decomposed it in (A.11). Similar to (A.3), we have

$$
\sup_{\tau\in[0,1]}\|B_{i,1}\| = O(h^2).
$$

We focus on $B_{i,2}$ below, and recall that Ξ_τ and Π_τ have been defined in the proof of this lemma.

$$
\begin{aligned}
B_{i,2} &= -\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}(\widehat{\mathbb{F}}_\tau \Pi_\tau^{-1} - K_\tau F)\gamma_i\\
&= -\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}(A_1(\tau)+\cdots+A_5(\tau)+A_7(\tau)+\cdots+A_9(\tau))\Xi_\tau\gamma_i\\
&:= -B_{i,21} - \cdots - B_{i,28}.
\end{aligned}
$$

In view of the development of $B_{i,22}$ below, it is easy to know that $B_{i,21}$ is negligible. Thus, we start our development with $B_{i,22}$, and write

$$
\begin{aligned}
&\frac{1}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} A_2(\tau)\Xi_\tau\gamma_i\\
&= \frac{1}{NT^2}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\sum_{\ell=1}^{N} K_\tau (DM_\ell - \mathbb{Z}\widetilde{m}_\ell(\tau))(F\gamma_\ell)^\top K_\tau \widehat{\mathbb{F}}_\tau \Xi_\tau \gamma_i\\
&\quad + \frac{1}{NT^2}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\sum_{\ell=1}^{N} K_\tau \mathbb{Z}(\widetilde{m}_\ell(\tau) - \widehat{\beta}_\ell)(F\gamma_\ell)^\top K_\tau \widehat{\mathbb{F}}_\tau \Xi_\tau \gamma_i\\
&= \frac{1}{NT^2}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\sum_{\ell=1}^{N} K_\tau (DM_\ell - \mathbb{Z}\widetilde{m}_\ell(\tau))\gamma_\ell^\top G_N^{-1}\gamma_i
\end{aligned}
$$

$$+\frac{1}{NT^2}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\sum_{\ell=1}^N K_\tau \mathbb{Z}(\widetilde{m}_\ell(\tau)-\widehat{\beta}_\ell)\gamma_\ell^\top G_N^{-1}\gamma_i$$
$$:= B_{i,221}+B_{i,222}.$$

Note that

$$\begin{aligned}\|B_{i,222}\| &\leq \left\|\frac{1}{NT^2}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\sum_{\ell=1}^N K_\tau \mathbb{Z}(\widetilde{m}_\ell(\tau)-\widehat{\beta}_\ell)(\gamma_\ell - E[\gamma_\ell])^\top G_N^{-1}\gamma_i\right\|_2 \\ &+ \left\|\frac{1}{NT^2}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}\sum_{\ell=1}^N K_\tau \mathbb{Z}(\widetilde{m}_\ell(\tau)-\widehat{\beta}_\ell)E[\gamma_\ell]^\top G_N^{-1}\gamma_i\right\|_2 \\ &= O_P\left(h^2+\frac{\sqrt{\log(NT)}}{\sqrt{NTh}}\right),\end{aligned}$$

where the equality follows from the first result of this lemma.

For $B_{i,23}$-$B_{i,26}$, using procedures similar to those given for B_{23}-B_{26} in the proof of the first result of this lemma, we know that they are negligible. Also, following the proof of (A.15), we know that

$$\sup_{\tau\in[0,1]}\|B_{i,28}\|_2 = O_P\left(\frac{1}{Th}\vee\frac{1}{Nh}\right).$$

Collecting the above results, we obtain that

$$\sup_{\tau\in[0,1]}\|\widehat{\beta}_i-\widetilde{m}_i(\tau)-B_{i,3}\| = O_P\left(h^2+\frac{\sqrt{\log(NT)}}{\sqrt{NTh}}\right),$$

which completes the proof. ■

Proof of Theorem 2.1:

By the second result of Lemma A.3, the result follows from Assumption 1.3 and Assumption 2.3. ■

Proof of Theorem 2.2:

Write

$$\begin{aligned}\widehat{\beta}^*_{\tau,i} &= \widehat{\beta}_{\tau,i}+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau(\widehat{\mathbb{U}}_i\circ\xi) \\ &= \widehat{\beta}_{\tau,i}+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau(\mathcal{E}_i\circ\xi) \\ &\quad+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau[D(M_i-\widehat{M}_i)\circ\xi] \\ &\quad+\left(\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau\mathbb{Z}\right)^{-1}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau}K_\tau[(F\gamma_i)\circ\xi] \\ &:= \widehat{\beta}_{\tau,i}+B_1+B_2+B_3,\end{aligned}\tag{A.16}$$

where $\widehat{M}_i$ is the estimated version of M_i, and the definitions of B_1, B_2 and B_3 are obvious.

