Nonparametric Estimation of the Density of a Change-Point*

By Marine Carrasco† (University of Montreal & CIREQ)

and Hugo Peltier‡ (INSEE)

January 2024

Abstract

The paper considers a panel model where the regression coeffi cients undergo changes at an unknown time point, different for each series. The timings of changes are assumed to be independent, identically distributed, and drawn from some com-mon distribution, the density of which we aim to estimate nonparametrically. The estimation procedure involves two steps. First, change-points are estimated indi-vidually for each series using the least-squares method. While these estimators are not consistent, they can be regarded as noisy signals of the true change-points. To address the inherent estimation error, a deconvolution kernel estimator is applied to estimate the density of the change-point. The paper establishes the consistency of this estimator and demonstrates that the rate of convergence of the Mean Inte-grated Squared error (MISE) is faster than that obtained with normal or Laplacian errors. Finally, using a Bayesian approach, we propose an estimator of the poste-rior means of the breakpoints, utilizing nonparametric estimates of the required densities. An application of the proposed methodology to portfolio returns reveals how quickly the markets responded to the Covid shock.

JEL: C13, C14, C23. Key words: deconvolution, kernel estimator, panel, structural break.

*Carrasco thanks FRQSC and SSHRC for partial financial support.

†marine.carrasco@umontreal.ca

‡hugo.jg.peltier@gmail.com

1 Introduction

Structural breaks frequently occur in economics as the result of a policy change or a shock like Covid. The time it takes for economic agents to respond to such shocks typically vary among individuals or markets. Assuming that the change times are drawn from the same distribution, it is of particular interest to estimate this distribution. Knowledge of the distribution of changes can be valuable in various other scenarios. In medicine, for example, it is often crucial to determine whether the introduction of a new drug affects the health state of a patient. Health indicators are recorded from a sample of patients, and due to individual heterogeneity, they will not all react at the same time. However, it is reasonable to assume that the reaction times follow the same distribution. Understanding this distribution can provide valuable insights into the impact of the new drug on patients' health.

The aim of this paper is to deduce the distribution of the break points from panel data. In the change-point literature, the behavior of the shift is often either not specified or assumed to be the realization of an Exponential or Weibull distribution (Pollak and Siegmund (1985)). It is quite surprising that the nonparametric estimation of the distribution of the change-point has received so little attention. However, related works include Joseph and Wolfson (1993a, b) who consider discrete-time observations and estimate a probability mass function over a finite period using the EM algorithm, and Joseph et al. (1997) who take a Bayesian approach to the estimation of the density of the change-point.

We consider a large panel of linear regressions where a subset of or all coefficients are affected by a single change which occurs at a random time point. The break points are assumed to be independent and identically distributed according to the same continuous distribution which we aim to determine. If the break-times were observed, the distribution could be estimated by the usual nonparametric kernel estimator (see Li and Racine, 2007). As the change-points are not known a priori, we adopt a two step approach.

Firstly, the change-points of each individual series are estimated by least-squares. In the case of a shrinking magnitude of the shift, Bai (1997) shows that the break point estimate converges to a distribution which is free of nuisance parameters when the time span increases.

Our second step involves leveraging these estimators to estimate the underlying density. Unfortunately, these estimators are not consistent and can be seen as noisy signal of the true break point. Consequently, we rely on a modification of the deconvolution kernel introduced by Stefanski and Carroll (1990). We show that the rates of convergence obtained here are much better than those obtained when data are observed with normal or even Laplace errors. The rapid rate of convergence can be attributed to the shape of the distribution of the errors which is very "peaked" at zero.

From a theoretical point of view, we make several contributions. First, we derived the characteristic function (CF) corresponding to the asymptotic distribution of the break point estimation error. We show that the tail behavior of this CF corresponds to that of ordinary smooth random variable (r.v), which are well-known to be easier to deconvolve than the supersmooth r.v. (see Fan, 1991). Second, we propose a deconvolution kernel estimator for heterogeneous panels, i.e. panels where the coefficients differ from one series to the next. Third, we derive its rate of convergence in the ideal situation where the amplitude of the shift is fixed and known. We show that the rate of the MISE is $n^{-1/2}$. Next, we investigate the case of shrinking amplitude and unknown parameters. We show that the rate of the MISE depends on the shrinking rate and is slower than before. Finally, using Bayes theorem, we propose an estimator of the posterior mean of the break points which uses nonparametric estimates of the necessary densities. This estimator exploits the knowledge that the break points are all drawn from the same distribution and hence is expected to give more accurate estimates of the break points than the least-squares. Its motivation is similar to that of the nonparametric empirical Bayes estimators proposed for instance by Gu and Koenker (2017) and Liu, Moon, and Schorfheide (2020).

From an applied point of view, our simulations show that the deconvolution kernel is capable of capturing features of the distribution which would be missed by the conventional kernel estimator. Moreover, we contribute to the modeling of financial returns by studying the impact of covid shock in 2020 and Ukraine war in 2022 on the coefficients in Fama-French five factor model. We show that both events yielded a shift in the coefficients and we estimate the density of the break points. While the density of the break point in 2020 is unimodal and quite concentrated, that for 2022 is bimodal suggesting the possibility of two shifts. Our empirical Bayes estimators permit to determine that the Asian markets moved first during the covid crisis and the US portfolios last. Our nonparametric estimator of the break point density could be used for policy evaluation. Indeed, it could be used to determine how fast agents react to a policy change.

Our estimation method draws from two different literatures: one focused on structural change and the other on deconvolution kernel. The literature concerning structural break estimation provides the asymptotic theory of the maximum likelihood estimator of the break point (Hinkley (1970), Picard (1985), Yao (1987)) and of the least squares estimator by Bai (1994, 1997). Recent work by Jiang et al. (2018) and Baek (2024) demonstrate that when the amplitude of the shift is small, of the order $O\left(T^{-1/2}\right)$, the asymptotic distribution of the break point estimator is trimodal (with modes at 0 and T) and asymmetric. In this paper, we consider a shrinking amplitude of the order $O\left(T^{\alpha-1/2}\right)$ for some $\alpha > 0$, as in Bai (1994, 1997); this ensures a symmetric unimodal distribution of the break point estimator. This aligns well with our application on the impact of covid, where the estimated break points lie far from the boundaries.

There is a vast literature on the deconvolution kernel estimator. This estimator was initially introduced by Carroll and Hall (1988) and Stefansky and Carroll (1990) with its properties studied by Fan (1991) for independent data and Masry (1993) for weakly dependent observations. Some extensions to the heteroscedastic case were explored by Meister (2006) and Delaigle and Meister (2008). A version robust to the presence of zeros in the characteristic function of the error was proposed by Carrasco and Florens (2011). Given that our estimators of break points can be considered as the sum of the true break point plus an additive error, our paper fits also in the literature on measurement error, see for instance Chen, Hong, and Nekipol (2011) for a survey. In our paper, the measurement error comes from the first step estimation of the break points. Our approach is further related to Arellano and Bonhomme's (2012) nonparametric estimation of the density of individual effects γ_i based on a deconvolution estimator using least-squares estimates of γ_i as input.

The paper is organized as follows: In Section 2, we present the model. Section 3 lists assumptions and describe the properties of the structural change estimator. Section 4 reviews the nonparametric deconvolution kernel estimator. Section 5 gives the rate of convergence of this estimator when the parameters are known and fixed. In Section 6, the case with unknown parameters is considered. In Section 7, we explain how to reestimate the break points using empirical Bayes. Simulation results are presented in Section 8. Section 9 contains an empirical application to Fama-French factor model. Finally, Section 10 concludes. All the proofs are collected in Appendix.

2 Model and overview

Consider a panel data model with a break in the regression coefficients:

$$\begin{cases} y_{jt} = w_{jt}^{'}\alpha_j + z_{jt}^{\prime}\delta_{1j} + \varepsilon_{jt},\ t = 1, 2, ..., k_j, \\ y_{jt} = w_{jt}^{'}\alpha_j + z_{jt}^{\prime}\delta_{2j} + \varepsilon_{jt},\ t = k_j + 1, 2, ..., T, \end{cases} \tag{1}$$

where $j = 1, 2, ..., n$. The shocks ε_{jt} are independent, identically distributed (i.i.d.) across individuals j and potentially cross-sectionally heteroscedastic with $E\left(\varepsilon_{jt}^2\right) = \sigma_j^2$. The variables $\{y_{jt}, w_{jt}, z_{jt}\}_{j,t}$ are observable. The regressors are assumed to be strictly exogenous. The coefficient of w_{jt} does not change over time, whereas that of z_{jt} varies at an unknown time k_j. The variables y_{jt} and z_{jt} could represent two economic variables whose relationship has been altered due to a shock, such as a policy change or an event like the Covid pandemic. We assume that the change-points k_j are random and independently drawn from an unknown distribution f_k. We want to estimate k_j, $j = 1, 2, ..., n$ and their distribution f_k.

Using the notation $x_{jt} = \left(w_{jt}^{\prime}, z_{jt}^{\prime}\right)^{\prime}$, $\beta_j = \left(\alpha_j^{\prime}, \delta_{1j}^{\prime}\right)^{\prime}$, and $\delta_j = \delta_{2j} - \delta_{1j}$, Model (1) can be rewritten as

$$\begin{cases} y_{jt} = x_{jt}^{'}\beta_j + \varepsilon_{jt},\ t = 1, 2, ..., k_j, \\ y_{jt} = x_{jt}^{'}\beta_j + z_{jt}^{\prime}\delta_j + \varepsilon_{jt},\ t = k_j + 1, 2, ..., T. \end{cases}$$

The estimation method proposed in this paper involves two steps:

Step 1. The change-point k_j is estimated using the least-squares estimator denoted $\widehat{k}_j$ for each individual series separately.

Step 2. Leveraging the set of estimators $\widehat{k}_j$, $j = 1, 2, ..., n$, we construct a deconvolution kernel estimator of f_k.

These two steps are detailed in Sections 3 and 4 below.

3 Estimation of the change-points

First, we estimate k_j by ordinary least-squares for each individual series j separately. To simplify the notation, we are going to drop the subscript j in this section and describe the estimation procedure for one typical series. Following Bai (1997), the least-squares estimator $\widehat{k}$ of k is defined as

$$\widehat{k} = \arg \min_{1 \le k \le T} \sum_{t=1}^{T} \left(y_t - x_t' \widehat{\beta}_k - z_t' \widehat{\delta}_k I \left(t > k \right) \right)^2$$

where $\widehat{\beta}_k$ and $\widehat{\delta}_k$ are the OLS estimators of β and δ for a given k.

The assumptions needed to obtain the asymptotic distribution of $\widehat{k}$ are listed below.

A1. $k = [\tau T]$, where $\tau \in (0, 1)$ and $[.]$ is the greatest integer value function.

A2. $(x_t, \varepsilon_t)_t$ is strictly stationary with $E(z_t z_t') = Q < \infty$, $E(\varepsilon_t^2 | x_t) = \sigma^2$, $E\left(\|x_t\|^{4+\nu}\right) \le C$ some $\nu > 0$, $C < \infty$.

A3. $X'X/T$ converges in probability to a nonrandom and positive definite matrix, where $X = (x_1, x_2, ..., x_T)'$.

A4. $\{\varepsilon_t, \mathcal{F}_t\}$ is a martingale difference sequence for $\mathcal{F}_t = \sigma\{\varepsilon_s, x_{s+1}, s \le t\}$. Moreover, $E\left(|\varepsilon_t|^{4+\nu}\right) < C$ for some $C < \infty$ and $\nu > 0$.

A5. $\delta_T = \delta_0 v_T$ where v_T is a positive scalar such that $v_T \to 0$ and $T^{1/2-\alpha} v_T \to \infty$ for some $\alpha \in (0, 1/2)$, $\delta_0 \neq 0$.

These assumptions are taken from Bai (1997). The martingale difference sequence in A4 could be replaced by a weak dependence at the cost of a more complicated estimator of the asymptotic variance. Assumption A5 stipulates that the amplitude of the shift

δ goes to zero as T goes to infinity. To make the dependence in T explicit, we use the notation δ_T. This shrinking shift is standard in the structural change literature and permits to obtain a relatively simple expression for the asymptotic distribution of $\widehat{k}$ (see Picard (1985), Yao (1987), and Bai (1997) among others).

