A Combination Forecast for Nonparametric Models with Structural Breaks*

Zongwu Cai   Gunawan

Department of Economics, University of Kansas, Lawrence, KS 66045, USA

September 25, 2023

Abstract: Structural breaks in time series forecasting can cause inconsistency in the conventional OLS estimator. Recent research suggests combining pre and post-break estimators for a linear model can yield an optimal estimator for weak breaks. However, this approach is limited to linear models only. In this paper, we propose a weighted local linear estimator for a nonlinear model. This estimator assigns a weight based on both the distance of observations to the predictor covariates and their location in time. We investigate the asymptotic properties of the proposed estimator and choose the optimal tuning parameters using multifold cross-validation to account for the dependence structure in time series data. Additionally, we use a nonparametric method to estimate the break date. Our Monte Carlo simulation results provide evidence for the forecasting outperformance of our estimator over the regular nonparametric post-break estimator. Finally, we apply our proposed estimator to forecast GDP growth for two countries and demonstrate its superior performance compared to the benchmark estimators using Diebold-Mariano tests.

Keywords: Combination Forecasting; Local Linear Fitting; Multifold Cross-Validation; Nonparametric Model; Structural Break Model;

1 Introduction

Econometric forecasting of time series data often assumes stationarity, and therefore the constancy of model parameters over time, such as mean, variance, frequency, trend, or combined. In practice, these parameters may change over time. For example, the US industrial production experienced slowdown during the financial crisis between 2007 and 2008, as well as the Covid-19 pandemic between 2020 and 2022, while it experiences expansion in other

*Contact information: caiz@ku.edu (Z. Cai) and gunawan@ku.edu (Gunawan).

time periods. Therefore, investigating structural instability is a long-standing issue in time series econometrics. These two different regimes are regarded as a consequence of parameter shift or varying smoothly over time. For the latter case, the reader is referred to the papers by Cai (2007), Sun, Hong, Lee, Wang, and Zhang (2021), and references therein. The point at which the regime change occurs is called a change point or structural break in statistics and econometrics literature, whereas the associated models are known as structural break models. In practice, breaks in the parameters of a forecasting model are caused by events that are essentially unknowable ex-ante and may be triggered by several factors, such as institutional, political, social, financial, legal, or technological change, may precipitate the breaks. These breaks will be understood better retrospectively rather than at the time of occurrence. Typically, it is assumed that the modeler does not have knowledge of the process determining the break (Clements and Hendry, 2011).

Structural breaks pose methodological challenges for forecasting exercise. In a time series model with a structural break in the conditional mean and/or conditional variance, a conventional OLS estimator based on full-sample observations might be inconsistent. A consistent estimator can be computed using post-break observations only if the post-break sample is sufficiently large. However, such forecasts may not be optimal in terms of the mean squared forecast error (MSFE) as the relatively small post-break sample size may induce large estimation uncertainty (Pesaran and Pick, 2011; Pesaran, Pick, and Pranovich, 2013; Rossi, 2013; Lee, Parsaeian, and Ullah, 2022). Therefore, pre-break observations may still be useful for forecast improvement depending on the magnitude of the break. If there is no break, the usual full-sample estimator is optimal. If the break is strong, the post-break estimator may be optimal. If the break is weak or moderate, a combined estimator of the full-sample estimator and the post-break estimator would be optimal, where a combination weight between 0 and 1 is chosen in a way that optimizes the trade-off between the bias and variance efficiency of the full-sample estimator.

The idea of combining information in producing the aforementioned forecast could be considered as frequentist model averaging, since we average the pre-break and post break estimators (Hjort and Claeskens, 2003; Hansen, 2007, 2008; Hansen and Racine, 2012; Sun et al., 2021; Lee et al., 2022). In this spirit, there is a large number of works that propose different forecast combination methods, particularly in the parametric literature (Clements and Hendry, 2006, 2011; Pesaran and Timmermann, 2005, 2007; Timmermann, 2006; Pesaran et al., 2013; Lee et al., 2022), and in the nonparametric setting (Sun et al., 2021). Another appealing approach that can be used for combining information from before and after the break is a semi-parametric kernel-based regression model. In particular, Lee et al. (2022) developed a weighted generalized least squares estimator (WGLS) for time series structural break models which exploits pre-break data in addition to the post-break data and uses leave-one-out cross-validation for choosing the tuning parameters.

This study is motivated by forecasting output growth using the slop of yield curve as a predictor. However, recent studies concluded that the forecasting relationship between output growth and yield curve may be subject to structural breaks; see, for example, Stock and Watson (1999), Giacomini and Rossi (2006), Estrella, Rodrigues, and Schich (2003), Schrimpf and Wang (2010), and references therein. The presence of structural breaks in turn leads to an instability of the model coefficients, which calls its usefulness for forecasting into question. Also, the relationship between output growth and yield curve seems to be nonlinear; see, for instance, Figure 1 in Section 4. Therefore, to conduct this empirical study, we need to develop a new nonparametric forecasting technique with structural breaks.

This paper contributes to the nonparametric forecasting with structural breaks literature by proposing a nonparametric method to exploit information contained in the dataset before breaks occur. While most cited previous studies use (semi-)parametric forecasting models, we approach this problem considering nonparametric mean regression. Our proposed estimator, inspired by the WGLS estimator by Lee et al. (2022), assigns weights to observations before and after the breaks. This weight is additional to the usual nonparametric weights that are given to observations based on how far they are located relative to the predictor covariates. Hence, it is termed as a "weighted local linear estimator". Also, the asymptotic properties, including the asymptotic bias and variance, of the proposed estimator are investigated and some discussions are provided to show that the asymptotic variance indeed can be smaller than that for the nonparametric estimator using only the post-break observations. Furthermore, the smoothing parameters are chosen using multifold cross-validation as in Cai, Fan, and Yao (2000), while the break date is estimated in a nonparametric way using the latest method proposed in the literature by Mohr and Selk (2020). In order to evaluate the forecasting performance, we perform Monte Carlo simulations with diverse schemes of data generating process. We apply this method in an empirical application for predicting gross domestic product (GDP) growth rate for two selected countries and compare its forecasting performance using Diebold-Mariano test. Both simulation and empirical application results suggest the outperformance of our proposed estimator over the usual post-break estimator and the weighted least squares estimator.

The remainder of the paper is organized as follows. The model, the break date estimator, as well as the weighted nonparametric regression predictor are introduced in Section 2. Section 3 presents the Monte Carlo simulation study and its results. Section 4 illustrates an empirical application, while Section 5 concludes the paper. Finally, the sketch proofs of theoretical results are relegated to Appendix, together with regularity conditions for deriving the asymptotic theories.