Next, we focus on the terms on the right hand side of (A.16). We start with B_1, and write

$$
\begin{aligned}
& E^*\left[\left(\frac{\sqrt{Th}}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau(\mathcal{E}_i\circ\xi)\right)\left(\frac{\sqrt{Th}}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau(\mathcal{E}_i\circ\xi)\right)^\top\right] \\
&= \frac{h}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau E^*[(\mathcal{E}_i\circ\xi)(\mathcal{E}_i\circ\xi)^\top] K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z} \\
&= \frac{h}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau(\mathcal{E}_i\mathcal{E}_i^\top) K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z} \\
&\quad + \frac{h}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau[(\mathcal{E}_i\mathcal{E}_i^\top)\circ(E[\xi\xi^\top]-1_T1_T^\top)] K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z},
\end{aligned}
$$

where the first term of the right hand is exactly the leading term when deriving the asymptotic distribution in Theorem 2.1. Below, we show

$$
\left|\frac{h}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau[(\mathcal{E}_i\mathcal{E}_i^\top)\circ(E[\xi\xi^\top]-1_T1_T^\top)] K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z}\right| = o_P(1), \tag{A.17}
$$

so we can conclude that

$$
\sqrt{Th}B_1\cdot\sqrt{Th}B_1^\top \to_P \Sigma_{1,\tau}^{-1}\Sigma_{2i,\tau}\Sigma_{1,\tau}^{-1} \tag{A.18}
$$

in connection with the facts that $E[f_t]=0$, and $\{f_t\}$ and $\{\varepsilon_{it}\}$ are independent.

First, we write

$$
\begin{aligned}
& \frac{h}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau[(\mathcal{E}_i\mathcal{E}_i^\top)\circ(E[\xi\xi^\top]-1_T1_T^\top)] K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z} \\
&= \frac{h}{T}\mathbb{Z}^\top K_\tau^2[(\mathcal{E}_i\mathcal{E}_i^\top)\circ(E[\xi\xi^\top]-1_T1_T^\top)] K_\tau^2 \mathbb{Z} + o_P(1),
\end{aligned}
$$

in which the equality follows from the facts that $E[f_t]=0$, and $\{f_t\}$ and $\{\varepsilon_{it}\}$ are independent. Also, note that

$$
\begin{aligned}
& E\left|\frac{h}{T}\mathbb{Z}^\top K_\tau^2[(\mathcal{E}_i\mathcal{E}_i^\top)\circ(E[\xi\xi^\top]-1_T1_T^\top)] K_\tau^2 \mathbb{Z}\right| \\
&= E\left|\frac{1}{Th}\sum_{t=1}^{d_T}\sum_{s=1}^{T-t} z_t z_{s+t} K_h(\tau_s-\tau)K_h(\tau_{s+t}-\tau)\varepsilon_{is}\varepsilon_{i,s+t}\left[a\left(\frac{t}{\ell}\right)-1\right]\right| \\
&\quad + E\left|\frac{1}{Th}\sum_{t=d_T+1}^{T}\sum_{s=1}^{T-t} z_t z_{s+t} K_h(\tau_s-\tau)K_h(\tau_{s+t}-\tau)\varepsilon_{is}\varepsilon_{i,s+t}\left[a\left(\frac{t}{\ell}\right)-1\right]\right| \\
&\le O(1)\sum_{t=1}^{d_T}\left|a\left(\frac{t}{\ell}\right)-a(0)\right| + O(1)\max_{i\ge 1}\sum_{t=d_T+1}^{T} E|\varepsilon_{i0}\varepsilon_{i,t}| \\
&\le O(1)\frac{d_T^2}{\ell} + O(1)\max_{i\ge 1}\sum_{t=d_T+1}^{T} E|\varepsilon_{i0}\varepsilon_{i,t}| = o(1),
\end{aligned}
$$

where the inequality follows from $a(w)$ being Lipschitz continuous on $[-1,1]$, and the last equality holds by letting $d_T^2/\ell \to 0$ and $d_T \to \infty$. Therefore, we have proved (A.17), which further yields (A.18).

