Multiscale Testing for Equality of Nonparametric Trend Curves

Marina Khismatullina[1] Michael Vogt[2]
University of Bonn University of Bonn

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

Key words: Multiscale statistics; nonparametric regression; time series errors; shape constraints; strong approximations; anti-concentration bounds.
AMS 2010 subject classifications: 62E20; 62G10; 62G20; 62M10.

1 The model

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

$$Y_{it} = \beta' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{1.1}$$

for $1 \le t \le T$, where β is a $d \times 1$ vector of unknown parameters, X_{it} is a $d \times 1$ vector of individual covariates, m_i is an unknown nonparametric trend function defined on $[0,1]$, α_i is a (deterministic or random) intercept term and $\mathcal{E}_i = \{\varepsilon_{it} : 1 \le t \le T\}$ is a zero-mean stationary error process. For identification, we normalize the functions m_i such that $\int_0^1 m_i(u) du = 0$ for all $1 \le i \le n$. The term α_i can also be regarded as an additional error component. In the econometrics literature, it is commonly called a fixed effect error term. It can be interpreted as capturing unobserved characteristics of the time series $\mathcal{Y}_i$ which remain constant over time. We allow the error terms α_i to be dependent across i in an arbitrary way. Hence, by including them in model equation (1.1), we allow the n time series $\mathcal{Y}_i$ in our sample to be correlated with each other.

[1]Address: Bonn Graduate School of Economics, University of Bonn, 53113 Bonn, Germany. Email: marina.k@uni-bonn.de.
[2]Corresponding author. Address: Department of Economics and Hausdorff Center for Mathematics, University of Bonn, 53113 Bonn, Germany. Email: michael.vogt@uni-bonn.de.

Whereas the terms α_i may be correlated, the error processes $\mathcal{E}_i$ are assumed to be independent across i. In addition, each process $\mathcal{E}_i$ is supposed to satisfy the conditions (C1)–(C3). Finally note that throughout the paper, we restrict attention to the case where the number of time series n in model (??) is fixed. Extending our theoretical results to the case where n slowly grows with the sample size T is a possible topic for further research.

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

(C1) For each i the variables ε_{it} allow for the representation $\varepsilon_{it} = G_i(\ldots, \eta_{it-1}, \eta_{it}, \eta_{it+1}, \ldots)$, where η_{it} are i.i.d. random variables across t and $G_i : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. Denote $\mathcal{Z}_{it} = (\ldots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$.

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

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

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

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

2 Testing for equality of time trends

In this section, we adapt the multiscale method developed in Section ?? to test the hypothesis that the trend functions in model (??) are all the same. More formally, we test the null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ in model (1.1). As we will see, the proposed multiscale method does not only allow to test whether the null hypothesis is violated. It also provides information on where violations occur. More specifically, it allows to identify, with a pre-specified confidence, (i) trend functions which are different from each other and (ii) time intervals where these trend functions differ.

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

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

$$Y_{it} = \beta_i' X_{it} + m_i\Big(\frac{t}{T}\Big) + \alpha_i + \varepsilon_{it} \tag{2.1}$$

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

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

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

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

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

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

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

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

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

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

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

$$\Delta Y_{it} = \beta_i' \Delta X_{it} + \Delta \varepsilon_{it} + O\Big(\frac{1}{T}\Big). \tag{2.2}$$

In particular, for each i we employ the least squares estimation method to estimate $\boldsymbol{\beta}_i$ in (2.2), treating $\Delta \mathbf{X}_{it}$ as the regressors and ΔY_{it} as the response variable. That is, we propose the following differencing estimator:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=1}^{T} \Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}^T \Big)^{-1} \sum_{t=1}^{T} \Delta \mathbf{X}_{it} \Delta Y_{it} \tag{2.3}$$