Under Assumptions A1 to A5, Bai (1997, Equation (14)) shows that as $T \to \infty$,

$$\frac{\delta_T' Q \delta_T}{\sigma^2} \left(\widehat{k} - k \right) | k \xrightarrow{d} \xi \sim \arg\max_{t \in \mathbb{R}} \left\{ W(t) - \frac{|t|}{2} \right\} \tag{2}$$

where $\{W(t) : -\infty < t < \infty\}$ is a two-sided standard Brownian motion with $W(0) = 0$. The limiting distribution ξ appeared first in Picard (1985) and Yao (1987), hence will be referred to as Picard-Yao distribution in the sequel. It follows from (2) that the asymptotic distribution of $\widehat{k} - k$ is known up to an unknown parameter $\omega = \sigma^2 / \delta_T' Q \delta_T$. In the sequel, ω will be estimated by $\hat{\omega} = \widehat{\sigma}^2 / \hat{\delta}_T' \widehat{Q} \hat{\delta}_T$ where the slope parameter δ_T is estimated by the least-squares estimator $\hat{\delta}_T$ (see Bai, 1997 for details), $\hat{Q} = \sum_{t=1}^{T} z_t z_t' / T,$ and $\hat{\sigma}^2 = \sum_{t=1}^{T} \left(y_t - x_t' \hat{\beta} - z_t' \hat{\delta}_T I \left(t > \hat{k} \right) \right)^2 / T$ where $\hat{\beta}$ is the least-squares estimate of β. Bai (1997, Corollary 1) shows that $\frac{\hat{\delta}_T' \widehat{Q} \hat{\delta}_T}{\widehat{\sigma}^2}$ is a $\sqrt{T}$ consistent estimator of $\frac{\delta_T' Q \delta_T}{\sigma^2}$. However, this is not the case for $\hat{\omega}$ because $\hat{\delta}_T$ goes to zero.

The estimator $\widehat{k}$ can be regarded as a noisy observation of k where the distribution of the measurement error is known up to the unknown parameter ω, so that

$$\widehat{k} = k + \omega \xi \tag{3}$$

where k and ξ are independent from each other. Equation (3) is an approximation which will hold for T large enough. Notice that $\widehat{k}$ is not a consistent estimator of k because $\widehat{k} - k$ is not multiplied by a function of T. However, the collection of $\widehat{k}_j$ will be useful to recover the distribution of k_j using a deconvolution kernel described in Section 4.

4 Estimation of the density of k_j

For each series j, we obtained an estimator $\hat{k}_j$ of k_j. Each estimator $\hat{k}_j$ is approximately the sum of the unobservable change-point k_j and an error ξ_j times a constant ω_j :

$$\hat{k}_j = k_j + \omega_j \xi_j \tag{4}$$

where

$$\xi_j \sim \arg\max_v \left\{ W(v) - \frac{1}{2} |v| \right\}$$

and $\omega_j = \sigma_j^2 / \delta'_{Tj} Q_j \delta_{Tj}$. Let us denote the density of $\hat{k}_j$ by $f_{\hat{k}_j}$ while the density of k_j, which we wish to estimate, is denoted f_k. First, the coefficients ω_j are assumed to be known. This assumption will be relaxed later. Since the distribution of ξ_j and that of the k_j are independent, the distribution of $\hat{k}_j$ is, in fact, the convolution of the distributions of k_j and $\omega_j \xi_j$.

$$f_{\hat{k}_j} = f_k * f_{\omega_j \xi}$$

To recover f_k, we need to deconvolve the two densities, f_k and $f_{\omega_j \xi}$. To deal with this problem, we are going to use a deconvolution kernel first introduced by Carroll and Hall (1988) and Stefanski and Carroll (1990) and whose properties have been studied by Fan (1991). With regards to this, we first impose some regularity conditions. Let φ_X denote the characteristic function of a random variable X with probability density function f, that is, $\varphi_X(t) = E\left(e^{itX}\right) = \int e^{itx} f(x) dx$.

B1. $|\varphi_\xi(t)| > 0 \quad \forall t \in \mathbb{R}$.

B2. The density f_k is bounded, twice differentiable, and $\int f_k''(x)\, dx < \infty$.

B3. The kernel K satisfies $\int K(t)dt = 1$ and $\int tK(t)dt = 0$.

And for every fixed $h > 0$ and $\omega_j > 0$:

B4. $\sup_t \left| \frac{\varphi_K(t)}{\varphi_\xi(\omega_j t/h)} \right| < \infty$.

B5. $\int \left| \frac{\varphi_K(t)}{\varphi_\xi(\omega_j t/h)} \right| dt < \infty$.

B6. $\int_0^\infty x^3 \varphi_K^2(x)\, dx < \infty$.

The estimation of f_k is essentially based on the properties of the characteristic function. Indeed, we have

$$\varphi_{\hat{k}_j}(t) = \varphi_k(t) \times \varphi_{\omega_j \xi}(t).$$

Note that $\varphi_{\omega_j \xi}(x) = \varphi_\xi(\omega_j x)$. According to the Fourier inversion theorem, the target density can be written as

$$f_k(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\varphi_{\hat{k}_j}(t)}{\varphi_\xi(\omega_j t)} dt.$$

The estimator of f_k is obtained by replacing $\varphi_{\hat{k}_j}(t)$ by its estimator. Assuming ω_j is the same for all j with $\omega_j = \omega$, Stefanski and Carroll's estimator of f_k is given by

$$\hat{f}_k(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(th) \frac{\frac{1}{n}\sum_{j=1}^{n} \exp\left(it\hat{k}_j\right)}{\varphi_{f_\xi}(\omega t)} dt$$

where h is a bandwidth such that $h \to 0$ as n goes to infinity.

To handle the heterogenous case where ω_j is different for each series j, we propose the following estimator inspired from Delaigle and Meister (2008):

$$\hat{f}_k(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(th) \frac{1}{n} \sum_{j=1}^{n} \frac{\exp\left(it\hat{k}_j\right)}{\varphi_\xi(\omega_j t)} dt. \quad (5)$$

This expression is actually simpler than that of Delaigle and Meister's estimator (see their equation (2.2)) because here the distributions of the error differ only through the multiplicative term ω_j, whereas Delaigle and Meister treat the more general case where the errors may have completely different distributions.

Using the notation $K_j^*(t) = \frac{1}{2\pi} \int e^{ity} \frac{\varphi_K(y)}{\varphi_\xi(\omega_j y/h)} dy$, the estimator given in (5) can be rewritten as a kernel estimator:

$$\hat{f}_k(x) = \frac{1}{nh} \sum_{j=1}^{n} K_j^* \left(\frac{\hat{k}_j - x}{h} \right).$$

The consistency and rate of convergence of this estimator will be studied in the subsequent sections. In Section 5, we will consider the special case where the relation (4) holds exactly with ω_j finite. Next, we will investigate the more realistic case where the density of ξ_j depends on T and ω_j is also a function of T and diverges when δ_{Tj} goes to zero. These two features will contribute to deteriorate the rate of convergence. We will first address the case where ω_j is known in Section 6.1 and then when ω_j is unknown in Section 6.2.

5 Rate of convergence of deconvolution kernel

In this section, we derive the rate of convergence of $\hat{f}_k(x)$ in the ideal setting where the relation (4) holds exactly with $\omega_j < \infty$. The purpose of investigating this simple case first is to gain some insights on the best possible rate we could hope to achieve. This rate crucially depends on the behavior of the characteristic function of ξ in the tails. The density f_ξ has a closed-form expression (see for instance Yao (1987)):

$$f_\xi(x) = \frac{3}{2}\exp(|x|)\,\Phi\left(-\frac{3\sqrt{|x|}}{2}\right) - \frac{1}{2}\Phi\left(-\frac{\sqrt{|x|}}{2}\right)$$

where Φ denotes the cumulative distribution function of the standard normal. Using this expression, the characteristic function of f_ξ can be calculated explicitly and is given in the following lemma.

Lemma 1.

$$\varphi_\xi(x) = \frac{9}{2\sqrt{2}}\frac{1}{1+x^2}\frac{\sqrt{1+\sqrt{1+64x^2}}}{\sqrt{1+64x^2}} + \frac{1}{2\sqrt{2}}\frac{8x^2-1}{|\,x\,|\,(1+x^2)}\frac{\sqrt{\sqrt{1+64x^2}-1}}{\sqrt{1+64x^2}} - \frac{3}{2}\frac{1}{1+x^2}.$$

From this result, it is clear that φ_ξ satisfies Assumption B1. Moreover, $\varphi_\xi(x)$ decays at an arithmetic rate $x^{-3/2}$ in the tails. So ξ belongs to the class of ordinary smooth random variables (r.v.) which typically yields faster rate of convergence than the supersmooth r.v. (see Fan (1991)).

The criterion of convergence that we consider is the Mean Integrated Squared Error (MISE):

$$\text{MISE}(\hat{f}_k) = E\int\{\hat{f}_k(x) - f_k(x)\}^2 dx \tag{6}$$

Theorem 1. Assume the relation (4) holds exactly, $\omega_j < \infty$ is fixed and (k_j, ξ_j), $j = 1, 2, ..., n$ are i.i.d. where the density of ξ_j is f_ξ. Assume that B1 to B6 hold and K is a second-order kernel $\left(\text{i.e. } \int x^2 K(x)\,dx \neq 0\right)$. Then, the asymptotic MISE is given by

$$\text{AMISE} = \frac{h^4}{4}\mu_2(K)^2 \int (f_k''(x))^2 dx + \frac{1}{n^2 \pi h^4}\left(\sum_{j=1}^{n} \omega_j^3\right)\int_0^{\infty} x^3 \varphi_K^2(x) dx$$

where $\mu_2(K) = \int x^2 K(x)\, dx$. The optimal bandwidth is then given by

$$\hat{h} = \left[\frac{4\left(\sum_{j=1}^{n} \omega_j^3\right)\int_0^{\infty} x^3 \varphi_K^2(x)\, dx}{n^2 \pi \mu_2^2(K) \int (f_k''(x))^2 dx}\right]^{\frac{1}{8}} \tag{7}$$

and the optimal MISE:

$$\inf_{h>0} \text{MISE}(h) \sim \frac{\mu_2(K)\sqrt{\left(\sum_{j=1}^{n} \omega_j^3\right)\int (f_k''(x))^2 dx \int_0^{\infty} x^3 \varphi_K^2(x)\, dx}}{n\sqrt{\pi}}$$

as n goes to infinity.

It follows from Theorem 1 that the speed of convergence of the MISE is $n^{-\frac{1}{2}}$ (note that $\sum_{j=1}^{n} \omega_j^3 = O(n)$). To make this result meaningful, we must compare it to other rates present in the literature. The deconvolution kernel has been developed to estimate the density of a variable that is measured with an error. When the measurement error follows a Laplace distribution, the speed of convergence of the MISE is $n^{-\frac{4}{9}}$ (Stefanski-Carroll (1990)) and when the error is normal, the rate is only $(\ln(n))^{-2}$ (Carroll-Hall (1988)). Optimal rates of convergence for general errors are given by Fan (1991). But none of the standard distribution gives a rate as fast as that obtained here. Therefore, $n^{-\frac{1}{2}}$ is a very good rate given the data are not directly observable. The speed of convergence depends on the tails of the characteristic function of the measurement error which itself is linked to the smoothness of the density. Figure 1 gives the plots of the pdf of the standard normal, the Laplace and the Picard-Yao distribution given in (2). We see that the smoother density is the normal which is therefore the most difficult density to deconvolve while the Picard-Yao pdf is very peaked at zero yielding a faster rate of convergence. However as expected, a speed of convergence of the MISE of order $n^{-\frac{1}{2}}$ is much slower than the rate obtained when the data are observable without error, i.e. $n^{-\frac{4}{5}}$, this is the price to pay for measurement error.

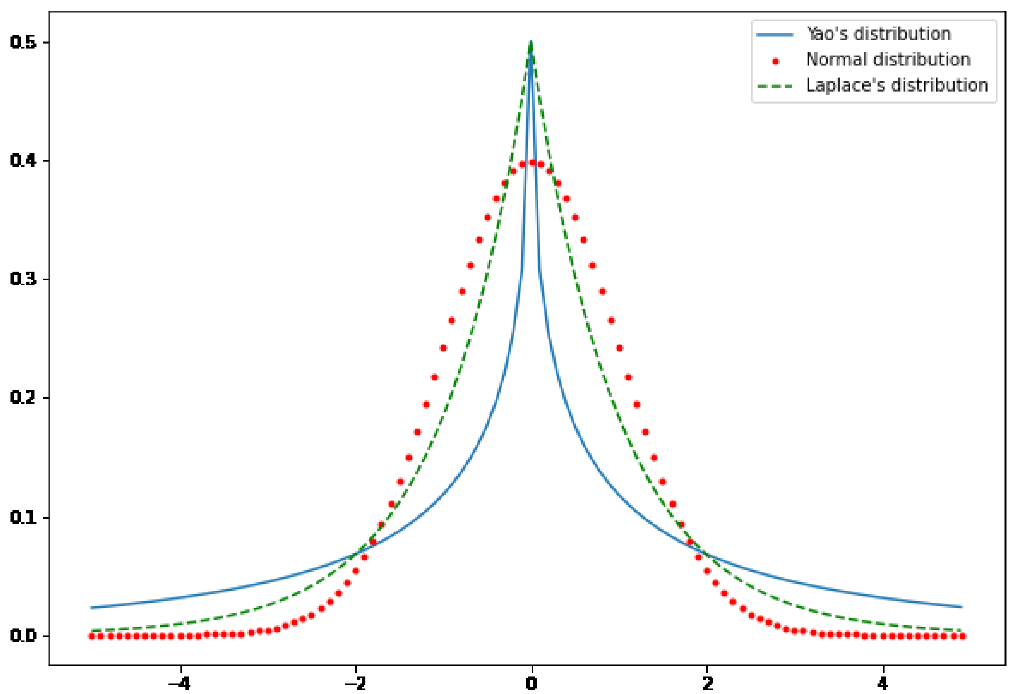

Figure 1: Shapes of densities

Theorem 1 assumes that the relation (4) holds exactly with fixed ω_j. In practice, for fixed value of T, the finite sample distribution of the measurement error may have a different shape from that of ξ. This introduces a misspecification of the distribution. When T goes to infinity, the distribution of the measurement error tends to the postulated distribution but another problem arises. When $T \to \infty$, $\delta_T \to 0$ and hence $\omega_j \to \infty$. This has for effect to blur the signal and hence slower the rate of convergence. In spite of these pitfalls, we will show in the next section that $\hat{f}_k$ is a consistent estimator even when taking into account the effect of T.