Similarly, we can show that

$$
\begin{aligned}
&E^*\left[\left(\frac{\sqrt{Th}}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau (F\gamma_i \circ \xi)\right)\left(\frac{\sqrt{Th}}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau (F\gamma_i \circ \xi)\right)^\top\right] \\
&= \frac{h}{T}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau (F\gamma_i\gamma_i^\top F^\top) K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \mathbb{Z} + o_P(1) = o_P(1).
\end{aligned}
$$

Now, we consider B_2, and write

$$
\begin{aligned}
&E^*\left\|\frac{1}{\sqrt{Th}}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau [D(M_i - \widehat{M}_i)\circ \xi]\right\|^2 \\
&\le \max_{\lfloor T(\tau-h)\rfloor \le t \le \lfloor T(\tau+h)\rfloor} \|\Delta_{M,it}\|^2 E^*\left\|\frac{1}{\sqrt{Th}}\mathbb{Z}^\top K_\tau M_{\widehat{\mathbb{F}}_\tau} K_\tau \xi\right\|^2 = O_P\left((h^4 + \frac{1}{Th})\cdot \ell\right) = o_P(1),
\end{aligned}
$$

where $\Delta_{M,it}$ stands for the t^{th} row of $D(M_i - \widehat{M}_i)$, and the first equality follows from Theorem 2.1 and the construction of ξ, and the last lines follows from $\frac{\ell}{\sqrt{Th}} \to 0$ and $Th^5 \to 0$.

To conclude the result, all we need is to show that

$$
B_1^* \equiv \frac{1}{\sqrt{Th}}\mathbb{Z}^\top K_\tau^2 (\mathcal{E}_i \circ \xi) \to_{D^*} N(0, \Sigma_{2i,\tau}),
$$

which can be done by verifying the Lindeberg condition using the large-block and the small-block technique. Using the Cramér-Wold device, let η be a $2(J+1)\times 1$ vector and $\|\eta\| = 1$. Thus, for $\forall i$, we consider

$$
\widetilde{B}_1^* \equiv \eta^\top B_1^* = \frac{\sqrt{h}}{\sqrt{T}}\eta^\top \mathbb{Z}^\top K_\tau^2(\mathcal{E}_i \circ \xi). \tag{A.19}
$$

The goal is to show that

$$
\widetilde{B}_1^* \to_{D^*} N(0, \eta^\top \Sigma_{2i,\tau}\eta), \tag{A.20}
$$

which in connection with Assumption 2.3 immediately yields the result.

We now rewrite $\widetilde{B}_1^*$ as follows.

$$
\widetilde{B}_1^* = \sum_{j=1}^K \nu_j^* + \sum_{j=1}^K \varpi_j^*, \tag{A.21}
$$

where

$$\nu_j^* = \sum_{t=B_j+1}^{B_j+r_1} \frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it}\xi_t K_h(\tau_t - \tau) \quad \text{and} \quad \varpi_j^* = \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2} \frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it}\xi_t K_h(\tau_t - \tau).$$

where $z_{\eta,t} \equiv \eta^\top z_t$. Moreover, $B_j = (j-1)(r_1+r_2)$, and without loss of generality we suppose that $K = T/(r_1+r_2)$ is an integer for simplicity. Otherwise, one needs to include the remaining terms in (A.21) which are negligible for an obvious reason. In addition, we let

$$(r_1, r_2) \to (\infty, \infty), \quad \left(\frac{r_2}{r_1}, \frac{r_1}{Th}\right) \to (0,0), \quad r_1 \geq \ell, \tag{A.22}$$

so the blocks ϖ_j^*'s are mutually independent by the construction of ξ_t's. Note that by $\frac{r_2}{r_1} \to 0$ of (A.22),

