Multiscale Inference for Nonparametric Time Trends

Marina Khismatullina [1] Michael Vogt [1]

July 12, 2018

[1]University of Bonn

Table of contents

Introduction

Motivation

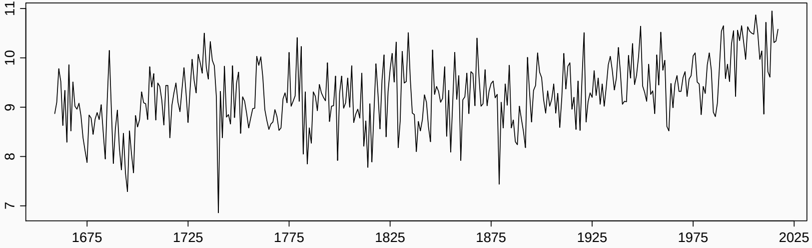


Figure 1: Yearly mean temperature in Central England from 1659 to 2017

- Does the observed time series have a trend at all?
- If so, which are the time regions where there is a strong trend?
- Is the trend decreasing or increasing in these regions?

Idea

Research question

Develop multiscale methods to test qualitative hypotheses about nonapametric time trends.

One time series is observed:

- Test the null hypothesis of the existence of a time trend.
- Identify time regions with upward or downward movement in the trend.

Multiple time series are observed:

- Detect which time trends are different and where.

Model

Model

We observe a single time series $\{Y_t : 1 \leq t \leq T\}$ of length T. The observations come from a following model:

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t$$

- m is an unknown trend function on $[0, 1]$;
- $\{\varepsilon_t : 1 \leq t \leq T\}$ is a zero-mean stationary error process.

Literature

Multiscale approaches for independent data

- SiZer method (Chaudhuri and Marron, 1999, 2000)
- Testing monotonicity of the trend function (Hall and Heckman, 2000, Dümbgen and Spokoiny, 2001)

Multiscale methods for dependent data

- Extensions to SiZer method (Park et al. 2004, 2009, Rondonotti et al. 2007)

The multiscale method

Testing

Testing problem:

$$
\begin{aligned}
H_0 : &\quad m = 0 \\
H_1 : &\quad m \neq 0
\end{aligned}
$$

Test Statistic

For a given location $u \in [0,1]$ and bandwidth h we construct the kernel averages

$$\widehat{\psi}_T(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h) Y_t,$$

where

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

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

$$S_{T,\ell}(u,h) = \frac{1}{Th}\sum_{t=1}^{T} K\Big(\frac{t/T-u}{h}\Big)\Big(\frac{t/T-u}{h}\Big)^{\ell}$$

for $\ell = 0,1,2$ and K is a kernel function.

Test Statistic

Test statistic is defined as follows

$$\widehat{\Psi}_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\},$$

where

- $\widehat{\sigma}^2$ is an appropriate estimator of the long-run variance σ^2 (Hall and Van Keilegom (2003));
- $\mathcal{G}_T$ is the set of points (u, h) that are taken into consideration;
- $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ is an additive correction term (Dümbgen and Spokoiny (2001)). Explanation

Test procedure

Gaussian version of the test statistic:

$$\Phi_T = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_T(u,h)}{\sigma} \right| - \lambda(h) \right\},$$

where

- $\phi_T(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\, \sigma Z_t$;
- Z_t are independent standard normal random variables;
- $q_T(\alpha)$ is $(1-\alpha)$ quantile of Φ_T.

Test procedure

For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_T > q_T(\alpha)$.

Assumptions

$\mathcal{C}1$ The variables ε_t are weakly dependent.

$\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.

$\mathcal{C}3$ Standard assumptions on the kernel function K.

$\mathcal{C}4$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

$$\mathcal{G}_T = \Big\{(u,h) : u = t/T \text{ for some } 1 \leq t \leq T \text{ and } h \in [h_{\min}, h_{\max}] \\ \text{with } h = t/T \text{ for some } 1 \leq t \leq T\Big\},$$

$\mathcal{C}5$ $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$ and $h_{\max} = o(1)$.

$\mathcal{C}6$ Assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$.