An assumption that we will maintain throughout the paper is the cross-sectional independence of the series. This assumption could be relaxed. Masry (1991) shows that if the dependence is not too strong, the rate of convergence of the deconvolution estimator remains the same.

6 Consistency of $\hat{f}_k$

In the previous section, we derived the rate of convergence of $\hat{f}_k$ in the ideal situation where (4) holds exactly and $\omega_j < \infty$ does not depend on T. We consider now the case where $\widehat{k}_j$ are obtained from the estimation of the structural break as described in Section 2. First, we assume that ω_j is known, then we assume ω_j is unknown.

6.1 Case where ω_j is known

For a given T, we denote $\frac{1}{\omega_j}\left(\widehat{k}_j - k_j\right)$ by ξ_T. The distribution of ξ_T converges to that of ξ as T goes to infinity where ξ is the Picard-Yao distribution given in (2).

Remark that the distribution of ξ_T does not have an explicit expression even in the simple case of a shift in the mean (see Hinkley, 1970). Moreover, the rate at which the density of ξ_T approaches that of ξ is unknown.

In this section, we assume that K has a characteristic function with bounded support [-1,1]. An example is given in Section 8. Such kernels are frequently used in deconvolution because they satisfy assumptions B3 to B6.

Theorem 2. Suppose ω_j is known. Assume that, for each series j, Assumptions A1 to A5 hold and $(x_{jt}, \varepsilon_{jt})_j$ are independent across j. Let K be a kernel which characteristic function has bounded support [-1,1]. Assume B1 and B2 hold. Let $a_T = \int \left(f_{\xi_T}(x) - f_{\xi}(x)\right)^2 dy$ with $a_T \to 0$ as T goes to infinity. The MISE is decomposed into variance and squared bias so that

$$\begin{aligned} Var &= O\left(\frac{\sum_j \omega_j^3}{n^2 h^4}\right) = O\left(\frac{1}{nh^4 v_T^6}\right), \\ Bias^2 &= O\left(\frac{a_T}{h^4}\frac{\sum_j \omega_j^2}{n}\right) + O\left(h^4\right) = O\left(\frac{a_T}{h^4}\frac{1}{v_T^4}\right) + O\left(h^4\right) \end{aligned}$$

where v_T was defined in A5.

When we compare Theorems 1 and 2, we see that Theorem 2 exhibits an extra term, $O\left(\frac{a_T}{h^4}\frac{\sum_j \omega_j^2}{n}\right)$, which is linked to the approximation of the error distribution (the use of φ_ξ instead of φ_{ξ_T} which is unknown). Moreover, we notice that both variance and bias are potentially affected by v_T. An interesting case is when $\frac{a_T}{h^4}\frac{1}{v_T^4}$ is negligible with respect to h^4. In that case, the bias does not depend on v_T and the optimal h is such that $h \sim n^{-1/8} v_T^{-3/4}$ and MISE$(\hat{f}_k) \sim n^{-1/2} v_T^{-3}$. Compared to Theorem 1, we obtain a slower rate because of the shrinking amplitude of the jump. Note that for this choice of h, a sufficient condition for $\frac{a_T}{h^4}\frac{1}{v_T^4}$ to be negligible with respect to h^4 is $n a_T v_T^2 \to 0$ which will be achieved provided T is large enough with respect to n.

6.2 Case where ω_j is estimated

Now, ω_j is replaced by its estimator $\hat{\omega}_j = \widehat{\sigma}_j^2 / \hat{\delta}_{Tj}' \widehat{Q}_j \hat{\delta}_{Tj}$. While $1/\omega_j$ can be consistently estimated at the $1/\sqrt{T}$ rate of convergence, it is not the case for ω_j because ω_j diverges with T. However, we can show that $\hat{\omega}_j/\omega_j$ converges to 1 at the rate $o_p\left(T^{-\alpha}\right)$ where α was defined in A5.

Theorem 3. Assume that, for each series j, Assumptions A1 to A5 hold. Then,

$$\frac{\hat{\omega}_j}{\omega_j} - 1 = o_p\left(T^{-\alpha}\right), \text{ for } j = 1, 2, ..., n.$$

To establish the rate of convergence of $\hat{f}_k$, some extra restrictions are needed. First, to simplify, we assume $\omega_j = \omega$ for all j and hence, ω can be estimated by $\hat{\omega} = \frac{1}{n}\sum_{j=1}^{n} \hat{\omega}_j$. Next, we impose the following assumption.

C.

$$E \int \left| \frac{\varphi_{\xi_T}(t)}{\varphi_\xi\left(\frac{\widehat{\omega} t}{\omega}\right)} - 1 \right|^2 dt \equiv b_T \to 0 \text{ as } T \text{ goes to infinity.}$$

Moreover $E\left((\hat{\omega}/\omega)^6\right) < \infty$.

Similarly to Meister (2006), we restrict the density of k to belong to a class $\mathcal{F}$ of densities with characteristic functions which decline at an arithmetic rate.

$$\mathcal{F} = \left\{ f \text{ density} : C_2 |t|^{-\beta} \geq |\varphi_k(t)| \geq C_1 |t|^{-\beta}, \text{ for all } t \text{ with } |t| > T_0 > 0 \right\}$$

with constants $C_2 > C_1 > 0$, $\beta > 1$, and $T_0 > 0$.

So, instead of assuming that f is twice differentiable as in Theorems 1 and 2, we make an assumption on the tail of its characteristic function.

Moreover, we assume K is a sinc-kernel, i.e. K is such that $\varphi_K(t) = I_{[-1,1]}(t)$. So that the estimator takes the form

$$\hat{f}_k(x) = \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \frac{1}{n} \sum_{j=1}^{n} \frac{\exp\left(it\hat{k}_j\right)}{\varphi_\xi(\hat{\omega}_j t)} dt$$

The following theorem establishes the rate of convergence of $\hat{f}_k$ toward f_k.

Theorem 4. Assume that, for each series j, Assumptions A1 to A5 hold and $(x_{jt}, \varepsilon_{jt})_j$ are independent across j. Assume $\omega_j = \omega$, $\forall j$, Assumption C holds, $f_k \in \mathcal{F}$, and K is a sinc kernel. Then,

$$\begin{aligned} \text{MISE}(\hat{f}_k) &= O\left(\frac{\omega^3}{h^4 n}\right) + O\left(h^{2\beta-1}\right) + O\left(\frac{b_T}{\omega}\right) \\ &= O\left(\frac{1}{nh^4 v_T^6}\right) + O\left(h^{2\beta-1}\right) + O\left(v_T^2 b_T\right). \end{aligned}$$

Remarks.

- If the term $O\left(v_T^2 b_T\right)$ is negligible with respect to the other terms (which is likely to happen if T is large enough compared to n), the optimal h is such as $h \sim n^{-\frac{1}{2\beta+3}} v_T^{-\frac{6}{2\beta+3}}$ and $\text{MISE}(\hat{f}_k) \sim n^{-\frac{2\beta-1}{2\beta+3}} v_T^{-\frac{6(2\beta-1)}{2\beta+3}}$.

- If we replace the class $\mathcal{F}$ by another class of densities (for instance the supersmooth densities), then the only thing that will change in Theorem 4 is the first term of the MISE, $O\left(h^{2\beta-1}\right)$, which depends on the tail behavior of the characteristic function of k. This term must be replaced by the appropriate rate.

7 Empirical Bayes

For each series j, we obtained an estimator of the break point $\widehat{k}_j$ by least-squares. However, this estimate does not take into account that the k_j are random and drawn from a common distribution. To take this information into account, we can adopt an approach borrowed from the empirical Bayes literature (see for instance, Gu and Koenker, 2017).

Let y_j denote the vector of all the observations $(y_{jt})_t$ and X_j the matrix of regressors $(x_{jt})_t$. Our objective is to find $\widetilde{k}_j$ which minimizes the squared error loss $E\left(\left(\widetilde{k}_j - k_j\right)^2 | (y_j, X_j)_{j=1,...,n}\right)$. The solution is simply $E\left(k_j | (y_j, X_j)_{j=1,...,n}\right)$. Let σ_j^2 be the variance of ε_{jt} and g denote the density of $\varepsilon_{jt}/\sigma_j$. To simplify, we assume that this density is the same for all j. Then, assuming independence of ε_{jt} across time, we have

$$E\left(k_j | (y_j, X_j)_{j=1,...,n}\right) = \frac{\int k \Pi_{t=1}^T g\left(\frac{y_{jt} - X_{jt}' \beta_{jk}}{\sigma_j}\right) \frac{1}{\sigma_j} f_k(k)\, dk}{\int \Pi_{t=1}^T g\left(\frac{y_{jt} - X_{jt}' \beta_{jk}}{\sigma_j}\right) \frac{1}{\sigma_j} f_k(k)\, dk}.$$

The traditional parametric approach would consist in replacing the density g but that of the standard normal. However, here we are able to estimate g nonparametrically. Let $\widehat{k}_j$ be the least-squares estimator of k_j, then we define $\hat{\varepsilon}_{jt} = y_j - X_j' \beta_{j\widehat{k}_j}$ and $\hat{\sigma}_j^2 = \sum_t \hat{\varepsilon}_{jt}^2 / T$. Using the rescaled residuals $\hat{\varepsilon}_{jt}/\hat{\sigma}_j$, $j = 1, 2, ..., n, t = 1, 2, ..., T$, we can estimate g by a kernel estimator $\widehat{g}$:

$$\widehat{g}(x) = \frac{1}{nT\widetilde{h}} \sum_j \sum_t K\left(\frac{\hat{\varepsilon}_{jt}/\hat{\sigma}_j - x}{\widetilde{h}}\right)$$

for some bandwidth $\widetilde{h}$ which goes to zero with the sample size. The density f_k can be estimated by $\widehat{f}_k$ given in Equation (5) with ω_j replaced by $\hat{\omega}_j$. Finally, we get an estimator of $E\left(k_j | \left(y_j, X_j\right)_{j=1,...,n}\right)$ given by

$$\hat{E}\left(k_j | \left(y_j, X_j\right)_{j=1,...,n}\right) = \frac{\int k\Pi_{t=1}^{T}\widehat{g}\left(\frac{y_{jt}-X'_{jt}\widehat{\beta}_{jk}}{\widehat{\sigma}_j}\right)\frac{1}{\widehat{\sigma}_j}\widehat{f}_k\left(k\right)dk}{\int \widehat{g}\left(\frac{y_{jt}-X'_{jt}\widehat{\beta}_{jk}}{\widehat{\sigma}_j}\right)\frac{1}{\widehat{\sigma}_j}\widehat{f}_k\left(k\right)dk}. \tag{8}$$

Equation (8) provides an estimator of the posterior mean of k_j.

8 Simulations

To evaluate the quality of our estimator, we generate simulations with $T = 150$, 300, 600 time periods, and $n = 50$, 100, 200 series. For each combination of n and T, we perform $B = 500$ replications, so that we can accurately assess the average error resulting from the estimation. We calibrate our simulations on real data from Fama-French, which will be studied in Section 9. The idea is to generate data with the same number of regressors as the Fama-French model. For each of the n series drawn, we draw X of dimension $T \times 5$ and ε of dimension $T \times 1$ according to normal distributions whose mean and variance are those of the real data for spring 2020 sample. In the same way, the break points are drawn according to a normal distribution with mean and variance equal to the empirical mean and variance found in the data. Finally, the coefficients β and δ of formula (1) are drawn in a normal distribution whose mean and variance are again given by the empirical mean and variance of the coefficients $\hat{\beta}$ and $\hat{\delta}$ estimated on the real data. This first step therefore allows us to generate data sets $(X, Y)_b$ for $b = 1, ..., B$ that are fairly similar to the actual observed data. We then apply to each of these draws the two-step density estimation method. First, we compute for each time series j the least-square estimate $\hat{k}_j$ of the break point k_j. Then, we estimate f_k by $\hat{f}_k$ given by (5) with the second-order kernel used in Delaigle and Meister (2008):

$$K_2\left(x\right) = \frac{48x(x^2 - 15)\cos x - 144\left(2x^2 - 5\right)\sin x}{\pi x^7}. \tag{9}$$

This kernel has the following bounded-support characteristic function

$$\varphi_K(t) = \left(1 - t^2\right)^3 I_{[-1,1]}(t).$$

Kernels that have a compactly supported Fourier transform are commonly used in deconvolution problems because it guarantees that assumption B3 is satisfied even for supersmooth error. Moreover, the integral calculation in (5) is faster in practice.