$$\frac{Kr_2}{T} \to 0 \quad \text{and} \quad \frac{Kr_1}{T} \to 1. \tag{A.23}$$

We now write

$$\begin{aligned}
&EE^*\left[\left(\sum_{j=1}^K \varpi_j^*\right)^2\right] = \sum_{j=1}^K EE^*[(\varpi_j^*)^2] \\
&\leq \frac{h}{T}\sum_{s=-r_2+1}^{r_2-1} a\left(\frac{s}{\ell}\right) \cdot |z_{\eta,t}z_{\eta,t+s}| \cdot E|\varepsilon_{it}\varepsilon_{i,t+s}| \sum_{j=1}^K \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-|s|} K_h(\tau_t-\tau)K_h(\tau_{t+|s|}-\tau) \\
&\leq O(1)\frac{1}{Th}\max_{0\leq s\leq r_2-1}\sum_{j=1}^K \sum_{t=B_j+r_1+1}^{B_j+r_1+r_2-|s|} K\left(\frac{\tau_t-\tau}{h}\right)K\left(\frac{\tau_{t+|s|}-\tau}{h}\right) \\
&\leq O(1)\frac{Kr_2}{T} = o(1),
\end{aligned}$$

where the second inequality follows from $a(\cdot)$ being bounded on $[-1,1]$, and Assumption 2.1. Therefore, the term $\sum_{j=1}^K \varpi_j^*$ of (A.21) is negligible.

Next, we employ the Lindeberg CLT to establish the asymptotic normality of $\sum_{j=1}^K \nu_j^*$. Recall that we have shown that $(\widetilde{B}_1^*)^2 = \eta^\top\Sigma_{2i,\tau}\eta + o_P(1)$ and $\sum_{j=1}^K \varpi_j^*$ of (A.21) is negligible, so it is easy to know that

$$E^*(\sum_{j=1}^K \nu_j^*)^2 = \eta^\top\Sigma_{2i,\tau}\eta + o_P(1).$$

Similar arguments can be seen in (A.8)-(A.9) of Chen et al. (2012a). That said, we just need to verify that for $\forall\epsilon > 0$

$$\sum_{j=1}^K E^*\left[(\nu_j^*)^2 \cdot I\left(|\nu_j^*| > \epsilon\right)\right] = o_P(1). \tag{A.24}$$

Before proceeding further, we point out that the series $\frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it}\xi_t K_h(\tau_t - \tau)$ is in fact a mixingale sequence mentioned in Definition 1 of Hansen (1991), where the term $|\frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it} K_h(\tau_t - \tau)|$ is equivalent to c_i in the notation of Hansen (1991). This is not hard to justify given $\{\xi_t\}$ is an ℓ-dependent series. When m in the notation of Hansen (1991) is greater than ℓ, all the requirements of Definition 1 of Hansen (1991) are fulfilled. Thus, it allows us to invoke the asymptotic properties associated to the mixingale sequence (i.e., Lemma A.2 of this paper) in the following development.

Write

$$
\begin{aligned}
&\sum_{j=1}^{K} E^*[(\nu_j^*)^2 \cdot I(|\nu_j^*| > \epsilon)] \\
&\le \sum_{j=1}^{K} \{E^*|(\nu_j^*)^2|^{\delta/2}\}^{2/\delta} \cdot \{E^*[I(|\nu_j^*| > \epsilon)]\}^{(\delta-2)/\delta} \le \sum_{j=1}^{K} \{E^*|(\nu_j^*)^2|^{\delta/2}\}^{2/\delta} \left\{ \frac{E^*|\nu_j^*|^\delta}{\epsilon^\delta} \right\}^{(\delta-2)/\delta} \\
&= \epsilon^{\delta-2} \sum_{j=1}^{K} E^*|\nu_j^*|^\delta = \epsilon^{\delta-2} \sum_{j=1}^{K} \left\{ E^* \left[\left(\sum_{t=B_j+1}^{B_j+r_1} \frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it}\xi_t K_h(\tau_t - \tau) \right)^\delta \right] \right\}^{\frac{1}{\delta}\cdot\delta} \\
&\le O(1)\epsilon^{\delta-2} \sum_{j=1}^{K} \left\{ \sum_{t=B_j+1}^{B_j+r_1} \left(\frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it} K_h(\tau_t - \tau) \right)^2 \right\}^{\frac{1}{2}\cdot\delta} \\
&\le O(1)\epsilon^{\delta-2} \sum_{j=1}^{K} r_1^{\delta/2-1} \sum_{t=B_j+1}^{B_j+r_1} \left(\frac{\sqrt{h}}{\sqrt{T}} z_{\eta,t}\varepsilon_{it} K_h(\tau_t - \tau) \right)^\delta \\
&\le O(1)\epsilon^{\delta-2} \frac{r_1^{\delta/2-1}}{(Th)^{\delta/2-1}} \cdot \frac{1}{Th} \sum_{t=1}^{T} \left(z_{\eta,t}\varepsilon_{it} K\left(\frac{\tau_t - \tau}{h} \right) \right)^\delta \\
&= O(1) \frac{r_1^{\delta/2-1}}{(Th)^{\delta/2-1}} = o(1), \qquad \text{(A.25)}
\end{aligned}
$$