2 Model and Its Forecasting Procedures

2.1 Model Setup

Let $\{(Y_t, \boldsymbol{X}_t) : t \in \mathbb{N}\}$ be a weakly dependent stochastic process in $\mathbb{R} \times \mathbb{R}^d$. We consider following the forecasting model

$$Y_{t+\tau} = m_t(\boldsymbol{X}_t) + u_t, \quad t \in \mathbb{N}, \tag{1}$$

where $\tau \geq 0$ is the forecasting horizon (τ-step ahead forecast), the idiosyncratic error u_t satisfies $\mathbb{E}[u_t|\mathcal{F}_t] = 0$ almost surely for the σ-field $\mathcal{F}_t = \sigma(u_{j-1}, \boldsymbol{X}_j : j \leq t)$. It is assumed that there exists a change point in the prediction function such that

$$m_t(\boldsymbol{x}) = m_{(1)}(\boldsymbol{x})\mathbb{1}(t \leq T_1) + m_{(2)}(\boldsymbol{x})\mathbb{1}(t > T_1) = m_{(1)}(\boldsymbol{x}) - \lambda(\boldsymbol{x})d_t \tag{2}$$

with $m_{(1)}(\boldsymbol{x}) \not\equiv m_{(2)}(\boldsymbol{x})$ and $\lambda(\boldsymbol{x}) = m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x})$, the break size function, where $1 < T_1 < T$ is the break point, which might be unknown, $d_t = \mathbb{1}(t > T_1)$, and both functions $m_{(1)}(\boldsymbol{x})$ and $m_{(2)}(\boldsymbol{x})$ are assumed to be continuous and satisfy some regularity conditions to ensure that $\{Y_t, \boldsymbol{X}_t\}$ is a stationary α-mixing time series. Here, $\boldsymbol{X}_t$ is allowed to include some lags of Y_t.[1] Also, it is assumed that $T_1 = \lfloor T s_0 \rfloor$ with $0 < s_0 < 1$, the portion of the pre-break observations. Finally, note that the expression in the right hand side of (2) can be regarded as a special case of a functional coefficient time series model proposed in Cai et al. (2000).

It is clear that when $m_t(\boldsymbol{x}) = \beta_t^\top \boldsymbol{x}$ in (1) with β_t changing smoothly over time, the model in (1) becomes to the models studied by Cai (2007) for estimation and forecasting

[1] For this regard, the reader is referred to the paper by Cai and Masry (2000) for details on the conditions and the theoretical justifications.

and Sun et al. (2021) for a model averaging. Furthermore, when β_t has structural change, the model in (2) was investigated by Pesaran et al. (2013) and Lee et al. (2022) for the weighted generalized least squares estimators for a conventional structural change linear model to combine the information from both pre-break and post-break. As argued in Pesaran et al. (2013) and Lee et al. (2022), the WGLS estimators proposed in Pesaran et al. (2013) and Lee et al. (2022) have an ability to reduce MSFE under the structural breaks by using the full-sample observations instead of using only the post-break observations, by deriving the optimal weight for the pre-break proportion of the full-sample. Note that in (2), our focus is only on one break and it is easy to generalize the model in (2) to a multiple break case.

2.2 Weighted Local Linear Estimation

Inspired by the work of Lee et al. (2022), we propose an estimator for nonparametric time series structural break model, where breaks may occur in the mean function and error variance. In particular, we are interested in estimating the mean function after the break by partly using information contained in the pre-break observations. Our starting point is the following nonparametric local linear regression problem. For $\boldsymbol{X}_t$ in a neighborhood of $\boldsymbol{x}$, a given grid point from the data domain, we can approximate locally the mean function by $m(\boldsymbol{X}_t) \approx \beta_0(\boldsymbol{x}) + \beta_1(\boldsymbol{x})^\top(\boldsymbol{X}_t - \boldsymbol{x})$ by ignoring the higher order term, where $\beta_0(\boldsymbol{x}) = m(\boldsymbol{x})$ and $\beta_1(\boldsymbol{x}) = m'(\boldsymbol{x})$, the first order derivative of $m(\boldsymbol{x})$. Then, the locally weighted least squares is given by

$$\min_{\beta_0,\beta_1} \sum_{t=1}^{T} \widetilde{K}(t,\gamma) K_h(\boldsymbol{x} - \boldsymbol{X}_t) \left(Y_{t+\tau} - \beta_0 - \beta_1^\top(\boldsymbol{X}_t - \boldsymbol{x})\right)^2, \tag{3}$$

where for some $0 \leq \gamma \leq 1$,

$$\widetilde{K}(t,\gamma) = \gamma \mathbb{1}(t \leq T_1) + \mathbb{1}(t > T_1) \tag{4}$$

is a discrete kernel.[2] We use the short notation $K_h(u) = K(u/h)/h^d$, where $K(\cdot)$ is a kernel function and h is the bandwidth. As mentioned in Cai et al. (2000), the estimation procedure and its asymptotic theory for the d-dimensional case are the same those for the case that $\boldsymbol{X}_t$ is the univariate case. Therefore, for ease notation, our next presentation is only for one-dimensional case; that is $d = 1$, so that $\boldsymbol{X}_t$ and $\boldsymbol{x}$ become to be X_t and x, respectively.

Equation (3) shows the weighting scheme used for this estimator, i.e. $\widetilde{K}(t, \gamma)$ to assign a weight γ or 1 based on where the observations lie in time t, and $K_h(x - X_t)$ to assign weights on each observation based on how close they are to a point x. In addition, we fit a local linear estimator instead of a local constant one in order to reduce boundary bias and to achieve the minimax efficiency (Fan and Gijbels, 1996). Based on (4), post-break observations receive a weight of 1, while a weight of $\gamma \in [0, 1]$ is assigned to pre-break observations, as information from recent data is considered more relevant for forecasting. If γ is close to zero, then the estimator is heavily weighted on the post-break observations. If $\gamma = 1$, then we would ignore any structural break and have a usual full-sample estimator. In other cases where $\gamma \in (0, 1)$, we thus have a combination of pre- and post-break observations for the estimator.