We then seek to evaluate the mean integrated squared error of $\hat{f}_k$. Since estimating the density f_k is computationally demanding, we approximate the integral in (6) by a grid. Hence, we simply calculate $\hat{f}_k$ for the points of a set $L = \{0, 2, 4, 6, ..., 260\}$. Then, the following quantity:

$$\widehat{MISE} = \frac{1}{Bcard(L)} \sum_{b=1}^{B} \sum_{x \in L} \left(\hat{f}_{k,b}(x) - f_k(x)\right)^2.$$

is used to estimate the MISE. Table 1 reports the values of $\widehat{MISE}$ for this first set of simulated data.

To assess the sensitivity to δ, we run a second set of simulations. The only modification is to consider δ following a normal distribution whose mean is now equal to 10 times the mean of the estimated $\widehat{\delta}_j$. Consequently, the structural shift will be more obvious for this new simulated data. The resulting MISEs are collected in Table 2.

In both cases, we observe that the magnitude $\widehat{MISE}$ decreases for increasing values of n and T. This suggests that the estimation error decreases as the number of time steps and the number of series in the panel increase. We also observe that for a small number of time steps ($T = 150$), the estimate is significantly better when the amplitude of the shift is large (see Table 2) than when it is small (see Table 1). For a small number of time periods, the extreme rupture is better captured by least squares, resulting in a better quality of the density estimate after deconvolution.

On the other hand, the estimate is not significantly better with a large shift as the number of time steps becomes large ($T = 600$). This can be explained by the result of Proposition 1 of Bai (1997), $\widehat{k} = k_0 + O_p\left(\|\delta_T\|^{-2}\right)$ which holds even for δ_T fixed. As we increase δ_T, the relative error committed by using $\widehat{k}_j$ instead of the true break point k_j becomes very small. As this estimation error becomes negligible, it is less relevant to apply the two-step deconvolution procedure. A kernel estimator applied directly on the least-squared estimates of k_j is likely to give a satisfactory estimate of f_k for a large T and large δ_T. Using the deconvolution kernel may add noise which results in a larger MISE.

	$n = 50$	$n = 100$	$n = 200$
$T = 150$	$4.050 * 10^{-6}$	$2.975 * 10^{-6}$	$2.190 * 10^{-6}$
$T = 300$	$2.774 * 10^{-6}$	$1.678 * 10^{-6}$	$1.123 * 10^{-6}$
$T = 600$	$2.557 * 10^{-6}$	$1.383 * 10^{-6}$	$8.456 * 10^{-7}$

Table 1: MISE when δ is conform to the data

	$n = 50$	$n = 100$	$n = 200$
$T = 150$	$3.563 * 10^{-6}$	$1.839 * 10^{-6}$	$1.383 * 10^{-6}$
$T = 300$	$3.204 * 10^{-6}$	$1.834 * 10^{-6}$	$1.105 * 10^{-6}$
$T = 600$	$2.969 * 10^{-6}$	$1.774 * 10^{-6}$	$1.033 * 10^{-6}$

Table 2: MISE for large δ

We perform a third simulation experiment meant to illustrate the importance of the deconvolution step when estimating the density. In this experiment, the k_j are drawn from a mixture of two equally-weighted normal distributions with means $\mu_1 = 100$ and $\mu_2 = 200$ and the same standard deviation, $\sigma = 16$. Therefore, the density f_k has two peaks. As before, we simulate panel data with $n = 400$ cross-sections and $T = 150$ time steps. Apart from the distribution of break points, the parameters are the same as for Table 1. Now we estimate f_k by two methods, a direct method (without applying deconvolution) using the conventional kernel estimator defined as

$$\widetilde{f}_{\hat{k}}(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - \hat{k}_j}{h}\right) \tag{10}$$

with a Gaussian kernel and a bandwidth h optimally chosen as in Wand and Jones (1994). The second method is the two-stage deconvolution with the K_2 kernel. Figure

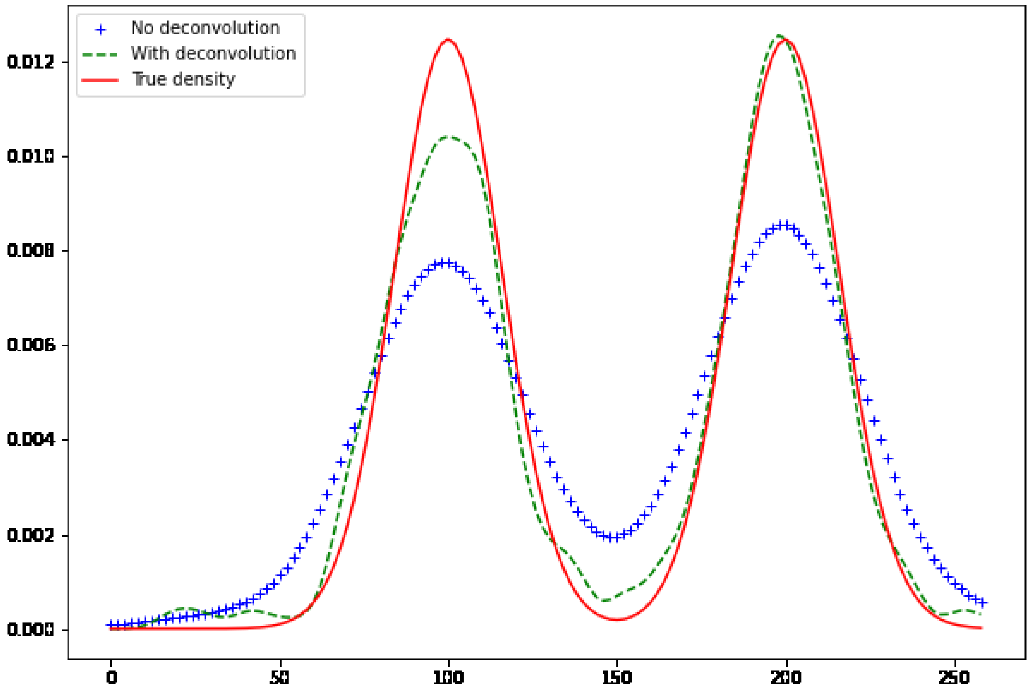

Figure 2: Estimation of bimodal density with and without deconvolution

2 displays the plot of the two densities. It is evident that when estimated by a direct method, the density is less accurately estimated compared to when applying a two-stage deconvolution method. In particular the density appears flattened for the conventional kernel estimator. Conversely, the deconvolution kernel allows us to better capture the two distinct peaks in the bimodal density.

9 Application to portfolio returns

In order to illustrate the method on real data, we have chosen to apply it to the Fama-French five-factor model during two extreme events. The selected periods are the first

Covid lockdown of spring 2020 and the Ukraine Invasion in spring 2022. Our hypothesis is that these events have had an impact on financial markets, which manifests itself in structural breaks in the regression coefficients.

9.1 Presentation of the Fama-French five-factor model

The Fama-French five-factor asset pricing model proposed by Fama and French (2015) is an empirical explanation of the return of a financial asset j at time t. It is written as follows:

$$R_{jt} - R_{Ft} = a_j + b_j(RM_t - RF_t) + s_j SMB_t + h_j HML_t + r_j RMW_t + c_j CMA_t + e_{it}$$

where R_{jt} is the return of asset j at time t, R_{Ft} is the risk-free return, R_{Mt} is the market return, SMB_t (Small Minus Big) is the difference between the return of a diversified portfolio of small company stocks and that of large company stocks, indicating the outperformance of small companies compared to large ones, and HML_t (High Minus Low) is the difference between the returns of a diversified portfolio of high-cap stocks and low-cap stocks. In addition to these factors, which form the foundation of the three-factor Fama-French model, the authors introduced RMW_t (Robust Minus Weak) and CMA_t (Conservative Minus Aggressive), which respectively measure the difference in returns between the most profitable and least profitable companies, and between the most conservative and most aggressive companies in their investments.

Throughout our work, we rely on data from Kenneth French's website. These data include the returns of a certain number of portfolios for different geographical areas, business types, etc. Since we aim for the portfolios under study to have the lowest possible correlations with each other, it is important to ensure that the intersection between two portfolios is empty. Therefore, we have chosen portfolios grouped by geographical area, company size, and momentum (size and momentum). Thus, we consider 25 Asian portfolios (excluding Japan), 25 European portfolios, 25 Japanese portfolios, and 25 North American portfolios, for a total of $n = 100$ portfolios.

In addition to these portfolio returns, the Kenneth French's website provides the 5 factors of the model for each of the considered geographical areas. For each portfolio j, we regress the excess returns on the 5 Fama-French factors. We allow for a single break point which affects at the same time the intercept and the regression coefficients. This break point, as well as the intercept and regression coefficients, are estimated separately for each portfolio.

9.2 Spring 2020: Effect of Covid pandemic

We apply our estimator to the Fama-French data around the spring lockdown of 2020. This period is particularly interesting as the lockdown measures were implemented at very close dates worldwide, and the pandemic affected all geographical areas. Consequently, we can apply our estimator to the $n = 100$ portfolios, assuming that the jump times are drawn from the same distribution, regardless of the geographical zone.

We consider daily observations ranging from 11/01/2019 to 09/01/2020, resulting in $T = 241$ dates, accounting for market closure days. This relatively broad range around the break dates allows for a good estimation of coefficients before and after the shocks. We start by ensuring that the jumps are statistically significant for the $n = 100$ considered portfolios. We apply the sup-Wald test of Andrews (1993) to test the null hypothesis $H_0 : \delta_j = 0$. We exclude $\pi = 5\%$ of the possible jump points at the beginning and at the end of the interval. The table of values provided by Andrews (2003) ensures that we can reject the null hypothesis (of no structural change) at the 5% level with $p = 6$ affected coefficients and $\pi = 5\%$ if the test statistics exceeds 21.56. We note that out of the 100 studied portfolios, only 4 do not allow us to reject the null hypothesis at the 5% level. Therefore, panel data indeed exhibit significant structural breaks during the period considered. We apply the method to all series including those for which we cannot reject the null hypothesis of parameter stability.

We implement two kernels:

The Gaussian kernel: $K_1(x) = 1/\sqrt{2\pi}\exp\left(-x^2/2\right)$, with $\varphi_{K_1} = exp(-x^2/2)$.

The K_2 kernel defined in (9).

We also choose the optimal bandwidth $\widehat{h}$ using the formula (7).

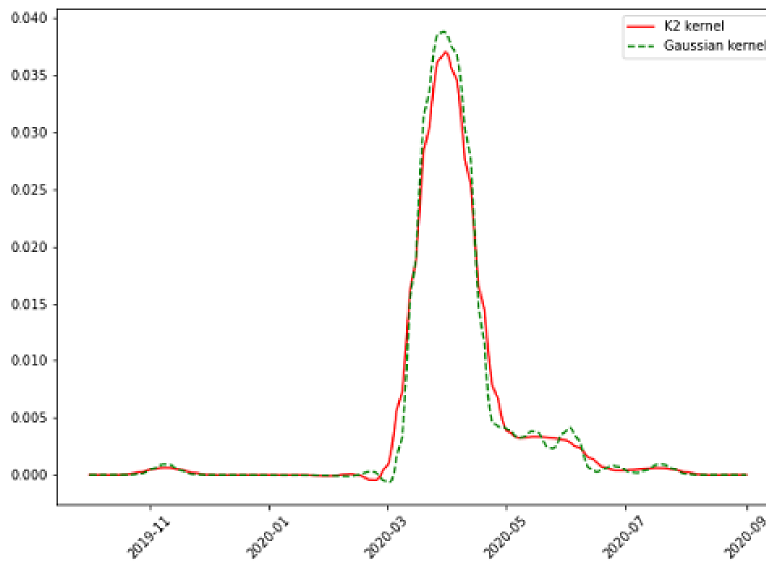

Figure 3: Density of break points during Covid

Figure 3 displays the estimated densities of the break points obtained for the two kernels. The two densities are very similar for the two kernels with K_2 being slightly smoother. Moreover, we observe that most of the breaks occur after March 13th, which corresponds to the initial lockdown. This suggests that the markets did not anticipate these events. Moreover, we observe a distribution of structural break points concentrated between March and April 2020. This is in line with the duration of the initial spring 2020 lockdown.