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

Theorem 2.1. *Under Assumptions (C1) - (C6), we have*

$$\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i = O_P\Big(\frac{1}{\sqrt{T}}\Big),$$

where $\widehat{\boldsymbol{\beta}}_i$ *is the differencing estimator given by* (2.3)*.*

2.2 Proof

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

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

We denote the collection of $\mathcal{L}^q$ *stochastic Lipschitz-continuous* over $[0, 1]$ classes by Lip_q. Define the *physical dependence measure* for the process $\mathbf{L}(v, \mathcal{F}_t)$ as the following:

$$\delta_q(\mathbf{L}, t) = \sup_{v\in[0,1]} ||\mathbf{L}(v, \mathcal{F}_t) - \mathbf{L}(v, \mathcal{F}'_t)||_q,$$

where $\mathcal{F}'_t = (\dots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \dots, \epsilon_{t-1}, \epsilon_t)$ is a coupled process with ϵ'_0 being an i.i.d. copy of ϵ_0.

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

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

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

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

We define the first-differenced regressors as follows.

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

where $v_t = t/T$. Similarly,

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

With these assumptions we can prove the following proposition.

Proposition 2.2. *Under Assumptions (C4) - (C6),*

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

Proof of Proposition 2.2. Note the following

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

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

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

and

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

Thus,

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

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

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

which completes the proof. □

To be able to prove the next propositions, we need additional assumptions on the relationship between the covariates and the error process.

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

(C8) Let $\zeta_{i,t} = (u_{it}, \eta_{it})^T$. Define $\mathcal{I}_{i,t} = (\ldots, \zeta_{i,t-1}, \zeta_{i,t})$ and $\mathbf{U}(v, \mathcal{I}_{i,t}) = \mathbf{H}_i(v, \mathcal{U}_{it}) G_i(\mathcal{Z}_{it})$. Then, $\sum_{s=0}^{\infty} \delta_2(\mathbf{U}, k) < \infty$.

Proposition 2.3. *Under Assumptions (C1) - (C8),*

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

Proof of Proposition 2.3. Let $\zeta_{i,t} = (u_{it}, \eta_{it})^T$. We need the following notation:

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

Then,

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

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

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

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

Moreover,

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

Thus, by Assumption ((C8))

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

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

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

Therefore, the proposition follows. □

Proof of Theorem 2.1. Recall the differencing estimator $\widehat{\boldsymbol{\beta}}_i$:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_i &= \Big(\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^T\Big)^{-1} \sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta Y_{it} = \\
&= \Big(\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^T\Big)^{-1} \sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Big(\Delta\mathbf{X}_{it}^T\boldsymbol{\beta}_i + \Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big)\Big) = \\
&= \boldsymbol{\beta}_i + \Big(\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^T\Big)^{-1} \sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it} + O\Big(\frac{1}{T}\Big)\Big(\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^T\Big)^{-1} \sum_{t=1}^{T} \Delta\mathbf{X}_{it}.
\end{aligned}$$

This leads to

$$\begin{aligned}
\sqrt{T}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) = &\Big(\frac{1}{T}\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^T\Big)^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\varepsilon_{it} + \\
&+ O\Big(\frac{1}{\sqrt{T}}\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} \Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^T\Big)^{-1} \frac{1}{T}\sum_{t=1}^{T} \Delta\mathbf{X}_{it}.
\end{aligned}$$

By Proposition 2.2, we know that for each $t \in \{1, \ldots, T\}$ $|\Delta\mathbf{X}_{it}| = O_P(1)$, which lead to

$$\left|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\right| = O_P(1). \tag{2.4}$$

Similarly, by Proposition 2.2

$$\left|\left|\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbf{X}_{it}\Delta\mathbf{X}_{it}^{T}\right|\right| = O_P(1), \tag{2.5}$$

where $||A||$ with A being a matrix is a matrix norm induced by the Euslidean norm on a vector.

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

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

By applyting Proposition 2.3, (2.4) and (2.5), the statement of the theorem follows. $\square$

References

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