The solution to (3) is a $\widehat{\beta}(x) = (\widehat{\beta}_0(x), \widehat{\beta}_1(x))^\top$, which gives $\widehat{m}(x) = \widehat{\beta}_0(x)$, the estimator of $m(x)$, and $\widehat{m}'(x) = \widehat{\beta}_1(x)$, the estimator of $m'(x)$. To express the estimator in matrix form, we introduce the following notations. Let $Y^\top = (Y_{(1)}^\top, Y_{(2)}^\top)$ be a $T \times 1$ vector of the dependent variable, where $Y_{(i)} = (Y_{T_{i-1}+1+\tau}, \dots, y_{T_i+\tau})^\top$, and $\boldsymbol{X}^\top = (\boldsymbol{X}_{(1)}^\top, \boldsymbol{X}_{(2)}^\top)$ be a $T \times 2$

[2]For more bout the discrete kernel, the reader is referred to the book by Li and Racine (2007).

matrix of the independent variables, where

$$\boldsymbol{X}_{(i)} = \begin{pmatrix} 1 & (X_{T_{i-1}+1} - x) \\ 1 & (X_{T_{i-1}+2} - x) \\ \vdots & \vdots \\ 1 & (X_{T_i} - x) \end{pmatrix}$$

with $i = 1$ and 2 and the convention that $T_0 = 0$, and $T_2 = T$. Now, define the $T \times T$ weighting matrix $\mathbf{W}(\gamma)$ as follows

$$\mathbf{W} = \mathbf{W}(\gamma) = \mathbf{W}_\gamma \mathbf{W}_k, \quad \text{where } \mathbf{W}_\gamma = \begin{pmatrix} \gamma \mathbf{I}_{T_1} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_{T-T_1} \end{pmatrix} \text{ and } \mathbf{W}_k = \begin{pmatrix} \mathbf{W}_{(1)} & \mathbf{O} \\ \mathbf{O} & \mathbf{W}_{(2)} \end{pmatrix}$$

with $\mathbf{W}_{(1)} = \mathrm{diag}(K_h(x - X_1), \ldots, K_h(x - X_{T_1}))$ and $\mathbf{W}_{(2)} = \mathrm{diag}(K_h(x - X_{T_1+1}), \ldots, K_h(x - X_T))$ as well as $\mathbf{I}_d$ denoting a $d \times d$ identity matrix. Thus, the minimizer of (3) is given by

$$\widehat{\beta}(x) = (\widehat{\beta}_0(x), \widehat{\beta}_1(x))^\top = (\boldsymbol{X}^\top \mathbf{W} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W} Y. \tag{5}$$

In particular, our weighted local linear (WLL) estimator for the mean function is given by

$$\widehat{m}_{\text{wll}}(x) = \widehat{\beta}_0(\boldsymbol{x}), \tag{6}$$

which reduces to the local linear estimator of $m_{(2)}(x)$ when $\gamma = 0$. Further, equation (5) can be rewritten as

$$\begin{aligned} \widehat{\beta}(x) &= \left[\gamma \boldsymbol{X}_{(1)}^\top \mathbf{W}_{(1)} \boldsymbol{X}_{(1)} + \boldsymbol{X}_{(2)}^\top \mathbf{W}_{(2)} \boldsymbol{X}_{(2)}\right]^{-1} \left(\gamma \boldsymbol{X}_{(1)}^\top \mathbf{W}_{(1)} Y_{(1)} + \boldsymbol{X}_{(2)}^\top \mathbf{W}_{(2)} Y_{(2)}\right) \\ &= \Gamma \widehat{\beta}_{(1)}(x) + (\mathbf{I}_2 - \Gamma) \widehat{\beta}_{(2)}(x), \end{aligned} \tag{7}$$

where

$$\Gamma = \Gamma(x, \gamma) = \left[\gamma \boldsymbol{X}_{(1)}^\top \mathbf{W}_{(1)} \boldsymbol{X}_{(1)} + \boldsymbol{X}_{(2)}^\top \mathbf{W}_{(2)} \boldsymbol{X}_{(2)}\right]^{-1} \left(\gamma \boldsymbol{X}_{(1)}^\top W_{(1)} \boldsymbol{X}_{(1)}\right).$$

Indeed, equation (7) can be viewed as the combined estimator of the pre-break and the post-break estimators, i.e., a combination of $\widehat{\beta}_{(1)}(x)$ for the estimator before the break and

$\widehat{\beta}_{(2)}(x)$ for the estimator after the break, where $\widehat{\beta}_{(1)}(x) = \left[\boldsymbol{X}_{(1)}^{\top}\mathbf{W}_{(1)}\boldsymbol{X}_{(1)}\right]^{-1}\left(\boldsymbol{X}_{(1)}^{\top}\mathbf{W}_{(1)}Y_{(1)}\right)$ and $\widehat{\beta}_{(2)}(x) = \left[\boldsymbol{X}_{(2)}^{\top}\mathbf{W}_{(2)}\boldsymbol{X}_{(2)}\right]^{-1}\left(\boldsymbol{X}_{(2)}^{\top}\mathbf{W}_{(2)}Y_{(2)}\right)$, respectively, with the combination weight Γ. A feasible version for (7) is

$$\widehat{\beta}(x) = \widehat{\Gamma}\widehat{\beta}_{(1)}(x) + \left(\mathbf{I}_2 - \widehat{\Gamma}\right)\widehat{\beta}_{(2)}(x),$$

where $\hat{\Gamma}$ is a consistent estimator of Γ, say,

$$\hat{\Gamma} = \left[\hat{\gamma}\boldsymbol{X}_{(1)}^{\top}\mathbf{W}_{(1)}\boldsymbol{X}_{(1)} + \boldsymbol{X}_{(2)}^{\top}\mathbf{W}_{(2)}\boldsymbol{X}_{(2)}\right]^{-1}\left(\hat{\gamma}\boldsymbol{X}_{(1)}^{\top}W_{(1)}\boldsymbol{X}_{(1)}\right),$$

where $\hat{\gamma}$ is chosen using multifold cross-validation discussed in Section 2.4. Of course, $\widehat{\beta}(x)$ involves the bandwidth h, which can be selected using multifold cross-validation.

2.3 Asymptotic Analyses

Now, we investigate the asymptotic properties of $\widehat{m}_{\text{wll}}(x)$. First, we evaluate Γ. To do so, consider $\boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{X} = (\gamma - 1)\boldsymbol{X}_{(1)}^{\top}\mathbf{W}_{(1)}\boldsymbol{X}_{(1)} + \boldsymbol{X}^{\top}\mathbf{W}_k\boldsymbol{X}$. For $j \geq 0$, let $\mu_j = \int K(u)u^j du$ and

$$S_j(x) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_t - x)\left(\frac{X_t - x}{h}\right)^j.$$