Theoretical properties

Proposition

Under our assumptions and under $H_0 : m = 0$ it holds that

$$\mathrm{P}\big(\widehat{\Psi}_T \le q_T(\alpha)\big) = (1 - \alpha) + o(1).$$

Proposition

Under our assumptions and under local alternatives we have

$$\mathrm{P}\big(\widehat{\Psi}_T \le q_T(\alpha)\big) = o(1).$$

Strategy of the proof

- Replace the statistic $\widehat{\Psi}_T$ under $H_0 : m = 0$ by a statistic $\widetilde{\Phi}_T$ with the same distribution and the property that

$$\big|\widetilde{\Phi}_T - \Phi_T\big| = o_p(\delta_T),$$

  where $\delta_T = o(1)$. To do so, we make use of strong approximation theory for dependent processes as derived in Berkes et al. (2014)

- Using the anti-concentration results for Gaussian random vectors (Chernozhukov et al. 2015), prove that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$, i.e.

$$\sup_{x \in \mathbb{R}} \mathrm{P}\big(|\Phi_T - x| \leq \delta_T\big) = o(1).$$

- Show that

$$\sup_{x \in \mathbb{R}} \big|\mathrm{P}(\widetilde{\Phi}_T \leq x) - \mathrm{P}(\Phi_T \leq x)\big| = o(1).$$

Testing for a constant trend function

Testing

Testing problem:

$$\begin{aligned} H_0 &: \quad m' = 0 \\ H_1 &: \quad m' \neq 0 \end{aligned}$$

Test Statistic

For a given location $u \in [0, 1]$ and bandwidth h we construct the kernel averages

$$\psi'_T(u, h) = \sum_{t=1}^{T} w'_{t,T}(u, h) Y_t,$$

where

$$w'_{t,T}(u, h) = \frac{\Lambda'_{t,T}(u, h)}{\{\sum_{t=1}^{T} \Lambda'_{t,T}(u, h)^2\}^{1/2}},$$

$$\Lambda'_{t,T}(u, h) = K\Big(\frac{t/T - u}{h}\Big)\Big[S_{T,0}(u, h)\Big(\frac{t/T - u}{h}\Big) - S_{T,1}(u, h)\Big]$$

$$S_{T,\ell}(u, h) = \frac{1}{Th} \sum_{t=1}^{T} K\Big(\frac{t/T - u}{h}\Big)\Big(\frac{t/T - u}{h}\Big)^{\ell}$$

for $\ell = 0, 1, 2$ and K is a kernel function.

Test Statistic

Test statistic is defined as follows

$$\widehat{\Psi}'_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} \Big| - \lambda(h) \Big\},$$

where

- $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ is an additive correction term;
- $\mathcal{G}_T$ is the set of points (u,h) that are taken into consideration;
- $\widehat{\sigma}^2$ is an appropriate estimator of the long-run variance σ^2.

Test procedure

Gaussian version of the test statistic:

$$\Phi'_T = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi'_T(u,h)}{\sigma} \right| - \lambda(h) \right\},$$

where

- $\phi'_T(u,h) = \sum_{t=1}^{T} w'_{t,T}(u,h)\, \sigma Z_t$;
- Z_t are independent standard normal random variables;
- $q'_T(\alpha)$ is $(1-\alpha)$ quantile of Φ'_T.

Test procedure

For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}'_T > q'_T(\alpha)$.

Theoretical properties

Proposition

Under our assumptions and under $H_0 : m' = 0$ it holds that

$$\mathrm{P}\big(\widehat{\Psi}'_T \le q'_T(\alpha)\big) = (1-\alpha) + o(1).$$

Proposition

Under our assumptions and under local alternatives, we have

$$\mathrm{P}\big(\widehat{\Psi}'_T \le q'_T(\alpha)\big) = o(1).$$

Theoretical properties

Define

$$\Pi_T^+ = \big\{ I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^+ \text{ and } I_{u,h} \subseteq [0,1] \big\}$$
$$\Pi_T^- = \big\{ I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^- \text{ and } I_{u,h} \subseteq [0,1] \big\}$$

with

$$\mathcal{A}_T^+ = \Big\{ (u,h) \in \mathcal{G}_T : \frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} > q'_T(\alpha) + \lambda(h) \Big\}$$
$$\mathcal{A}_T^- = \Big\{ (u,h) \in \mathcal{G}_T : -\frac{\widehat{\psi}'_T(u,h)}{\widehat{\sigma}} > q'_T(\alpha) + \lambda(h) \Big\}$$

Theoretical properties

Proposition

Under our assumptions, for events
$E_T^+ = \Big\{ \forall I_{u,h} \in \Pi_T^+ : m'(v) > 0 \textit{ for some } v \in I_{u,h} \Big\}$ *it holds that and*
$E_T^- = \Big\{ \forall I_{u,h} \in \Pi_T^- : m'(v) < 0 \textit{ for some } v \in I_{u,h} \Big\}$ *it holds that*

$$\mathrm{P}\big(E_T^+\big) \geq (1-\alpha) + o(1)$$
$$\mathrm{P}\big(E_T^-\big) \geq (1-\alpha) + o(1)$$

Theoretical properties

Minimal intervals

An interval $I_{u,h} \in \Pi_T^+$ is called **minimal** if there is no other interval $I_{u',h'} \in \Pi_T^+$ with $I_{u',h'} \subset I_{u,h}$.