where the first inequality follows from the Hölder inequality, the second inequality follows from the Chebyshev's inequality, the third inequality follows from Lemma 2 of Hansen (1991), and the last equality follows from $r_1/(Th) \to 0$ and $\delta > 2$. Thus, we can conclude the validity of (A.24).

Based on the above development, we are readily to conclude that (A.20) holds. The proof is now completed. ■

Proof of Corollary 2.2:

The proof follows from a procedure very much similar to (A.7), we thus omit the details. ■

Proof of Theorem 2.3:

Similar to (A.10), we can show that

$$\sup_{\tau\in[0,1]}\left\|\widetilde{V}_\tau-\frac{\widetilde{\mathbb{F}}_\tau^\top K_\tau F}{T}\cdot\frac{\Gamma^\top\Gamma}{N}\cdot\frac{FK_\tau\widetilde{\mathbb{F}}_\tau}{T}\right\|_2$$
$$=O_P\left(\frac{1}{\sqrt{N}}\|\widetilde{\mathbb{B}}-\mathbb{B}_{0,\tau}\|+h^2+\frac{1}{\sqrt{(N\wedge T)h}}\right)$$
$$=O_P\left(h^2+\frac{1}{\sqrt{(N\wedge T)h}}\right),$$

where $\widetilde{V}_\tau=\mathrm{diag}\{\lambda_{\tau,1},\ldots,\lambda_{\tau,d_{\max}}\}$, and the second equality follows from Corollary 2.2.

To proceed further, we define a few notations. Decompose $\widetilde{\mathbb{F}}_\tau$ as $\widetilde{\mathbb{F}}_\tau=(\widetilde{\mathbb{F}}_{\tau,1},\ldots,\widetilde{\mathbb{F}}_{\tau,d_{\max}})$. Let $\widetilde{\Pi}_\ell$ be the ℓ^{th} column of $\frac{\Gamma^\top\Gamma}{N}\cdot\frac{FK_\tau\widetilde{\mathbb{F}}_\tau}{T}\cdot\widetilde{V}_{\tau,1:d_f}^{-1}$, where $\widetilde{V}_{\tau,1:d_f}$ is the leading $d_f\times d_f$ principal sub-matrix of $\widetilde{V}_\tau$. Further we let

$$\Sigma_\tau=\frac{1}{NT}K_\tau F\Gamma^\top\Gamma F^\top K_\tau.$$

In what follows, we consider $\widetilde{\mathbb{F}}_{\tau,\ell}$ for $\ell\in[d_f]$ first. Following the proof similar to Lemma 2.1.3, we obtain that for $\ell\in[d_f]$

$$\sup_{\tau\in[0,1]}\frac{1}{\sqrt{T}}\|\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell}\|=O_P\left(h^2+\frac{1}{\sqrt{(N\wedge T)h}}\right).$$