It is easy to see that

$$\boldsymbol{X}^{\top}\mathbf{W}_k\boldsymbol{X} = T\mathbf{H}\begin{pmatrix} S_0(x) & S_1(x) \\ S_1(x) & S_2(x) \end{pmatrix}\mathbf{H},$$

where $\mathbf{H} = \text{diag}\{1, h\}$. Under some regularity conditions given in Appendix; see, for example, Assumptions (A1) - (A5), it follows from (C.1) in Appendix that as $T \to \infty$, $S_j(x) \xrightarrow{p} \mu_j f(x)$, where $f(x)$ is the density of X_t, $\mu_j = \int u^j K(u)$, and $\xrightarrow{p}$ denotes the convergence in probability. Therefore, $\boldsymbol{X}^{\top}\mathbf{W}_k\boldsymbol{X} = Tf(x)\mathbf{H}\mu\mathbf{H}(1 + o_p(1))$, where $\mu = \text{diag}\{1, \mu_2\}$ and $A_T = o_p(B_T)$ means that $A_T/B_T \xrightarrow{p} 0$ as $T \to \infty$. Similarly, we have $\boldsymbol{X}_{(1)}^{\top}\mathbf{W}_{(1)}\boldsymbol{X}_{(1)} = f(x)T_1\mathbf{H}\mu\mathbf{H}(1 + o_p(1)) = s_0 f(x)T\mathbf{H}\mu\mathbf{H}(1 + o_p(1))$. Hence,

$\boldsymbol{X}^\top \mathbf{W} \boldsymbol{X} = [1 + (\gamma - 1)s_0] f(x) T \mathbf{H} \mu \mathbf{H}(1 + o_p(1))$, which implies that $\Gamma = s_b \mathbf{I}_2$, where $s_b = s_0 \gamma [1 + (\gamma - 1)s_0]^{-1}$, which depends on both γ and s_0.

2.3.1 Asymptotic Bias

Next, we evaluate the asymptotic bias for $\widehat{m}_{\text{wll}}(x)$. For this purpose, (7) is re-expressed as $\widehat{\beta}(x) = \widehat{\beta}_{(2)}(x) + \Gamma \left[\widehat{\beta}_{(1)}(x) - \widehat{\beta}_{(2)}(x) \right]$, so that $\widehat{m}_{\text{wll}}(x) - m_{(2)}(x) = \widehat{\beta}_0(\boldsymbol{x}) - m_{(2)}(x) \approx \widehat{\beta}_{0,(2)}(x) - m_{(2)}(x) + s_b \left[\widehat{\beta}_{0,(1)}(x) - \widehat{\beta}_{0,(2)}(x) \right]$, where $\widehat{\beta}_{0,(1)}(x)$ and $\widehat{\beta}_{0,(2)}(x)$ are the first component of $\widehat{\beta}_{(1)}(x)$ and $\widehat{\beta}_{(2)}(x)$, respectively. Indeed, $\widehat{\beta}_{0,(1)}(x)$ is the local linear estimator for $m_{(1)}(x)$ using only the pre-break observations and $\widehat{\beta}_{0,(2)}(x)$ is the local linear estimator for $m_{(2)}(x)$ using only the post-break observations. Also, under regularity conditions given in Appendix, we show in Appendix that the asymptotic biases for $\widehat{\beta}_{0,(1)}(x)$ and $\widehat{\beta}_{0,(2)}(x)$ are $B_1(x) = h^2 m''_{(1)}(x) \mu_2 / 2$ and $B_2(x) = h^2 m''_{(2)}(x) \mu_2 / 2$, respectively. Therefore, the asymptotic bias for $\widehat{m}_{\text{wll}}(x)$ is

$$B_{\text{wll}}(x) = B_2(x) + s_b \left[\lambda(x) + \frac{h^2}{2} \mu_2 \lambda''(x) \right], \tag{8}$$

where $\lambda(x)$ is defined in (2). Clearly, the second term in the right hand side of $B_{\text{wll}}(x)$ is extra by comparing with that for $\widehat{\beta}_{0,(2)}(x)$ due to the weighted estimation procedure and it is negative if $\lambda(x) < 0$ by ignoring the higher order term. Finally, one can see that for a linear model ($m_t(\boldsymbol{X}_t) = \beta_t^\top \boldsymbol{X}_t$), $B_{\text{wll}}(x)$ reduces to $s_b \lambda(x)$, which is similar to those in Pesaran et al. (2013) and Lee et al. (2022).

One might be interested in a bias correction version to reduce the asymptotic bias of $\widehat{m}_{\text{wll}}(x)$ to be in the order of $O_p(h^2)$ instead of $\lambda(x)$. To do so, it is easy to see that $\hat{\lambda}(x) = \widehat{\beta}_{0,(1)}(x) - \widehat{\beta}_{0,(2)}(x)$ is a consistent estimate of $\lambda(x)$. Therefore, a bias corrected version of $\widehat{m}_{\text{wll}}(x)$ is

$$\widehat{m}_{\text{wll,c}}(x) = \widehat{m}_{\text{wll}}(x) - \hat{s}_b \left[\widehat{\beta}_{0,(1)}(x) - \widehat{\beta}_{0,(2)}(x) \right], \tag{9}$$

where $\hat{s}_b = s_0\hat{\gamma}[1 + (\hat{\gamma} - 1)s_0]^{-1}$ if $\hat{\gamma}$ is a consistent estimate of γ. Then, the asymptotic bias for $\widehat{m}_{\text{wll,c}}(x)$ should be

$$\frac{1}{2}h^2\mu_2\left[m''_{(2)}(x) + s_b\lambda''(x)\right] = \frac{1}{2}h^2\mu_2\left[(1 - s_b)m''_{(2)}(x) + s_b m''_{(1)}(x)\right],$$

which is the conventional asymptotic bias for a local linear estimator; see, for instance, Cai et al. (2000) for details.

2.3.2 Asymptotic Variance

Finally, addition to the asymptotic bias given in (8), we consider the asymptotic variance of $\widehat{m}_{\text{wll}}(x)$. To this end, express

$$\boldsymbol{X}^\top \mathbf{W} U = \gamma \sum_{t=1}^{T_1} K_h(X_t - x)\begin{pmatrix} 1 \\ X_t - x \end{pmatrix} u_t + \sum_{t=T_1+1}^{T} K_h(X_t - x)\begin{pmatrix} 1 \\ X_t - x \end{pmatrix} u_t = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix},$$

where U is defined in the same way as Y, which is the main term that contributes to the asymptotic variance of $\widehat{m}_{\text{wll}}(x)$, and A_1 and A_2 are defined in a clear manner. Clearly,

$$C_0(\gamma) = \sqrt{\frac{h}{T}}A_1 = \gamma\sqrt{s_0}\,C_1 + \sqrt{1 - s_0}\,C_2,$$

where

$$C_1 = \sqrt{\frac{h}{T_1}}\sum_{t=1}^{T_1} K_h(X_t - x)u_t \quad \text{and} \quad C_2 = \sqrt{\frac{h}{T - T_1}}\sum_{t=T_1+1}^{T} K_h(X_t - x)u_t$$