We now turn our attention to the posterior means of structural break points. This approach leverages the assumption that the break points follow the same distribution. To perform the Bayesian reestimation of structural break points, we first compute the deconvolution estimator $\widehat{f}_k$ using the K_2 kernel. We then incorporate this estimator $\widehat{f}_k$ into Equation (8), yielding new values for $\widehat{k}_j$. Figure 4 plots the histograms of the posterior means of break points for the 100 studied portfolios grouped into 4 regions.

We observe that the Asian markets (excluding Japan) exhibit a highly concentrated

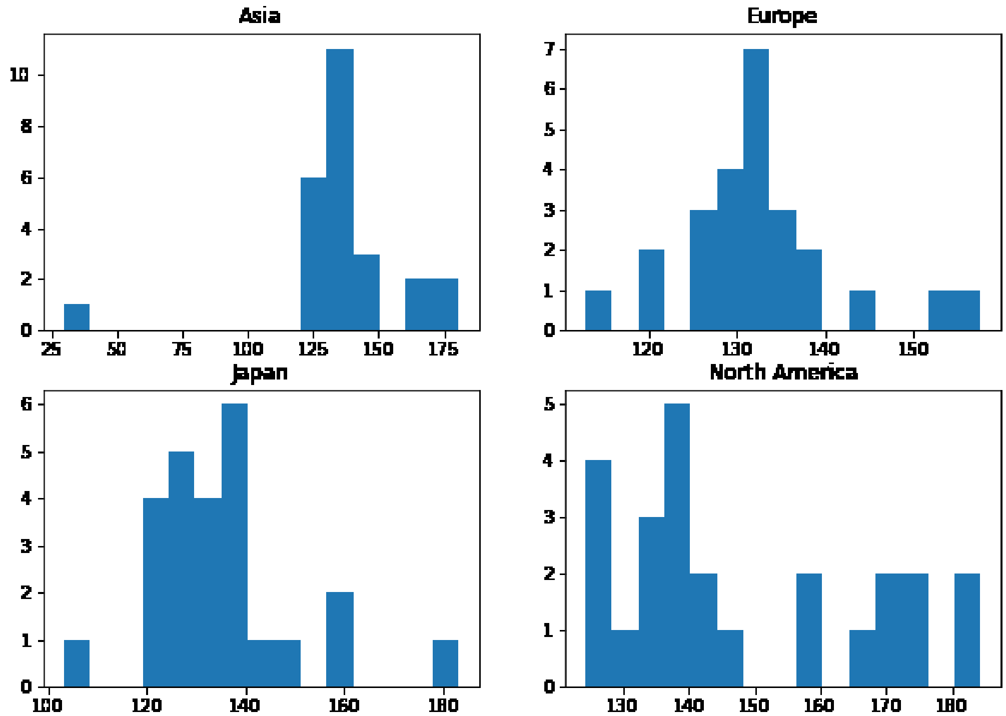

Figure 4: Posterior means of breakpoints by regions

distribution of structural break points between time steps 120 and 140 (corresponding to dates from March 17th to April 14th, 2020). The same pattern is evident for European and Japanese data. This does indeed seem to align with the impact of the initial lockdown measures. In contrast, the American data reveals a much less localized distribution of break point instances, with a primary break around time step 140 (mid-April) and subsequent break points occurring between time steps 160 and 185 (from May 12th to June 16th). This suggests a delayed effect of the crisis, with break points occurring later in North America than in other regions of the world.

9.3 Spring 2022: Effect of the Ukraine invasion

We apply the same methodology as before to data from spring 2022, assuming again that an international political event, the invasion of Ukraine in this case, will impact financial markets, leading to changes in the Fama-French coefficients. We analyze data from 06/01/2021 to 01/01/2023, covering a span of one and a half years.

Similar to before, we conduct the sup-Wald test while excluding 5% of observations at the beginning and end of the interval. This time, the null hypothesis, indicating the absence of structural breaks, is rejected slightly less often compared to the case of the 2020 lockdown. Specifically, out of the 100 time series in the panel, 13 do not reject the null hypothesis at the 5% level. However, a structural break in the Fama-French model in the spring of 2022 is plausible for nearly 90% of the series, which is still satisfactory. Here again, we keep all series when we apply the estimation procedure of the change-point density.

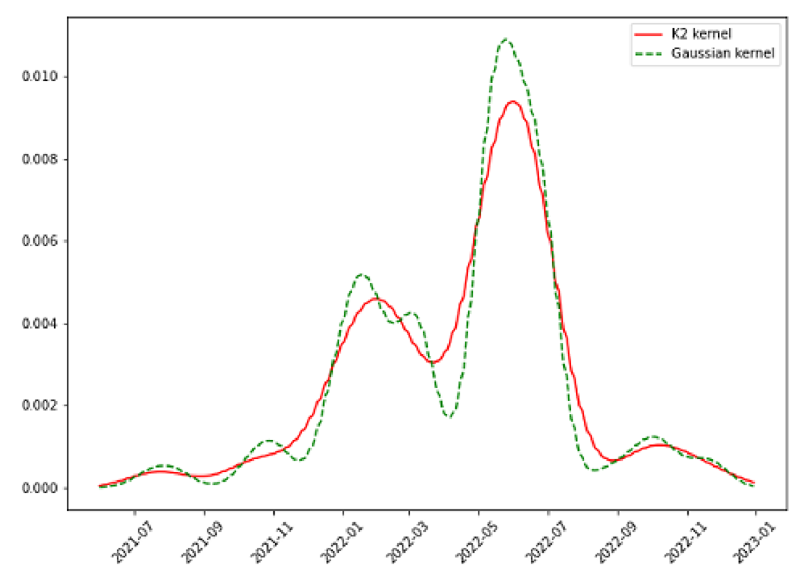

Figure 5: Density of break points during 2022

Figure 5 shows the plots of the density estimators for the two kernels K_1 and K_2, with optimal bandwidths. Once more, we observe that the results produced by the two kernels are close to each other. However, we note that the shock is much less localized than in the spring of 2020. Specifically, we observe two peaks in the distribution of jump times. One begins in November 2021 and peaks in February 2022, while the second, more pronounced, peak reaches its maximum in June 2022.

Further analysis of the data reveals that the first of the two peaks in the distribution is primarily observed in the Asian data. This peak appears to correspond to China's reinforcement of its zero-Covid policy following the arrival of Omicron end of 2021. On the other hand, the second peak is more pronounced in Europe and the United States. This period corresponds to a period of high uncertainty in financial markets linked to the surge of oil prices resulting from the invasion of Ukraine.

This observation challenges the assumption of homogeneity of panel data since Asian and European data exhibit distinct behaviors. Consequently, the assumption that structural break points are drawn from the same distribution for all portfolios is questionable. Generally, it is important to note that multiple shocks can overlap and cause differentiated structural breaks in a panel dataset. When the estimated distribution of structural break points shows multiple peaks, it is necessary to consider possible heterogeneity and the presence of multiple structural breaks.

As our estimator estimates precisely the density f_k, it permits to identify the presence of distinct peaks. Figure 6 plots the estimated densities using the deconvolution estimator and the Gaussian kernel without deconvolution. We observe that without the deconvolution step, the estimated density is much broader, making it quite challenging to conclude regarding the existence of one or multiple distinct shocks. However, with the application of deconvolution, the two peaks are clearly visible, making it easy to recognize the presence of two distinct break points.

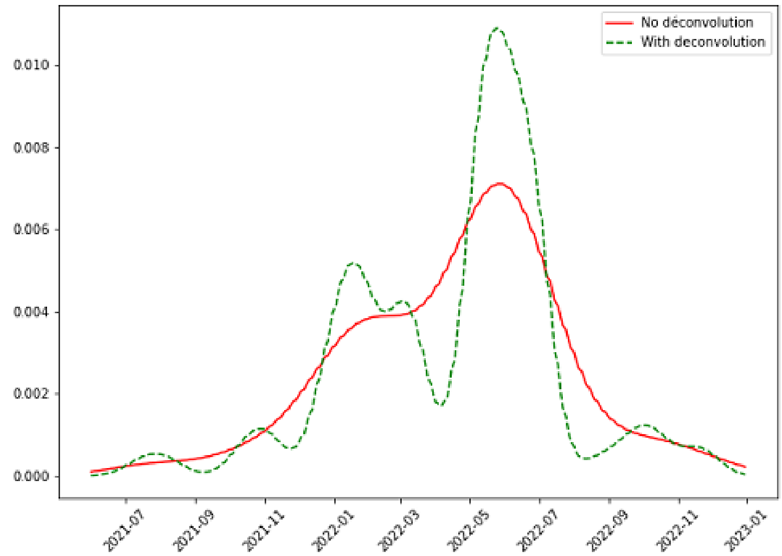

Figure 6: Density of breakpoints during 2022 with and without deconvolution

10 Conclusion

The paper shows how to estimate the density of break points using deconvolution kernel. Deconvolution kernel estimation has often been criticized dues to its slow rate of convergence when the error is normally distributed. Here, in this case, the error is much more concentrated around 0, resulting in a faster rate of convergence. Moreover, by employing empirical Bayes, we can estimate posterior means of the break points leveraging our density estimator.

Simulations with a bimodal density shows that our method allows us to capture the two humps in the density, which may be overlookedwhen using a standard kernel estimator. The application on Fama-French five-factor model reveals that Covid and the Ukraine war had an impact on the coefficients.

Our method could be employed for policy evaluation. For instance, it could be used to observe how quickly the agents react to a policy change. The presence of a peak following a policy decision can indicate an effect of that measure.

Some of our results could be applied to alternative models, particularly the Threshold Autoregressive and sample splitting models. Indeed, the Picard-Yao distribution (2) emerges as the asymptotic distribution of the threshold estimation error in Hansen (2000). Therefore, the deconvolution kernel presented here could be adapted to the estimation of the density of thresholds assuming a panel of observations is available and the thresholds are drawn from some common distribution.

11 Proofs

Proof of Lemma 1. Given the density f_ξ is even, its characteristic function is given by

$$\begin{aligned}\varphi_\xi(t) &= 2\int_0^\infty \cos(tx) f_\xi(x)dx \\ &= 3\int_0^\infty \cos(tx)e^x \Phi\left(-\tfrac{3}{2}\sqrt{x}\right)dx - \int_0^\infty \cos(tx)\Phi\left(-\tfrac{\sqrt{x}}{2}\right)dx.\end{aligned}$$

First consider the second term. Integration by parts gives:
$\int_0^\infty \cos(tx)\Phi\left(-\frac{\sqrt{x}}{2}\right)dx = \left[\frac{\sin(tx)}{t}\Phi\left(-\frac{\sqrt{x}}{2}\right)\right]_0^\infty + \int_0^\infty \frac{1}{4t\sqrt{x}}\phi\left(\frac{\sqrt{x}}{2}\right)\sin(tx)dx$ where ϕ denotes the pdf of the standard normal. The term within brackets is equal to zero as x goes to infinity since $|\sin(tx)| \leq 1$. Using the change of variables $u = \sqrt{x}$, we obtain
$= \frac{1}{2t}\int_0^\infty \phi\left(\frac{u}{2}\right)\sin(tu^2)du$
$= \frac{1}{4it\sqrt{2\pi}}\left(\int_0^\infty \left\{e^{-\frac{u^2}{8}(1-8it)} - e^{-\frac{u^2}{8}(1+8it)}\right\}du\right).$

Denote $g(t) = \int_0^\infty e^{-\frac{u^2}{8}(1-8it)}du$. Deriving g and an integration by parts show that $g(t)$ satisfies a differential equation

$$g'(t) = \frac{4i}{1-8it}g(t) = \frac{4i-32t}{1+64t^2}g(t)$$

Let $g(t) = g_1(t) + ig_2(t)$, g_1 and g_2 are real and satisfy the following system

$$\begin{cases} g_1'(t) = \frac{-4g_2(t)-32tg_1(t)}{1+64t^2} \\ g_2'(t) = \frac{4g_1(t)-32tg_2(t)}{1+64t^2} \end{cases}$$

Then $g_1(t) = \frac{\sqrt{1+\sqrt{1+64t^2}}}{\sqrt{2}\sqrt{1+64t^2}}g(0)$ and $g_2(t) = \frac{\sqrt{\sqrt{1+64t^2}-1}}{\sqrt{2}\sqrt{1+64t^2}}g(0)$, with $g(0) = \int_0^\infty e^{-\frac{u^2}{8}}du = \sqrt{2\pi}$. Moreover, $g(t)$ takes the simple form $\frac{\sqrt{2\pi}}{\sqrt{1-8it}}$. This shows that

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{u^2}{8}(1-8it)}du &= \frac{1}{\sqrt{1-8it}} = A + iB \text{ with} \\ A &= \frac{\sqrt{1+\sqrt{1+64t^2}}}{\sqrt{2}\sqrt{1+64t^2}} \text{ and } B = \frac{\sqrt{\sqrt{1+64t^2}-1}}{\sqrt{2}\sqrt{1+64t^2}}.\end{aligned}$$

By the same way, we obtain
$\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{u^2}{8}(1+8it)}du = \frac{1}{\sqrt{1+8it}} = A - iB.$