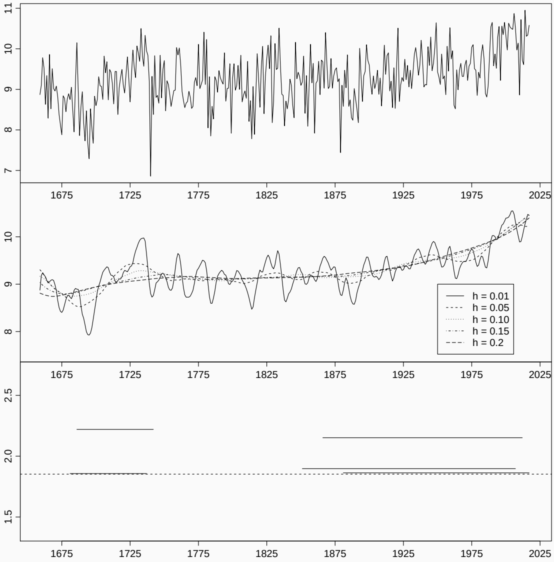
Define

$$\Pi_T^{min,+} = \text{ set of minimal intervals from } \Pi_T^+,$$
$$E_T^{min,+} = \Big\{ \forall I_{u,h} \in \Pi_T^{min,+} : m'(v) > 0 \text{ for some } v \in I_{u,h} \Big\}$$

Since $E_T^{min,+} = E_T^+$, we have

$$\mathrm{P}\big(E_T^{min,+}\big) \geq (1-\alpha) + o(1).$$

Testing for presence of time trend in temeprature data



Simulation results of the multiscale test for constant trend function

Table 1: Size of the multiscale test.

	nominal size α		
T	0.01	0.05	0.1
250	0.004	0.039	0.092
350	0.012	0.051	0.069
500	0.006	0.047	0.094
1000	0.014	0.058	0.105

Power of the multiscale test for different slope parameters β.

Table 2: $\beta = 1.25$

	nominal size α		
T	0.01	0.05	0.1
250	0.085	0.252	0.341
350	0.236	0.396	0.470
500	0.315	0.577	0.669
1000	0.763	0.900	0.936

Table 3: $\beta = 1.875$

	nominal size α		
T	0.01	0.05	0.1
250	0.318	0.621	0.714
350	0.648	0.796	0.865
500	0.793	0.943	0.967
1000	0.997	1.000	1.000

Table 4: $\beta = 2.5$

	nominal size α		
T	0.01	0.05	0.1
250	0.693	0.898	0.937
350	0.929	0.981	0.990
500	0.986	1.000	1.000
1000	1.000	1.000	1.000

Conclusion

Conclusion

We developed multiscale methods to test qualitative hypothese about nonparametric time trends:

- whether the trend is present at all;
- whether the trend function is constant;
- in which time regions there is an upward/downward movement in the trend.

We derived asymptotic theory for the proposed tests.

As an application of our method, we analyzed the behavior of the yeraly mean temperature in Central England from 1659 to 2017.

Thank you!

Testing for equality of the time trends

Model

We observe n time series $\mathcal{Y}_i = \{Y_{it} : 1 \leq t \leq T\}$ of length T for $1 \leq i \leq n$

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

where

- m_i is an unknown trend function on $[0, 1]$, that are Lipschitz continuous and normalized such that $\int_0^1 m_i(u)du = 0$;
- α_i is an intercept term;
- $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary error process;
- $\mathcal{E}_i$ are independent across i.

Model

Not observed variables

$$Y_{it}^{o} = Y_{it} - \alpha_i = m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}$$

can be approximated by

$$\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i = Y_{it} - \frac{1}{T}\sum_{i=1}^{T} Y_{it}.$$

By construction

$$Y_{it}^{o} - \widehat{Y}_{it} = \widehat{\alpha}_i - \alpha_i = \frac{1}{T}\sum_{i=1}^{T}\varepsilon_{it} + \frac{1}{T}\sum_{i=1}^{T} m_i(t/T) = O_P(T^{-1/2}).$$

Test Statistic

For a given location $u \in [0,1]$, bandwidth h and a pair of time series i and j we construct the kernel averages

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

where

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

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

$$S_{T,\ell}(u,h) = \frac{1}{Th}\sum_{t=1}^{T} K\Big(\frac{t/T-u}{h}\Big)\Big(\frac{t/T-u}{h}\Big)^{\ell}$$

for $\ell = 0,1,2$ and K is a kernel function.