One can show in Appendix that under regularity conditions given in Appendix; see, for example, Assumptions (B1) - (B4),

$$C_1 \xrightarrow{d} N\left(0, \sigma^2_{m,1}(x)\right) \quad \text{and} \quad C_2 \xrightarrow{d} N\left(0, \sigma^2_{m,2}(x)\right),$$

where $\xrightarrow{d}$ denotes the convergence in distribution, $\sigma^2_{m,1}(x) = \nu_0\sigma^2_1(x)f(x)$ and $\sigma^2_{m,2}(x) = \nu_0\sigma^2_2(x)f(x)$ with $\nu_j = \int u^{2j}K^2(u)du$ $(j \geq 0)$, $\sigma^2_1(x) = E\left(u^2_t | X_t = x\right)$ for $t \leq T_1$ and $\sigma^2_2(x) =$

$E\left(u_t^2 \middle| X_t = x\right)$ for $t \geq T_1$, if the conditional variance of u_t given $X_t = x$ has the same break date as the mean function. Also, it is not difficult to show that $\text{Cov}(C_1, C_2) \to 0$ as $T \to \infty$. Therefore,

$$C_0(\gamma) = \sqrt{\frac{h}{T}} \left[\gamma \sum_{t=1}^{T_1} K_h(X_t - x) u_t + \sum_{t=T_1+1}^{T} K_h(X_t - x) u_t \right]$$
$$= \gamma \sqrt{s_0}\, C_1 + \sqrt{1 - s_0}\, C_2 \ \xrightarrow{d}\ N\left(0, \sigma^2_{m,0}(x)\right),$$

where $\sigma^2_{m,0}(x) = \nu_0 \left[s_0 \gamma^2 \sigma_1^2(x) + (1 - s_0)\sigma_2^2(x)\right] f(x)$, which implies that

$$\sqrt{Th}\left[\widehat{m}_{\text{wll}}(x) - B_{\text{wll}}(x) + o_p(h^2)\right] \ \xrightarrow{d}\ N\left(0, \sigma^2_{\text{wll}}(x)\right), \tag{10}$$

where $\sigma^2_{\text{wll}}(x) = \sigma^2_{m,0}(x)[1 + (\gamma - 1)s_0]^{-2}/f^2(x)$, which is regarded as the asymptotic variance of $\widehat{m}_{\text{wll}}(x)$. If there is no break in the variance function; that is, $\sigma^2(x) = E\left(u_t^2 \middle| X_t = x\right) = \sigma_1^2(x) = \sigma_2^2(x)$, then, it is reduced to $\sigma^2_{\text{wll}}(x) = \nu_0\, s_{\text{wll}} \sigma^2(x)/f(x)$, where $s_{\text{wll}} = [\gamma^2 s_0 + (1 - s_0)]/[1 + (\gamma - 1)s_0]^2$. By the same token, it is not difficult to derive the asymptotic variance of $\widehat{\beta}_{0,(2)}(x)$, which is $\sigma^2_{(2)}(x) = \nu_0 s_{(2)} \sigma^2(x)/f(x)$, where $s_{(2)} = 1/[1 - s_0]$. Evidently, $s_{\text{wll}} < s_{(2)}$ so that the asymptotic variance for $\widehat{m}_{\text{wll}}(x)$ is smaller than that for $\widehat{\beta}_{0,(2)}(x)$. Note that (10) provides the asymptotic normality for $\widehat{m}_{\text{wll}}(x)$. Also, note that when γ is consistently estimated as $\hat{\gamma}$,

$$C_0(\hat{\gamma}) = C_0(\gamma) + (\hat{\gamma} - \gamma)\sqrt{s_0}\, C_1 = C_0(\gamma) + o_p(1) \ \xrightarrow{d}\ N\left(0, \sigma^2_{\text{wll}}(x)\right)$$

by Slutsky theorem, $\sigma^2_{wll}(x)$ is defined in (10), which indicates that the asymptotic normality for $\widehat{m}_{\text{wll}}(x)$ is the same for both known γ and the consistent estimate $\hat{\gamma}$, as long as γ can be consistently estimated.

It is clear that from (8) and (10), the mean squared error (MSE) of $\widehat{m}_{\text{wll}}(x)$ is given by

$$\text{MSE}\left(\widehat{m}_{\text{wll}}(x)\right) = B^2_{\text{wll}}(x) + \frac{\sigma^2_{wll}(x)}{Th},$$

which provides a criterion for choosing the optimal h and γ simultaneously, described as follows.

2.4 Selection of Tuning Parameters

Our weighted local linear estimator (7) relies on two parameters that require selection: the bandwidth h and the weight γ. To determine appropriate values for these parameters, we employ the multifold cross-validation method, which allows us to effectively choose optimal values for both h and γ.

A popular method used for choosing tunning parameters is the leave-one-out cross-validation as used in Lee et al. (2022). However, as pointed out by Shao (1993) and Cai et al. (2000), this leave-one-out cross-validation method would fail for time series data, since adjacent points might be highly dependent. In our study, we use a modified multifold cross-validation proposed by Cai et al. (2000) to be attentive to the structure of stationary time series data, which is different from that in Lee et al. (2022).

Let m and Q be two given positive integers such that $T > mQ$. The idea is first to use Q sub-series of lengths $T - qm$ $(q = 1, \ldots, Q)$ to estimate the unknown mean functions and then compute the one-step forecasting errors of the next section of the time series of length m based on the estimated models. More precisely, we choose the optimal weight γ and the optimal bandwidth that minimize the average mean squared (AMS) error

$$\mathrm{AMS}(h, \gamma) = \frac{1}{Q} \sum_{q=1}^{Q} \mathrm{AMS}_q(h, \gamma), \tag{11}$$

where for $(q = 1, \ldots, Q)$,

$$\mathrm{AMS}_q(h, \gamma) = \frac{1}{m} \sum_{t=T_1-qm+1}^{T_1-qm+m} \left(Y_{t+\tau} - \widehat{m}_{\mathrm{wll},q}(\boldsymbol{X}_t)\right)^2$$