Then, we have

$\int_0^\infty \cos(tx)\Phi\left(-\frac{\sqrt{x}}{2}\right)dx = \frac{1}{4it}\left[\frac{1}{\sqrt{1-8it}} - \frac{1}{\sqrt{1+8it}}\right] = B/(2t).$

Now consider the first part of the characteristic function:

$\int_0^\infty \cos(tx)e^x\Phi\left(-\frac{3}{2}\sqrt{x}\right)dx$

$=\left[\frac{e^x}{1+t^2}(\cos(tx)+t\sin(tx))\Phi\left(-\frac{3}{2}\sqrt{x}\right)\right]_0^\infty + \int_0^\infty \frac{e^x}{1+t^2}(\cos(tx)+t\sin(tx))\frac{3}{4\sqrt{x}}\phi\left(-\frac{3}{2}\sqrt{x}\right)dx$

using an integration by parts. The term in brackets tends to zero as x tends to infinity, and is equal to $-\frac{1}{2(1+t^2)}$ as $x=0$, so we focus our attention on the second term which, using the change of variables $u=\sqrt{x}$, becomes

$\frac{1}{1+t^2}\int_0^\infty e^{u^2}\left(\cos(tu^2)+t\sin(tu^2)\right)\frac{3}{2}\phi\left(\frac{3}{2}u\right)du$

$=\frac{3}{4}\frac{1}{1+t^2}\frac{1}{\sqrt{2\pi}}\left\{\left(1+\frac{t}{i}\right)\int_0^\infty e^{-\frac{u^2}{8}(1-8it)}du + \left(1-\frac{t}{i}\right)\int_0^\infty e^{-\frac{u^2}{8}(1+8it)}du\right\}$

$=\frac{3}{4}\frac{1}{1+t^2}\left\{\left(1+\frac{t}{i}\right)\frac{1}{\sqrt{1-8it}} + \left(1-\frac{t}{i}\right)\frac{1}{\sqrt{1+8it}}\right\}$

After simplification, we obtain the form given in Lemma 1.

The proof of Theorem 1 uses the following result.

Lemma 2. Assume ω fixed or ω/h goes to infinity. Then,

$$\int_0^\infty \frac{\varphi_K(x)^2}{\varphi_\xi\left(\frac{\omega x}{h}\right)^2}dx \sim \frac{\omega^3}{h^3}\int_0^\infty x^3\varphi_K(x)^2dx \quad \text{as } h \text{ goes to zero.}$$

Proof of Lemma 2. Without loss of generality, let $\omega=1$. The relation that we want to show is the following:

$$\int_0^\infty \frac{\varphi_K^2(x)}{\varphi_\xi\left(\frac{x}{h}\right)^2}dx \sim \frac{1}{h^3}\int_0^\infty x^3\varphi_K^2(x)dx \qquad \text{as } h \text{ goes to zero.}$$

From the expression of φ_ξ, one can deduce immediately an equivalent when x goes to infinity:

$$\varphi_\xi(x) \sim x^{-\frac{3}{2}}.$$

Let x be fixed such that $a<x<b$, for any $0<a<b$ and ε, $0<\varepsilon<1$, there is some number h^* such that for all $h \le h^*$, we have

$$(1-\varepsilon)\frac{x^3}{h^3} \le \frac{1}{\varphi_\xi\left(\frac{x}{h}\right)^2} \le (1+\varepsilon)\frac{x^3}{h^3}$$

It follows that

$$(1-\varepsilon)\frac{1}{h^3}\int_a^b x^3\varphi_K^2(x)\,dx \le \int_a^b \frac{\varphi_K^2(x)}{\varphi_\xi\left(\frac{x}{h}\right)^2}\,dx \le (1+\varepsilon)\frac{1}{h^3}\int_a^b x^3\varphi_K^2(x)\,dx$$

Since $\int_0^\infty x^3\varphi_K^2(x)\,dx$ converges by Assumption B6, we can make a tend to zero and b tend to infinity. Then, the result follows.

Proof of Theorem 1.

Assumptions B4 and B5 imply that K_j^* is bounded, so $\hat{f}_k$ is also bounded and its expectation necessarily exists. We examine first the bias, second the variance.

Bias. By Fubini's theorem and B4-B5, we can interchange the expectation and integration to obtain

$$\begin{aligned}
E\left[K_j^*\left(\frac{\hat{k}_j-x}{h}\right)|k_j\right] &= \frac{1}{2\pi}\int E\left(\exp iy\left(\frac{\hat{k}_j-x}{h}\right)|k_j\right)\frac{\varphi_K(y)}{\varphi_\xi\left(\omega_j y/h\right)}dy \\
&= \frac{1}{2\pi}\int e^{-iyx/h}E\left(\exp iy\left(\frac{k_j+\omega_j\xi_j}{h}\right)|k_j\right)\frac{\varphi_K(y)}{\varphi_\xi\left(\omega_j y/h\right)}dy \\
&= \frac{1}{2\pi}\int e^{iy(k_j-x)/h}\varphi_K(y)dy \\
&= K\left(\frac{k_j-x}{h}\right)
\end{aligned}$$

where the last equality follows from the Fourier inversion formula. Hence, we have

$$\begin{aligned}
E\left(\hat{f}_k\left(x\right)\right) &= \frac{1}{nh}\sum_{j=1}^n E\left[K_j^*\left(\frac{\hat{k}_j-x}{h}\right)\right] \\
&= \frac{1}{nh}\sum_{j=1}^n E\left[K\left(\frac{k_j-x}{h}\right)\right] \\
&= \frac{1}{h}\int K\left(\frac{u-x}{h}\right)f_k\left(u\right)du \\
&= \left(K_h * f_k\right)(x)
\end{aligned}$$

where $K_h\left(.\right)=K\left(./h\right)/h$. Therefore, the bias of $\hat{f}_k$ depends neither on ω_j nor on the distribution of ξ. It coincides with the bias of the ordinary kernel estimator. Therefore, we can exploit well-known results on the bias of kernel estimators which yield the following rate provided the second-order derivative of f_k is squared integrable (see for instance Wand and Jones (1995)):

$$\int \left[E\left(\hat{f}_k(x)\right) - f_k(x)\right]^2 dx = \frac{h^4}{4}\mu_2(K)^2 \int f_k''(x)^2 dx + o\left(h^4\right).$$

Variance.

$$\int V\left(\hat{f}_k(x)\right) dx = \frac{1}{n^2h^2}\sum_{j=1}^{n}\int V\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right) dx$$

where

$$V\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right) = E\left[\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right)^2\right] - \left[E\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right)\right]^2.$$

We have $E\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right) = h\left(K_h * f_k\right)(x) = O(h)$, therefore the second term on the r.h.s. will be negligible. We focus on the first term. By a change of variables $t = \left(\hat{k}_j - x\right)/h$, we have

$$\begin{aligned}
\frac{1}{n^2h^2}\sum_{j=1}^{n}\int E\left[\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right)^2\right] dx &= \frac{1}{n^2h^2}\sum_{j=1}^{n} E\int\left(K_j^*\left(\frac{\hat{k}_j - x}{h}\right)\right)^2 dx \\
&= \frac{1}{n^2h}\sum_{j=1}^{n}\int K_j^*(t)^2\, dt.
\end{aligned}$$

By Parseval's Identity, we have

$$\begin{aligned}
\int K_j^*(t)^2\, dt &= \int\left(\frac{1}{2\pi}\int e^{ity}\frac{\varphi_K(y)}{\varphi_\xi\left(\omega_j y/h\right)} dy\right)^2 dt \\
&= \frac{1}{2\pi}\int\frac{\varphi_K^2(y)}{\varphi_\xi^2\left(\omega_j y/h\right)} dy.
\end{aligned}$$

Hence,

$$\int V\left(\hat{f}_k\left(x\right)\right)dx = \frac{1}{n^2 h 2\pi}\sum_{j=1}^{n}\int \frac{\varphi_K^2(y)}{\varphi_\xi^2\left(\omega_j y/h\right)}dy + O\left(\frac{1}{n}\right).$$

We obtain

$$\begin{aligned} MISE &= \int\left(E\left(\hat{f}_k\left(x\right) - f\left(x\right)\right)\right)^2 dx + \int V\left(\hat{f}_k\left(x\right)\right)dx \\ &= \frac{h^4}{4}\mu_2(K)^2\int f_k''(x)^2 dx + \frac{1}{n^2\pi h 2}\sum_{j=1}^{n}\int \frac{\varphi_K^2(x)}{\varphi_\xi^2\left(\omega_j x/h\right)}dx + o\left(h^4\right) + o\left(\frac{1}{nh}\right). \end{aligned}$$

Then, using Lemma 2, the asymptotic MISE is given by

$$\text{AMISE} = \frac{h^4}{4}\mu_2(K)^2\int (f_k''(x))^2 dx + \frac{1}{n^2\pi h^4}\left(\sum_{j=1}^{n}\omega_j^3\right)\int_0^\infty x^3\varphi_K^2(x)dx.$$

The result follows.

Proof of Theorem 2.

We use the notation $\widehat{k}_j = k_j + \omega_j\xi_T$. The density of ξ_T is denoted f_{ξ_T}. We have

$$K_j^*\left(\frac{\widehat{k}_j - x}{h}\right) = \frac{1}{2\pi}\int \exp\left(iy\left(\frac{\widehat{k}_j - x}{h}\right)\right)\frac{\varphi_K(y)}{\varphi_\xi\left(\omega_j y/h\right)}dy.$$

Hence,

$$\begin{aligned} E\left[K_j^*\left(\frac{\widehat{k}_j - x}{h}\right)|k_j\right] &= \frac{1}{2\pi}\int e^{iy\left(\frac{k_j - x}{h}\right)}\frac{\varphi_K(y)}{\varphi_\xi\left(\omega_j y/h\right)}\varphi_{\xi_T}\left(\omega_j y/h\right)dy \\ &= \frac{1}{2\pi}\int e^{iy\left(\frac{k_j - x}{h}\right)}\frac{\left[\varphi_{\xi_T}\left(\omega_j y/h\right) - \varphi_\xi\left(\omega_j y/h\right)\right]}{\varphi_\xi\left(\omega_j y/h\right)}\varphi_K(y)dy \\ &\quad + \frac{1}{2\pi}\int e^{iy\left(\frac{k_j - x}{h}\right)}\varphi_K(y)dy. \\ &= \frac{1}{2\pi}\int e^{iy\left(\frac{k_j - x}{h}\right)}\frac{\left[\varphi_{\xi_T}\left(\omega_j y/h\right) - \varphi_\xi\left(\omega_j y/h\right)\right]}{\varphi_\xi\left(\omega_j y/h\right)}\varphi_K(y)dy \\ &\quad + K\left(\frac{k_j - x}{h}\right). \end{aligned}$$

Squared bias.

By Fubini's theorem, we have

$$\begin{aligned} & E\left(\hat{f}_k(x)\right) \\ = \; & \frac{1}{hn}\frac{1}{2\pi}\sum_j \int\int e^{iy\left(\frac{k-x}{h}\right)} f_k(k)\, dk \frac{\left[\varphi_{\xi_T}(\omega_j y/h) - \varphi_\xi(\omega_j y/h)\right]}{\varphi_\xi(\omega_j y/h)}\varphi_K(y)dy \qquad (11) \\ & + (K_h * f_k)(x) \qquad (12) \end{aligned}$$

where $K_h = K(./h)/h$. Hence

$$\int \left(E\left(\hat{f}_k(x)\right) - f_k(x)\right)^2 dx \leq 2\int (11)^2 dx + 2\int \left((K_h * f_k)(x) - f_k(x)\right)^2 dx.$$

We have

$$\begin{aligned} (11) &= \frac{1}{2\pi nh}\sum_j \int e^{-iyx/h}\varphi_k(y/h)\frac{\left[\varphi_{\xi_T}(\omega_j y/h) - \varphi_\xi(\omega_j y/h)\right]}{\varphi_\xi(\omega_j y/h)}\varphi_K(y)dy \\ &= \frac{1}{2\pi nh}\sum_j \int e^{itx/h}\varphi_k(-t/h)\frac{\left[\varphi_{\xi_T}(-\omega_j t/h) - \varphi_\xi(-\omega_j t/h)\right]}{\varphi_\xi(-\omega_j t/h)}\varphi_K(-t)dt \end{aligned}$$

by a change of variables $t = -y$. By Parseval's identity,

$$\begin{aligned} \int (11)^2 dx &\leq \frac{1}{2\pi nh^2}\sum_j \int \varphi_k(-y/h)^2 \frac{\left[\varphi_{\xi_T}(-\omega_j y/h) - \varphi_\xi(-\omega_j y/h)\right]^2}{\varphi_\xi(-\omega_j y/h)^2}\varphi_K(-y)^2 dy \\ &\leq \frac{1}{2\pi nh^2}\sum_j \sup_{t\in[-1,1]}\frac{1}{\varphi_\xi(-\omega_j t/h)^2}\int_{-1}^{1}\left[\varphi_{\xi_T}(-\omega_j y/h) - \varphi_\xi(-\omega_j y/h)\right]^2\varphi_K(-y)^2 dy \end{aligned}$$

using the compactness of the support of $\varphi_K(-y)$. Moreover,
$\int_{-1}^{1}\left[\varphi_{\xi_T}(-\omega_j y/h) - \varphi_\xi(-\omega_j y/h)\right]^2\varphi_K(-y)^2 dy \leq 4\int_{-1}^{1}\varphi_K(-y)^2 dy < \infty$. So by the

dominated convergence theorem,

$$\begin{aligned}
& \lim_{T\to\infty}\int_{-1}^{1}\left[\varphi_{\xi_T}\left(-\omega_j y/h\right)-\varphi_{\xi}\left(-\omega_j y/h\right)\right]^2\varphi_K(-y)^2dy \\
= & \int_{-1}^{1}\left\{\lim_{T\to\infty}\left[\varphi_{\xi_T}\left(-\omega_j y/h\right)-\varphi_{\xi}\left(-\omega_j y/h\right)\right]^2\right\}\varphi_K(-y)^2dy \\
= & \ 0.
\end{aligned}$$

However, the term $\frac{1}{h^2}\sup_{t\in[-1,1]}\frac{1}{\varphi_\xi(\omega_j t/h)^2}=\frac{\omega_j^3}{h^5}$ by Lemma 2 and hence diverges when $h\to 0$. In absence of a rate of convergence of the characteristic function, it is not possible to characterize the rate of convergence of the bias.