For $\ell\in[d_f]$, we are now ready to write

$$\sup_{\tau\in[0,1]}\left|\lambda_{\tau,\ell}-\lambda^*_{\tau,\ell}\right|=\sup_{\tau\in[0,1]}\left|\frac{1}{T}\widetilde{\mathbb{F}}_{\tau,\ell}^\top K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau\widetilde{\mathbb{F}}_{\tau,\ell}-\frac{1}{T}(K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top\Sigma_\tau K_\tau F\widetilde{\Pi}_{\tau,\ell}\right|$$
$$=\sup_{\tau\in[0,1]}\frac{1}{T}|(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell}+K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top(K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau-\Sigma_\tau+\Sigma_\tau)(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell}+K_\tau F\widetilde{\Pi}_{\tau,\ell})$$
$$-(K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top\Sigma_\tau K_\tau F\widetilde{\Pi}_{\tau,\ell}|$$
$$=\sup_{\tau\in[0,1]}\frac{1}{T}|(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top(K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau-\Sigma_\tau)(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell})|$$
$$+\sup_{\tau\in[0,1]}\frac{2}{T}|(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top(K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau-\Sigma_\tau)K_\tau F\widetilde{\Pi}_{\tau,\ell}|$$
$$+\sup_{\tau\in[0,1]}\frac{1}{T}|(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top\Sigma_\tau(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell})|$$
$$+\sup_{\tau\in[0,1]}\frac{2}{T}|(\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top\Sigma_\tau K_\tau F\widetilde{\Pi}_{\tau,\ell}|$$
$$+\sup_{\tau\in[0,1]}\frac{1}{T}|(K_\tau F\widetilde{\Pi}_{\tau,\ell})^\top(K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau-\Sigma_\tau)K_\tau F\widetilde{\Pi}_{\tau,\ell}|$$
$$:=A_1+2A_2+A_3+2A_4+A_5.$$

Note that $|A_1|=o_P(|A_5|)$, $|A_2|=o_P(|A_5|)$, and $|A_3|=o_P(|A_4|)$. Thus, we focus on the leading terms A_4 and A_5 below. Simple algebra shows that

$$|A_4|\le\sup_{\tau\in[0,1]}\frac{1}{\sqrt{T}}\|\widetilde{\mathbb{F}}_{\tau,\ell}-K_\tau F\widetilde{\Pi}_{\tau,\ell}\|_2\cdot\frac{1}{\sqrt{T}}\|\Sigma_\tau K_\tau F\widetilde{\Pi}_{\tau,\ell}\|_2=O_P\left(h^2+\frac{1}{\sqrt{(N\wedge T)h}}\right),$$

$$|A_5| = O_P(1) \sup_{\tau\in[0,1]} \|K_\tau \widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau - \Sigma_\tau\| = O_P\left(h^2 + \frac{1}{\sqrt{(N\wedge T)h}}\right),$$

where the development of A_4 and A_5 follow from (A.10).

Next, we consider $\lambda_{\tau,\ell}$ for $\ell = d_f+1,\ldots,T$. Let F^* be a $T\times(T-d_f)$ matrix such that

$$\frac{1}{T}(F^*, K_\tau F\mathfrak{D})^\top (F^*, K_\tau F\mathfrak{D}) = \begin{pmatrix} I_{T-r} & 0 \\ 0 & I_r \end{pmatrix},$$

where $\mathfrak{D}$ is an $d_f\times d_f$ rotation matrix such that $\frac{1}{T}\mathfrak{D}^\top F^\top K_\tau^2 F\mathfrak{D} = I_{d_f}$. Additionally, let

$$K_\tau \widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau = \Sigma_\tau + (K_\tau \widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau - \Sigma_\tau) := \Sigma_\tau + \Delta\Sigma_\tau.$$

Having introduced the above variables, we are now ready to proceed further. Note that F^*, $K_\tau F\mathfrak{D}$, Σ_τ and $\Delta\Sigma_\tau$ are corresponding to Q_1, Q_2, A and E of Lemma A.1. Thus, using Lemma A.1, we obtain that

$$\widetilde{F}^* := \frac{1}{\sqrt{T}}(F^* + K_\tau F\mathfrak{D}P)(I_{T-d_f} + P^\top P)^{-1/2},$$

which is corresponding Q_1^0 of Lemma A.1. Moreover,

$$\begin{aligned}
\sup_{\tau\in[0,1]} \|P\|_2 &\le \sup_{\tau\in[0,1]} \frac{4}{\mathrm{sep}(0, \frac{1}{T}F^\top K_\tau\Sigma_\tau K_\tau F)} \cdot \|\Delta\Sigma_\tau\| \le O_P(1) \sup_{\tau\in[0,1]} \|\Delta\Sigma_\tau\| \\
&= O_P\left(h^2 + \frac{1}{\sqrt{(N\wedge T)h}}\right).
\end{aligned}$$