and $\{\widehat{m}_{\mathrm{wll},q}(\cdot)\}$ is the weighted local linear mean estimate computed using (7) from the sample $\{(Y_{t+\tau}, \boldsymbol{X}_t), 1 \leq t \leq T_1 - qm\}$. To account for both pre- and postbreak samples, we need to use an optimal bandwidth for each subsample, denoted as $h_{(1)}$ and $h_{(2)}$, respectively. To determine these bandwidths, we utilize Gaussian kernel for the regression and perform multifold cross-validation as well. Ten candidate values for each bandwidth are chosen to be equidistant within the range $[10^{-2}, 10] \cdot \tilde{h}_{(i)}$, where $\tilde{h}_{(i)}$ represents the theoretically optimal bandwidth for subsample i under Gaussian kernel. Specifically, $\tilde{h}_{(1)}$ is calculated as $1.06 \cdot \sigma(\boldsymbol{X}_{(1)}) T_1^{-1/(4+d)}$, while $\tilde{h}_{(2)}$ is calculated as $1.06 \cdot \sigma(\boldsymbol{X}_{(2)})(T - T_1)^{-1/(4+d)}$, where T_1 represents the size of the prebreak subsample, T represents the total sample size, and d represents the number of covariates. Additionally, we adopt $m = [0.1T]$ and $Q = 4$ as recommended in Cai et al. (2000). It is worth noting that the selected bandwidths are not particularly sensitive to the choice of m and Q, provided that the product mQ is large enough to ensure stable prediction error estimation. Alternatively, one can use the nonparametric type Akaike information criterion selector as proposed in Cai and Tiwari (2000); see, for instance, Cai and Tiwari (2000) for details.

2.5 Estimation of Break Date

When the break date T_1 is unknown, it can be estimated using the method proposed by Mohr and Selk (2020). The objective is to estimate the rescaled change point s_0. The estimator itself is based on a Kolmogorov-Smirnov functional of the marked empirical process of residuals; that is

$$\hat{\mathcal{T}}_T(s, \boldsymbol{z}) = \frac{1}{T} \sum_{t=1}^{\lfloor Ts \rfloor} (Y_t - \hat{m}_T(\boldsymbol{X}_t))\, \omega_T(\boldsymbol{X}_t) \mathbb{1}\left(\boldsymbol{X}_t \leq \boldsymbol{z}\right)$$

for $s \in [0, 1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, where $\boldsymbol{x} \leq \boldsymbol{y}$ is short for $x_j \leq y_j$ for all $j = 1, \ldots, d$, $\omega_T(\bullet) = \mathbb{1}\{\bullet \in [-(\log T)^{\frac{1}{d+1}}, (\log T)^{\frac{1}{d+1}}]^d\}$ and for simplicity, $\hat{m}_T(\cdot)$ is the Naradaya-Watson estimator[3], namely

$$\hat{m}_T(\boldsymbol{x}) = \frac{\sum_{t=1}^T K_h(\boldsymbol{x} - \boldsymbol{X}_t) Y_t}{\sum_{t=1}^T K_h(\boldsymbol{x} - \boldsymbol{X}_t)}.$$

The truncation of the domain of $\boldsymbol{X}_t$ to a compact set within $\mathbb{R}^d$ by the function $\omega_T(\bullet)$ is motivated by the fact that kernel estimators only perform well in regions where there are many observations and rather poorly on the edges and outside of the sample space. Therefore, the nice asymptotic properties can not be expected on the whole domain of $\mathbb{R}^d$. Then, s_0 is estimated by

$$\hat{s}_T := \min\left\{ s : \sup_{\boldsymbol{z}\in\mathbb{R}^d} |\hat{\mathcal{T}}_T(s, \boldsymbol{z})| = \sup_{\bar{s}\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} |\hat{\mathcal{T}}_T(\bar{s}, \boldsymbol{z})| \right\}. \tag{12}$$

Note that $\hat{s}_T = \lfloor T\hat{s}_T \rfloor / T$. Under some regularity conditions; see, for instance, Assumptions I - TX.2 in Mohr and Selk (2020), it follows from Mohr and Selk (2020) that $\hat{s}_T$ is a consistent estimate of s_0 with the convergence rate T. The reader is referred to the paper by Mohr and Selk (2020) for details. Therefore, $\hat{s}_T$ in (12) is used in our simulation and empirical studies conducted in Sections and 3 and 4, respectively.

3 Monte Carlo Simulation Studies

In order to evaluate to finite sample performance of our proposed estimator, we consider two basic models; that is

(IID) $Y_{t+\tau} = m_t(X_t) + \sigma(X_t)\varepsilon_t$, where $X_t, \varepsilon_t \sim \mathcal{N}(0, 1)$ i.i.d.

(TS) $Y_{t+\tau} = m_t(X_t) + \sigma(X_t)\varepsilon_t$, where $X_t = 0.4X_{t-1} + \eta_t$, and $\eta_t, \varepsilon_t \sim \mathcal{N}(0, 1)$ i.i.d.

(AR) $Y_{t+\tau} = m_t(X_t) + \sigma(X_t)\varepsilon_t$, where $X_t = Y_{t-1}$ and $\eta_t, \varepsilon_t \sim \mathcal{N}(0, 1)$ i.i.d.

[3] Of course, one can use the local linear fitting scheme.

with a break in variance $\sigma^2(x) = (0.1 + \sigma_0|x|)\,\mathbb{1}\,(t \leq T_1) + (0.2 + \sigma_0|x|)\,\mathbb{1}\,(t > T_1)$. We generate data for both the homoscedastic case $\sigma_0 = 0$ and the heteroscedastic case $\sigma_0 = 1$. The prediction function is modeled in six different scenarios

$$m_t(x) = m_{(1)}(x)\mathbb{1}\,(t \leq T_1) + m_{(2)}(x)\mathbb{1}\,(t > T_1), \tag{M1}$$

where $m_{(1)}(x) = \sin(x)$ and $m_{(2)}(x) = b_1 \sin(x)$ with $b_1 = 0.9$, 0.7 and 0.5, respectively, so that the break size function $\lambda_1(x) = (1 - b_1)\sin(x)$ characterized by b_1, and

$$m_t(x) = m_{(1)}(x)\mathbb{1}\,(t \leq T_1) + m_{(2)})(x)\mathbb{1}\,(t > T_1), \tag{M2}$$

where $m_{(1)} = x(1 + \cos(x))$ and $m_{(2)}(x) = x(1 + \cos(b_2 x))$ with $b_2 = 1.1$, 1.3 and 1.5, respectively, so that the break size function $\lambda_2(x) = x(\cos(x) - \cos(b_2 x))$ characterized by b_2. The pre-break sample size is defined as a proportion of the full-sample, $T_1 = \lfloor T s_0 \rfloor$ with $s_0 \in \{0.7, 0.8, 0.9\}$, with sample sizes of $T \in \{500, 1000\}$. The simulation is iterated $M = 1000$ times. Both sets of scenarios M1 and M2 represent two different mean functions with different sizes of break. We shall evaluate whether the size of the break in both the mean and variance influences the forecasting performance of our proposed estimator. We distinguish the cases when s_0 is known, or unknown and estimated by $\hat{s}_T$ using (12). We use the Gaussian kernel for estimating the mean function $\widehat{m}(\cdot)$, together with the bandwidth h and the weight γ determined by the multifold cross-validation in (11). As mentioned in Section 2, the estimator $\widehat{m}_{\text{wll}}(\cdot)$ exhibits bias. Consequently, in this simulation exercise, we implement the proposed bias correction to address this issue.