Denote $\int_{-1}^{1}\left[\varphi_{\xi_T}\left(-\omega_j y/h\right)-\varphi_{\xi}\left(-\omega_j y/h\right)\right]^2\varphi_K(-y)^2dy=\widetilde{a}_{Tj}$. We have

$$\begin{aligned}
\widetilde{a}_{Tj} &\leq C\int\left[\varphi_{\xi_T}\left(-\omega_j y/h\right)-\varphi_{\xi}\left(-\omega_j y/h\right)\right]^2dy \\
&= C\frac{h}{\omega_j}\int\left[\varphi_{\xi_T}\left(u\right)-\varphi_{\xi}\left(u\right)\right]^2du \\
&= C'\frac{h}{\omega_j}\int\left[f_{\xi_T}\left(x\right)-f_{\xi}\left(x\right)\right]^2dx \\
&= C'\frac{h}{\omega_j}a_T
\end{aligned}$$

for some constants $C,C'>0$, where the first equality comes from a change of variables and the second equality from Parseval's identity. Then the squared bias is $O\left(\frac{a_T}{h^4}\frac{\sum_j\omega_j^2}{n}\right)+O\left(h^4\right)=O\left(\frac{a_T}{h^4}\frac{1}{v_T^4}\right)+O\left(h^4\right)$. Under the assumption, $\frac{a_T}{h^4}\frac{1}{v_T^4}\to 0$, the bias goes to zero. Given h depends on n only, this imposes restrictions on the growth rate of T relative to n.

Variance.

$$\begin{aligned}
V\left(\hat{f}_k\left(x\right)\right) &= \frac{1}{n^2h^2}\sum_j V\left(K_j^*\left(\frac{\widehat{k}_j-x}{h}\right)\right) \\
&\leq \frac{1}{n^2h^2}\sum_j E\left[\left(K_j^*\left(\frac{\widehat{k}_j-x}{h}\right)\right)^2\right].
\end{aligned}$$

$$
\begin{aligned}
E\left[K_j^{*2}\left(\frac{\widehat{k}_j - x}{h}\right)\right] &= E\left[E\left[K_j^{*2}\left(\frac{\widehat{k}_j - x}{h}\right)|k_j = k\right]\right] \\
&= \int\left(\int K_j^{*2}\left(\frac{k+\omega_j\xi - x}{h}\right) f_{\xi_T}(\xi)\, d\xi\right) f_k(k)\, dk \\
&= \int\left(\int K_j^{*2}\left(\frac{k+\omega_j\xi - x}{h}\right) f_k(k)\, dk\right) f_{\xi_T}(\xi)\, d\xi
\end{aligned}
$$

by Fubini. By a change of variables $t = (k + \omega_j \xi - x)/h$, we obtain

$$
E\left[K_j^{*2}\left(\frac{\widehat{k}_j - x}{h}\right)\right] = h\int\left(\int K_j^{*2}(t)\, f_k(th - \omega_j\xi + x)\, dt\right) f_{\xi_T}(\xi)\, d\xi.
$$

Denote

$$
A(h,a) = \frac{\int K_j^{*2}(t)\, f_k(th+a)\, dt}{\int K_j^{*2}(t)\, dt}
$$

and notice that $A(h,a) \leq \sup_x f_k(x) \equiv B_{f_k} < \infty$ by B2. Hence, we have

$$
\begin{aligned}
E\left[K_j^{*2}\left(\frac{\widehat{k}_j - x}{h}\right)\right] &= h\left(\int A(h, x-\omega_j\xi)\, f_{\xi_T}(\xi)\, d\xi\right)\int K_j^{*2}(t)\, dt \\
&\leq hB_{f_k}\int K_j^{*2}(t)\, dt \\
&= hB_{f_k}\frac{1}{2\pi}\int \frac{\varphi_K^2(y)}{\varphi_\xi^2(\omega_j y/h)} dy
\end{aligned}
$$

by Parseval's identity. Therefore,

$$
\begin{aligned}
V\left(\hat{f}_k(x)\right) &\leq \frac{B_{f_k}}{n^2 h}\sum_j \frac{1}{2\pi}\int\frac{\varphi_K^2(y)}{\varphi_\xi^2(\omega_j y/h)} dy \\
&\leq C\frac{1}{n^2h^4}\sum_j \omega_j^3 = O\left(\frac{1}{nh^4 v_T^6}\right)
\end{aligned}
$$

by Lemma 2. So combining these results, we obtain

$$
MISE = O\left(\frac{a_T}{(hv_T)^4}\right) + O\left(h^4\right) + O\left(\frac{1}{nh^4 v_T^6}\right).
$$

Proof of Theorem 3.

From Corollary 1 of Bai (1997), $\hat{\beta} - \beta = O_p(\frac{1}{\sqrt{T}})$ and $\hat{\delta} - \delta_T = O_p(\frac{1}{\sqrt{T}})$. Moreover, Proposition 1 of Bai (1997) states that

$$\hat{k} = k_0 + O_p(||\delta_T||^{-2}). \tag{13}$$

Limit of $\hat{\sigma}^2$:

Let $Y = (y_1, y_2, ..., y_T)'$, $X = (x_1', x_2', ..., x_T')'$, and $Z = (z_1', z_2', ..., z_T')'$. As in Bai (1997), we define R as the full ranked matrix such that $z_t = R'x_t$. Let $\hat{X} = \left(0, ..., 0, x_{\widehat{k}+1}, ..., x_T\right)'$ and $\widehat{Z} = \widehat{X}R$.

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T}(Y - X\hat{\beta} - \hat{Z}\hat{\delta})'(Y - X\hat{\beta} - \hat{Z}\hat{\delta}) \\
&= \frac{1}{T}(Y - X\beta - Z\delta_T - X(\hat{\beta} - \beta) - (\hat{Z} - Z)\delta_T - \hat{Z}(\hat{\delta} - \delta_T))' \\
&\quad (Y - X\beta - Z\delta_T - X(\hat{\beta} - \beta) - (\hat{Z} - Z)\delta_T - \hat{Z}(\hat{\delta} - \delta_T)) \\
&= \frac{1}{T}(Y - X\beta - Z\delta_T)'(Y - X\beta - Z\delta_T) \\
&\quad - \frac{2}{T}(Y - X\beta - Z\delta_T)'(X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T)) \\
&\quad + \frac{1}{T}(X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T))'(X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T)).
\end{aligned}$$

By the law of large numbers, $\frac{1}{T}(Y - X\beta - Z\delta_T)'(Y - X\beta - Z\delta_T) \to \sigma^2$ in probability. Moreover,

$$\begin{aligned}
& \left\|(Y - X\beta - Z\delta_T)'(X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T))\right\| \\
\leq\ & \|Y - X\beta - Z\delta_T\| \ \left\|X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T)\right\| \\
\leq\ & \|Y - X\beta - Z\delta_T\| \ \left(\left\|X(\hat{\beta} - \beta)\right\| + \left\|(\hat{Z} - Z)\delta_T\right\| + \left\|\hat{Z}(\hat{\delta} - \delta_T)\right\|\right).
\end{aligned} \tag{14}$$

Regarding the first term,

$$\left\|\frac{1}{\sqrt{T}}(X(\hat{\beta} - \beta))\right\| \leq \frac{1}{\sqrt{T}}||X|| \ ||\hat{\beta} - \beta||.$$

As $\frac{1}{\sqrt{T}}||X||$ converges in probability (by Assumptions A2 and A3) and $\hat{\beta} - \beta =$

$O_p(\frac{1}{\sqrt{T}})$, Slutsky's theorem yields:

$$\left\| \frac{1}{\sqrt{T}}(X(\hat{\beta} - \beta)) \right\| \leq \frac{1}{\sqrt{T}} ||X|| \; ||\hat{\beta} - \beta|| = O_p(\frac{1}{\sqrt{T}}). \tag{15}$$

Similarly,

$$\begin{aligned} \left\| \frac{1}{\sqrt{T}}(\hat{Z}(\hat{\delta} - \delta)) \right\| &\leq \frac{1}{\sqrt{T}} ||\hat{Z}|| \; ||\hat{\delta} - \delta|| \\ &\leq \frac{1}{\sqrt{T}} ||X|| \; ||\hat{\delta} - \delta|| \end{aligned}$$

where the second inequality comes from the fact that $\hat{Z}$ is a linear transformation of X. It follows that

$$\left\| \frac{1}{\sqrt{T}}(\hat{Z}(\hat{\delta} - \delta_T)) \right\| = O_p(\frac{1}{\sqrt{T}}). \tag{16}$$

For the third term, we use Equation (13):

$$\begin{aligned} \left\| \frac{1}{\sqrt{T}}(\hat{Z} - Z)\delta_T \right\| &= \left\| \frac{1}{\sqrt{T}} \sum_{i=\hat{k}+1}^{k_0} z_i \delta_T \right\| \leq \frac{1}{\sqrt{T}} \sum_{i=\hat{k}+1}^{k_0} ||z_i|| \; ||\delta_T|| \\ &= \frac{1}{\sqrt{T}} ||\delta|| \; O_p(||\delta_T||^{-2}) = \frac{1}{\sqrt{T}} O_p(||\delta_T||^{-1}). \end{aligned}$$

As ${\delta_T}^{-1} = o_p(T^{1/2-\alpha})$, we obtain

$$\left\| \frac{1}{\sqrt{T}}(\hat{Z} - Z)\delta_T \right\| = o_p\left(\frac{1}{T^\alpha}\right). \tag{17}$$

From Equations (15), (16), and (17), we obtain by Inequality (14) that as $T \to \infty$:

$$\frac{1}{T}(Y - X\beta - Z\delta_T)'(X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T)) = o_p\left(\frac{1}{T^\alpha}\right).$$

Similarly, we have $\frac{1}{T}||X(\hat{\beta} - \beta) + (\hat{Z} - Z)\delta_T + \hat{Z}(\hat{\delta} - \delta_T)||^2 = o_p\left(\frac{1}{T^{2\alpha}}\right)$.

It follows that for $T \to \infty$:

$$\hat{\sigma}^2 = \sigma^2 + o_p\left(\frac{1}{T^\alpha}\right).$$

Limit of $\hat{Q}$:

By the law of large numbers,

$$\hat{Q} = \frac{1}{T}\sum_{t=1}^{T} z_t z_t' \xrightarrow{P} \mathbf{E}(z_t z_t')$$

Moreover, by Assumption A2 (existence of the fourth moment of z_t) and the central limit theorem $Q - \hat{Q} = O_p(\frac{1}{\sqrt{T}})$.