Since $\widetilde{F}^*$ is an orthonormal basis for a subspace that is invariant for $\Sigma_\tau + \Delta\Sigma_\tau = K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau$, studying λ_ℓ for $\ell = d_f+1,\ldots,T$ is equivalent to investigating $\widetilde{F}^*$. Then we write

$$\begin{aligned}
&\sup_{\tau\in[0,1]} \left\|\widetilde{F}^* - \frac{1}{\sqrt{T}}F^*\right\|_2 \\
&= \sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}} \|[F^* + K_\tau F\mathfrak{D}P - F^*(I_{T-d_f} + P^\top P)^{1/2}](I_{T-d_f} + P^\top P)^{-1/2}\|_2 \\
\le &\sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}} \|F^*(I_{T-d_f} - (I_{T-d_f} + P^\top P)^{1/2})(I_{T-d_f} + P^\top P)^{-1/2}\|_2 \\
&+ \sup_{\tau\in[0,1]} \frac{1}{\sqrt{T}} \|K_\tau F\mathfrak{D}P(I_{T-d_f} + P^\top P)^{-1/2}\|_2 \\
\le &\sup_{\tau\in[0,1]} \|(I_{T-d_f} - (I_{T-d_f} + P^\top P)^{1/2})(I_{T-d_f} + P^\top P)^{-1/2}\|_2 \\
&+ \sup_{\tau\in[0,1]} \|P(I_{T-d_f} + P^\top P)^{-1/2}\|_2 \\
\le &\sup_{\delta\in[0,1]} \|I_{T-d_f} - (I_{T-d_f} + P^\top P)^{1/2}\|_2 + \sup_{\tau\in[0,1]} \|P\|_2 \le 2\sup_{\tau\in[0,1]} \|P\|_2
\end{aligned}$$

$$= O_P\left(h^2 + \frac{1}{\sqrt{Nh}} \vee \frac{1}{\sqrt{Th}}\right), \tag{A.26}$$

where the third and fourth inequalities follow from Exercise 1 of page 231 of Magnus & Neudecker (2007).

Thus for $\ell = 1, \ldots, T - d_f$, we write

$$\begin{aligned}
&\sup_{\tau\in[0,1]} |\lambda_{\tau,d_f+\ell}| = \sup_{\delta\in[0,1]} \widetilde{F}_\ell^{*\top} K_\tau \widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau) K_\tau \widetilde{F}_\ell^* \\
&= \sup_{\tau\in[0,1]} \left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^* + \frac{1}{\sqrt{T}}F^*\right)^\top (K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau - \Sigma_\tau + \Sigma_\tau) \left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^* + \frac{1}{\sqrt{T}}F^*\right) \\
&= \sup_{\tau\in[0,1]} \left|\left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^*\right)^\top (K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau - \Sigma_\tau) \left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^*\right)\right| \\
&\quad +2 \sup_{\tau\in[0,1]} \left|\left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^*\right)^\top (K_\tau\widehat{\Sigma}(\widetilde{\mathbb{B}}_\tau)K_\tau - \Sigma_\tau) \cdot \frac{1}{\sqrt{T}}F^*\right| \\
&\quad + \sup_{\delta\in[0,1]} \left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^*\right)^\top \Sigma_\tau \left(\widetilde{F}^* - \frac{1}{\sqrt{T}}F^*\right) = O_P\left(h^4 + \frac{1}{\sqrt{(N\wedge T)h}}\right),
\end{aligned} \tag{A.27}$$

where $\widetilde{F}_\ell^*$ and F_ℓ^* stand for the ℓ^{th} columns of $\widetilde{F}^*$ and F^* respectively, and the last equality follows from (A.26) and the development of (A.10).

Up to this point, we can conclude that

$$\begin{aligned}
&\sup_{\tau\in[0,1]} \frac{\lambda_{\tau,1}}{\lambda_{\tau,0}} \asymp 1, \\
&\sup_{\tau\in[0,1]} \frac{\lambda_{\tau,\ell+1}}{\lambda_{\tau,\ell}} \asymp 1 \text{ for } \ell = 1, \ldots, d_f - 1, \\
&\sup_{\delta\in[0,1]} \frac{\lambda_{\tau,d_f+1}}{\lambda_{\tau,d_f}} = O_P\left(h^4 + \frac{1}{(N\wedge T)h}\right),
\end{aligned}$$

which in connection with (A.27) and the construction of ϵ_{NT} immediately yields the result. The proof is now completed. ■