In order to evaluate forecasting performance, we employ the mean squared forecasting error of one to four step ahead forecasts by comparing our weighted local linear estimator ("wll") and the forecast using post-break estimator ("pb"). The τ-step ahead forecast for Y_t computed at time T using method i is denoted as $\widehat{Y}_{i,T+\tau}$ with $\tau = 1, 2, 3, 4$, and $i =$ wll or, pb. That is,

$$\widehat{Y}_{\text{wll},T+\tau} = \widehat{m}_{\text{wll,c}}(X_T),$$

where $\widehat{m}_{\text{wll,c}}(\cdot)$ is computed using (9), while $\widehat{Y}_{\text{pb},T+\tau}$ is based on local linear estimator using post-break observations only. We use a fixed estimation window from $t = 1, \ldots, T$. The MSFE for each method is calculated as

$$\text{MSFE}_{i,\tau} = \frac{1}{M} \sum_{m=1}^{M} \left(Y_{i,T+\tau}^{(m)} - \widehat{Y}_{i,T+\tau}^{(m)} \right)^2,$$

where $\widehat{Y}_{i,T+\tau}^{(m)}$ is the forecasted value for $Y_{T+\tau}$ computed using method i for the m-th replication.

Tables 1 (homoscedastic errors) and 2 (heteroscedastic errors) present simulation results for IID data. In contrast, Tables 3 (homoscedastic errors) and 4 (heteroscedastic errors) display results for TS data, where the regressor follows an AR(1) process. Additionally, Tables 5 (homoscedastic errors) and 6 (heteroscedastic errors) showcase results for AR data, where the first-lag dependent data serves as the regressor. In all cases, the sample size is $T = 500$ or $T = 1000$. These tables contain the MSFE of the weighted local linear estimator relative to the post-break estimator; that is $\text{MSFE}_j/\text{MSFE}_i$, where j stands for the weighted local linear estimator and i represents the post-break estimator ($\gamma = 0$). If the number is less than 1, then the former estimator performs better than the latter, and vice versa, while a relative MSFE of 1 shows the equal forecasting performance of both estimators.

Table 1 presents the relative MSFEs for IID data with a break in homoscedastic variance. Across all forecast horizons, we observe that our proposed estimator consistently outperforms the post-break estimator, as evidenced by relative MSFEs lower than 1. In some cases, the relative MSFEs are slightly above 1, indicating that our proposed estimator performs at least as well as the post-break estimator. When comparing across s_0, we notice a decreasing trend in relative MSFEs as s_0 increases. This suggests that as the post-break sample size decreases,

DGP	s_0	s_0 known				s_0 estimated			
		$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$
$T=500$									
$b_1=0.9$	0.7	0.967	1.006	0.973	0.988	0.984	0.995	0.984	1.000
	0.8	0.937	0.922	0.916	0.935	0.993	1.001	1.000	0.995
	0.9	0.879	0.728	0.734	0.745	0.998	0.912	0.996	0.967
$b_1=0.7$	0.7	0.988	0.978	0.976	0.981	0.979	0.994	0.998	0.984
	0.8	0.968	0.941	0.942	0.600	0.997	1.001	0.996	0.988
	0.9	0.837	0.853	0.639	0.308	0.989	0.975	1.004	0.951
$b_1=0.5$	0.7	0.999	0.995	0.983	1.001	0.996	0.996	0.995	1.001
	0.8	1.003	0.979	0.977	0.986	1.003	0.996	0.997	1.000
	0.9	0.890	0.916	0.843	0.826	0.991	1.000	0.979	0.999
$b_2=1.1$	0.7	0.994	0.983	0.987	0.977	0.997	0.998	0.997	0.995
	0.8	0.959	0.979	0.976	0.965	0.999	1.000	0.998	0.990
	0.9	0.968	0.894	0.924	0.653	1.000	1.000	0.998	0.988
$b_2=1.3$	0.7	0.994	0.996	0.993	0.992	1.001	0.999	1.002	0.999
	0.8	0.970	0.985	0.968	0.986	0.997	0.998	0.992	0.995
	0.9	0.940	0.970	0.792	0.885	1.001	0.999	0.998	0.999
$b_2=1.5$	0.7	0.994	0.997	0.997	1.002	1.001	0.998	1.002	1.000
	0.8	0.988	0.970	0.997	0.984	1.001	0.939	0.996	0.995
	0.9	0.972	0.971	0.935	0.977	0.995	0.998	1.000	1.001
$T=1000$									
$b_1=0.9$	0.7	0.972	0.986	0.968	0.982	0.996	1.004	0.916	0.982
	0.8	0.944	0.968	0.942	0.930	0.992	0.992	0.588	0.986
	0.9	0.806	0.565	0.668	0.812	0.994	0.943	0.939	0.986
$b_1=0.7$	0.7	0.977	0.998	0.982	0.984	0.993	1.004	0.993	0.987
	0.8	0.974	0.953	0.866	0.890	0.995	0.985	0.933	0.973
	0.9	0.785	0.663	0.703	0.793	0.986	0.960	1.005	0.976
$b_1=0.5$	0.7	0.986	0.984	0.983	0.985	1.000	0.998	0.999	1.000
	0.8	0.993	0.973	0.998	0.978	0.999	0.997	1.000	1.003
	0.9	0.909	0.682	0.875	0.714	0.992	1.002	0.995	0.969
$b_2=1.1$	0.7	0.997	1.001	0.998	0.984	0.997	0.999	1.000	0.995
	0.8	0.958	0.946	0.986	0.984	0.997	0.997	1.000	0.997
	0.9	0.861	0.851	0.908	0.926	0.952	1.000	0.965	0.989
$b_2=1.3$	0.7	0.992	0.993	0.991	0.982	0.996	0.999	0.997	0.987
	0.8	0.971	0.886	0.972	0.989	0.996	0.981	0.997	0.995
	0.9	0.963	0.958	0.897	0.906	0.998	0.996	0.996	1.002
$b_2=1.5$	0.7	1.003	0.993	1.001	0.994	0.997	1.000	0.996	0.999
	0.8	0.991	0.996	0.981	1.005	1.000	1.000	0.999	1.000
	0.9	0.974	0.982	0.969	0.935	0.997	0.998	1.001	0.998

Table 1: MSFE for WLL estimator relative to the post-break estimator. IID data with a break in homoscedastic variance $\sigma^2(x) = 0.1 \cdot \mathbb{1}(t \leq T_1) + 0.2 \cdot \mathbb{1}(t > T_1)$, both known (the left panel) and estimated s_0 (the right panel). Forecast horizon $\tau = 1, 2, 3, 4$. Sample size $T = 500$ (the top panel) and $T = 1000$ (the bottom panel) with $M = 1000$ Monte-Carlo iterations.