Limit of $\hat{\sigma}^2/(\hat{\delta}^T \hat{Q}\hat{\delta})$:

By Assumption A5, we have

$$\begin{aligned}\hat{\delta} &= \delta_T + O_p(\frac{1}{\sqrt{T}}) = \delta_0 \upsilon_T + O_p(\frac{1}{\sqrt{T}}) = \upsilon_T(\delta_0 + O_p(\frac{\upsilon_T^{-1}}{\sqrt{T}})) \\ &= \upsilon_T(\delta_0 + o_p(\frac{1}{T^\alpha})).\end{aligned}$$

Hence,

$$\begin{aligned}\frac{\frac{\hat{\sigma}^2}{\hat{\delta}'\hat{Q}\hat{\delta}}}{\frac{\sigma^2}{\delta_T'Q\delta_T}} &= \frac{\hat{\sigma}^2}{\sigma^2}\frac{\delta_T'Q\delta_T}{\hat{\delta}'\hat{Q}\hat{\delta}} = \frac{\hat{\sigma}^2}{\sigma^2}\frac{\upsilon_T^2\delta_0'Q\delta_0}{\upsilon_T^2(\delta_0 + o_p(\frac{1}{T^\alpha}))'\hat{Q}(\delta_0 + o_p(\frac{1}{T^\alpha}))} \\ &= \frac{\hat{\sigma}^2}{\sigma^2}\frac{\delta_0'Q\delta_0}{(\delta_0 + o_p(\frac{1}{T^\alpha}))'\hat{Q}(\delta_0 + o_p(\frac{1}{T^\alpha}))}.\end{aligned}$$

Therefore,

$$\frac{\frac{\hat{\sigma}^2}{\hat{\delta}'\hat{Q}\hat{\delta}}}{\frac{\sigma^2}{\delta'Q\delta}} \xrightarrow{P} 1.$$

Rate of convergence:

$$\begin{aligned}\frac{\frac{\hat{\sigma}^2}{\hat{\delta}'\hat{Q}\hat{\delta}} - \frac{\sigma^2}{\delta'Q\delta}}{\frac{\sigma^2}{\delta'Q\delta}} &= \frac{\frac{\sigma^2 + o_p(\frac{1}{T^\alpha})}{(\delta_0 + o_p(\frac{1}{T^\alpha}))'(Q + O_p(\frac{1}{\sqrt{T}}))(\delta_0 + o_p(\frac{1}{T^\alpha}))} - \frac{\sigma^2}{\delta_0'Q\delta_0}}{\frac{\sigma^2}{\delta_0^TQ\delta_0}} \\ &= \frac{\delta_0'Q\delta_0}{\sigma^2}\frac{\delta_0'Q\delta_0\sigma^2 + o_p(\frac{1}{T^\alpha}) - \delta_0'Q\delta_0\sigma^2 + o_p(\frac{1}{T^\alpha})}{(\delta_0'Q\delta_0)(\delta_0'Q\delta_0 + o_p(\frac{1}{T^\alpha}))}.\end{aligned} \tag{18}$$

It follows from Slutsky's theorem that

$$\frac{\frac{\hat{\sigma}^2}{\hat{\delta}'\hat{Q}\hat{\delta}} - \frac{\sigma^2}{\delta' Q\delta}}{\frac{\sigma^2}{\delta' Q\delta}} = \frac{o_p(\frac{1}{T^\alpha})}{\sigma^2(\delta_0' Q\delta_0 + o_p(\frac{1}{T^\alpha}))} = o_p\left(\frac{1}{T^\alpha}\right). \tag{19}$$

Proof of Theorem 4.

$$\begin{aligned} \hat{f}_k(x) - f_k(x) &= \frac{1}{2\pi}\int_{-1/h}^{1/h} e^{-itx}\left[\frac{1}{n}\sum_{j=1}^{n}\frac{\exp\left(it\hat{k}_j\right)}{\varphi_\xi(\hat{\omega}t)}dt - \varphi_k(t)\right]dt && (20)\\ &\quad + \frac{1}{2\pi}\int_{|t|>1/h} e^{-itx}\varphi_k(t)\,dt. && (21)\end{aligned}$$

By Parseval's inequality, we have

$$\begin{aligned} &\int\left[\hat{f}_k(x) - f_k(x)\right]^2 dx \\ \leq\ & \frac{2}{\pi}\int_{-1/h}^{1/h}\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\exp\left(it\hat{k}_j\right)}{\varphi_\xi(\hat{\omega}t)}dt - \varphi_k(t)\right|^2 dt && (22)\\ & + \frac{2}{\pi}\int_{|t|>1/h}|\varphi_k(t)|^2\,dt. && (23)\end{aligned}$$

First, we consider the term (23). Using the tail behavior of f_k in $\mathcal{F}$, we have

$$\frac{1}{\pi}\int_{|t|>1/h}|\varphi_k(t)|^2\,dt \leq Ch^{2\beta-1} \tag{24}$$

where $C > 0$ is a generic constant.

Now, we consider the term (22).

$$
\begin{aligned}
(22) &= \frac{2}{\pi}\int_{-1/h}^{1/h}\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\exp\left(it\hat{k}_j\right)}{\varphi_\xi\left(\hat{\omega}t\right)}dt-\frac{\varphi_{\widehat{k}}\left(t\right)}{\varphi_\xi\left(\hat{\omega}t\right)}+\frac{\varphi_{\widehat{k}}\left(t\right)}{\varphi_\xi\left(\hat{\omega}t\right)}-\varphi_k\left(t\right)\right|^2 dt \\
&\leq \frac{4}{\pi}\int_{-1/h}^{1/h}\frac{1}{\varphi_\xi\left(\hat{\omega}t\right)^2}\left|\frac{1}{n}\sum_{j=1}^{n}\exp\left(it\hat{k}_j\right)-\varphi_{\widehat{k}}\left(t\right)\right|^2 dt \qquad (25)\\
&\quad +\frac{4}{\pi}\int_{-1/h}^{1/h}\left|\frac{\varphi_{\widehat{k}}\left(t\right)}{\varphi_\xi\left(\hat{\omega}t\right)}-\varphi_k\left(t\right)\right|^2 dt. \qquad (26)
\end{aligned}
$$

Consider (25). By Fubini and Cauchy-Schwarz, we have

$$
\begin{aligned}
& E\int_{-1/h}^{1/h}\frac{1}{\varphi_\xi\left(\hat{\omega}t\right)^2}\left|\frac{1}{n}\sum_{j=1}^{n}\exp\left(it\hat{k}_j\right)-\varphi_{\widehat{k}}\left(t\right)\right|^2 dt \\
=& \int_{-1/h}^{1/h}E\left\{\frac{1}{\varphi_\xi\left(\hat{\omega}t\right)^2}\left|\frac{1}{n}\sum_{j=1}^{n}\exp\left(it\hat{k}_j\right)-\varphi_{\widehat{k}}\left(t\right)\right|^2\right\}dt \\
\leq& \int_{-1/h}^{1/h}\sqrt{E\left(\frac{1}{\varphi_\xi\left(\hat{\omega}t\right)^4}\right)E\left[\left|\frac{1}{n}\sum_{j=1}^{n}\exp\left(it\hat{k}_j\right)-\varphi_{\widehat{k}}\left(t\right)\right|^4\right]}dt
\end{aligned}
$$

Using Lemma 2, we have

$$
\begin{aligned}
\sqrt{E\left(\frac{1}{\varphi_\xi\left(\hat{\omega}t\right)^4}\right)} &\leq C\sqrt{E\left(\hat{\omega}^6\right)}t^3 \\
&= C\sqrt{E\left(\left(\frac{\hat{\omega}}{\omega}\right)^6\right)}\omega^3 t^3.
\end{aligned}
$$

Moreover, by Assumption C, $E\left(\left(\frac{\hat{\omega}}{\omega}\right)^6\right)<\infty$ and by the independence of $\left\{\widehat{k}_j\right\}$, $j=1,2,...,n$, $E\left[\left|\frac{1}{n}\sum_{j=1}^{n}\exp\left(it\hat{k}_j\right)-\varphi_{\widehat{k}}\left(t\right)\right|^4\right]=O\left(1/n^2\right)$. Combining these terms, we obtain

$$
E\int_{-1/h}^{1/h}\frac{1}{\varphi_\xi\left(\hat{\omega}t\right)^2}\left|\frac{1}{n}\sum_{j=1}^{n}\exp\left(it\hat{k}_j\right)-\varphi_{\widehat{k}}\left(t\right)\right|^2 dt=\frac{\omega^3}{h^4}O\left(\frac{1}{n}\right).
$$

Now consider (26). Using $\varphi_{\widehat{k}}(t) = \varphi_k(t)\,\varphi_{\xi_T}(\omega t)$ and $|\varphi_k(t)|^2 \leq 1$, we have

$$\begin{aligned}
(26) &= \frac{4}{\pi}\int_{-1/h}^{1/h} |\varphi_k(t)|^2 \left|\frac{\varphi_{\xi_T}(\omega t)}{\varphi_\xi(\hat{\omega} t)} - 1\right|^2 dt \\
&\leq \frac{4}{\pi}\int_{-1/h}^{1/h} \left|\frac{\varphi_{\xi_T}(\omega t)}{\varphi_\xi(\hat{\omega} t)} - 1\right|^2 dt.
\end{aligned}$$

A change of variables, $u = \omega t$, yields

$$\begin{aligned}
\int_{-1/h}^{1/h} \left|\frac{\varphi_{\xi_T}(\omega t)}{\varphi_\xi(\hat{\omega} t)} - 1\right|^2 dt &= \frac{1}{\omega}\int_{-\omega/h}^{\omega/h} \left|\frac{\varphi_{\xi_T}(u)}{\varphi_\xi\left(\frac{\hat{\omega}}{\omega}u\right)} - 1\right|^2 du \\
&\leq \frac{1}{\omega}\int \left|\frac{\varphi_{\xi_T}(u)}{\varphi_\xi\left(\frac{\hat{\omega}}{\omega}u\right)} - 1\right|^2 du \\
&= \frac{1}{\omega} O(b_T)
\end{aligned}$$

by Assumption C.

REFERENCES

Andrews, D. (1993) Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61, p. 821-856.

Andrews, D. (2003) Tests for Parameter Instability and Structural Change with Unknown Change Point : A Corrigendum, *Econometrica*, 71, 395-397.

Arellano, M. and S. Bonhomme (2012) Identifying Distributional Characteristics in Random Coefficients Panel Data Models, *Review of Economic Studies*, 79, 987-1020.

Baek, Y. (2024) Estimation of a Structural Break Point in Linear Regression Models, J*ournal of Business & Economic Statistics*, 42, 95-108.

Bai, J (1994) Least Squares Estimation of a Shift in Linear Processes, *Journal of Time Series Analysis*, 15, 453-71.

Bai, J. (1997) Estimation of a Change Point in Multiple Regression Models, *The Review of Economics and Statistics*, 79, 551-563.

Carrasco, M. and J.P. Florens (2011) Spectral method for deconvolving a density, *Econometric Theory*, 27, 546-581.

Carroll, R. and P. Hall (1988) Optimal Rates of Convergence for Deconvolving a Density, *Journal of American Statistical Association*, 83, No.404, 1184-1186.

Chen, X., H. Hong, and D. Nekipelov (2011) Nonlinear Models of Measurement Errors, *Journal of Economic Literature*, 49, 901-937.

Delaigle, A. and A. Meister (2008) Density estimation with heteroscedastic error, *Bernoulli*, 14, N.2, 562-579.

Fama, E. and K. French (2015) A five-factor asset pricing model, *Journal of Financial Economics*, 116, 1-22.

Fan, J. (1991) On the optimal rates of convergence for nonparametric deconvolution problems, *The Annals of Statistics*, 19, No.3, 1257-1272.

Gu, J. and R. Koenker (2017) Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data, *Journal of Applied Econometrics*, 32, 575-599.

Hansen, B. (2000) Sample Splitting and Threshold Estimation, *Econometrica*, 68, 3, 575-603.

Hinkley, D.V. (1970) Inference About the Change-Point in a Sequence of Random Variables, *Biometrika*, 57, 1, 1-17.

Jiang L., X. Wang, and S. Yu (2018) New Distribution Theory for the Estimation of Structural Break, *Journal of Econometrics*, 205, 156-176.

Joseph, L. and D. Wolfson (1993a) Estimation in the Multi-path Change-point Problems, *Commun. Statist.-Theory Meth.*, 21(4), 897-913.

Joseph, L. and D. Wolfson (1993b) Maximum Likelihood Estimation in the Multi-path Change-point Problem, *Ann. Inst. Statist. Math.*, 45, No. 3, 511-530.

Joseph L., D. Wolfson, R. Berger, and R. Lyle (1997) Analysis of Panel Data with Change-Points, *Statistica Sinica*, **7**, 687-703.

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

Liu, L., H.R. Moon, and F. Schorfheide (2020) Forecasting With Dynamic Panel Data Models, *Econometrica*, 88, 171-201.

Masry, E. (1991) Multivariate Probability Density Deconvolution for Stationary Random Processes, *IEEE Transactions on Information Theory*, 37, 1105-1115.

Meister, A. (2006) Density estimation with normal measurement error with unknown variance, *Statistica Sinica*, Vol. 16. 195-211.

Picard, D. (1985) Testing and estimating change-points in time series, *Adv. in Appl. Probab.*, 17, 841-867.

Pollak, M. and D. Siegmund (1985) A diffusion process and its applications to detecting a change in the drift of a Brownian motion, *Biometrika*, 72, 2, 267-280.

Stefanski, L. and R. Carroll (1990) Deconvoluting Kernel Density Estimators, *Statistics*, 2, 169-184.

Wand, M.P. and M.C. Jones (1995) *Kernel Smoothing*, Chapman & Hall.

Yao, Y. (1987) Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables, *The Annals of Statistics*, 15, No. 3, 1321-1328.