Test Statistic

Our multiscale statistic is defined as follows

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

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

where

- $\lambda(h) = \sqrt{2 \log\{1/(2h)\}}$ is an additive correction term;
- $\mathcal{G}_T$ is the set of points (u, h) that are taken into consideration;
- $\widehat{\sigma}_i^2$ is an appropriate estimator of the long-run variance σ_i^2.

Test procedure

Testing problem:

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

Gaussian version of the test statistic:

$$\Phi_{n,T} = \max_{1 \le i < j \le n} \Phi_{ij,T},$$

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

where

$$\phi_{ij,T}(u,h) = \sum\nolimits_{t=1}^{T} w_{t,T}(u,h) \left\{ \widehat{\sigma}_i \Big(Z_{it} - \frac{1}{T} \sum_{t=1}^{T} Z_{it} \Big) - \widehat{\sigma}_j \Big(Z_{jt} - \frac{1}{T} \sum_{t=1}^{T} Z_{jt} \Big) \right\};$$

Z_t are independent standard normal random variables;

$q_{n,T}(\alpha)$ is $(1-\alpha)$ quantile of $\Phi_{n,T}$.

Test procedure

For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$.

Theoretical properties

Proposition

Supose that $\mathcal{E}_i$ are independent across i and satisfy $\mathcal{C}1 - \mathcal{C}2$ for each i. Under our remaining assumptions and under $H_0 : m_1 = m_2 = \ldots = m_n$ it holds that

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

Proposition

Let the conditions of previous proposition be satisfied. Under local alternatives we have

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

Clustering, group structure

- The null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is violated.
- There exist sets or groups of time series $G_1, \ldots, G_N$ with $N \leq n$ and $\{1, \ldots, n\} = \dot{\bigcup}_{\ell=1}^{N} G_\ell$ such that for each $1 \leq \ell \leq N$ we have $m_i = g_\ell \quad$ for all $i \in G_\ell$, where g_ℓ are group-specific trend functions.
- For any $\ell \neq \ell'$, the trends $g_{\ell,T}$ and $g_{\ell',T}$ differ in the following sense: There exists $(u, h) \in \mathcal{G}_T$ with $[u-h, u+h] \subseteq [0, 1]$ such that $g_{\ell,T}(w) - g_{\ell',T}(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u-h, u+h]$ or $g_{\ell',T}(w) - g_{\ell,T}(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u-h, u+h]$, where $0 < c_T \to \infty$.

Clustering, algorithm

Dissimilarity measure between two sets of time series S and S':

$$\widehat{\Delta}(S, S') = \max_{\substack{i \in S, \\ j \in S'}} \widehat{\Psi}_{ij,T}.$$

Clustering algorithm

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

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

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

and merge them into a new cluster.

Clustering, theoretical properties

The estimator of the number of groups is

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

Proposition

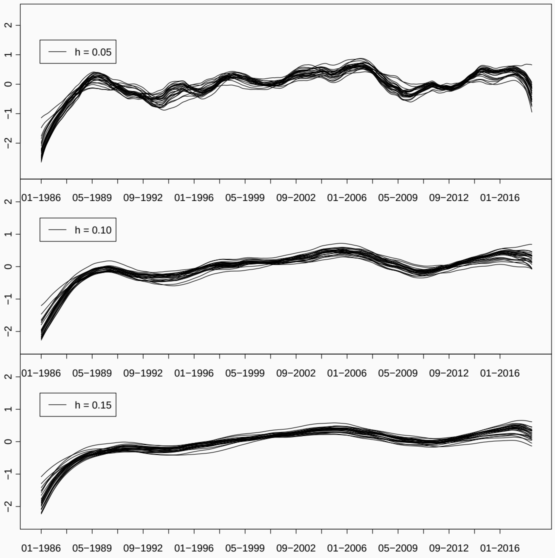
Let the conditions of previous propositions be satisfied. Then

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

and

$$\mathrm{P}\big(\widehat{N} = N\big) \ge (1 - \alpha) + o(1).$$

Testing for equality of different temperature time trends



Idea behind the additive correction

Consider the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \right|$$

under the null hypothesis $H_0 : m = 0$ and under simplifying assumptions:

- the errors ε_i are i.i.d. normally distributed;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1\leq l\leq L} \max_{1\leq k\leq 1/2h_l} \left| \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} \right|$$

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1\leq l\leq L} \max_{1\leq k\leq 1/2h_l} \left| \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} \right|$$

$\Rightarrow \quad \max_k \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} = \sqrt{2\log(1/2h_l)} + o_P(1) \to \infty$ as $h \to 0$ and the stochastic behavior of $\widehat{\Psi}_{T,\text{uncorrected}}$ is dominated by $\frac{\widehat{\psi}_T(u_k,h_l)}{\sigma}$ for small bandwidths h_l.