DGP	s_0	s_0 known				s_0 estimated			
		$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$
$T = 500$									
$b_1 = 0.9$	0.7	0.977	0.976	0.969	0.980	0.991	1.003	0.987	0.989
	0.8	0.918	0.888	0.906	0.938	1.000	0.999	0.998	1.013
	0.9	0.800	0.632	0.842	0.706	1.006	0.993	0.989	0.998
$b_1 = 0.7$	0.7	0.993	0.983	0.939	0.973	0.988	0.983	0.987	0.998
	0.8	0.960	0.907	0.914	0.919	1.000	0.972	0.962	0.983
	0.9	0.866	0.692	0.526	0.481	0.995	0.976	1.005	0.991
$b_1 = 0.5$	0.7	0.972	0.989	0.973	0.990	0.999	1.000	0.982	0.937
	0.8	0.995	0.917	0.924	0.939	0.587	0.995	0.997	0.514
	0.9	0.802	0.787	0.770	0.816	0.988	0.994	1.002	0.995
$b_2 = 1.1$	0.7	0.996	0.988	0.996	0.997	0.993	0.999	1.001	0.970
	0.8	0.956	0.993	0.972	0.945	0.999	1.002	0.998	1.000
	0.9	0.874	0.929	0.907	0.922	0.997	0.994	1.000	0.985
$b_2 = 1.3$	0.7	0.996	0.985	0.980	1.000	0.998	1.002	0.995	0.994
	0.8	0.988	0.988	0.981	0.986	1.000	1.002	0.997	0.999
	0.9	0.939	0.650	0.968	0.888	0.999	1.003	0.999	1.000
$b_2 = 1.5$	0.7	0.998	0.988	0.997	0.999	1.001	0.999	1.000	0.999
	0.8	0.983	0.965	0.994	0.987	0.998	0.997	0.998	1.002
	0.9	0.982	0.941	0.836	0.932	1.001	1.001	0.996	1.001
$T = 1000$									
$b_1 = 0.9$	0.7	0.982	0.989	0.965	0.954	0.989	1.005	0.986	0.991
	0.8	0.939	0.952	0.852	0.650	0.994	0.991	1.003	0.885
	0.9	0.801	0.868	0.792	0.665	0.999	0.980	1.000	0.996
$b_1 = 0.7$	0.7	0.977	0.995	0.985	0.983	0.996	0.946	0.990	0.987
	0.8	0.957	0.963	0.865	0.854	1.003	0.996	0.929	1.000
	0.9	0.805	0.781	0.651	0.653	0.982	0.995	1.001	0.924
$b_1 = 0.5$	0.7	0.990	0.977	0.978	0.992	1.004	0.997	1.006	1.001
	0.8	0.984	0.968	0.902	0.927	1.001	0.993	0.923	0.949
	0.9	0.858	0.681	0.623	0.720	0.984	0.987	1.002	0.974
$b_2 = 1.1$	0.7	0.995	0.994	0.991	0.996	0.998	0.993	0.998	0.989
	0.8	0.979	0.931	0.983	0.980	0.995	0.998	0.996	0.997
	0.9	0.948	0.950	0.904	0.930	0.998	0.997	1.000	0.997
$b_2 = 1.3$	0.7	0.995	0.997	0.981	0.989	0.991	0.998	0.982	0.996
	0.8	0.988	0.759	0.984	0.992	0.992	0.989	0.995	0.997
	0.9	0.948	0.953	0.829	0.939	0.996	0.995	0.999	0.999
$b_2 = 1.5$	0.7	0.985	0.992	0.998	0.987	0.999	1.001	1.001	0.993
	0.8	0.994	0.999	0.993	1.002	0.999	0.999	0.999	1.000
	0.9	0.948	0.963	0.903	0.856	1.000	1.005	1.000	1.000

Table 2: MSFE for WLL estimator relative to the post-break estimator. IID data with a break in heteroscedastic variance $\sigma^2(x) = (0.1 + |x|)\mathbb{1}(t \leq T_1) + (0.2 + |x|)\mathbb{1}(t > T_1)$, both known (the left panel) and estimated s_0 (the right panel). Forecast horizon $\tau = 1, 2, 3, 4$. Sample size $T = 500$ (the top panel) and $T = 1000$ (the bottom panel) with $M = 1000$ Monte-Carlo iterations.

DGP	s_0	s_0 known				s_0 estimated			
		$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$
$T=500$									
$b_1=0.9$	0.7	0.982	0.972	0.991	0.984	1.003	0.984	0.985	0.986
	0.8	0.914	0.965	0.886	0.890	0.996	0.985	0.991	0.998
	0.9	0.786	0.657	0.688	0.816	0.987	0.985	0.993	0.992
$b_1=0.7$	0.7	1.000	0.916	0.986	0.991	0.994	0.993	0.992	0.999
	0.8	0.969	0.971	0.957	0.819	0.998	0.952	1.004	0.991
	0.9	0.776	0.761	0.611	0.604	0.988	0.986	0.996	0.993
$b_1=0.5$	0.7	1.021	0.992	0.888	0.992	0.998	0.996	0.993	0.997
	0.8	1.001	0.950	0.988	0.954	1.000	0.995	0.981	0.998
	0.9	0.941	0.779	0.854	0.883	1.004	0.999	1.001	0.998
$b_2=1.1$	0.7	0.994	0.994	0.994	0.994	0.982	0.997	0.996	0.995
	0.8	0.883	0.978	0.985	0.983	1.001	1.001	1.002	1.004
	0.9	0.957	0.947	0.927	0.954	0.996	0.999	0.994	0.997
$b_2=1.3$	0.7	0.990	0.999	0.996	0.990	0.995	0.996	0.996	0.999
	0.8	0.994	0.993	0.985	0.975	0.989	0.997	0.997	0.997
	0.9	0.981	0.950	0.905	0.970	0.998	1.000	1.001	0.987
$b_2=1.5$	0.7	1.015	0.997	0.999	1.000	1.001	0.999	0.993	0.998
	0.8	1.028	0.997	0.995	0.996	1.002	1.000	1.000	1.000
	0.9	1.043	0.980	0.932	0.936	1.000	1.001	1.000	0.993
$T=1000$									
$b_1=0.9$	0.7	0.975	0.993	0.981	0.973	0.997	1.001	1.001	0.993
	0.8	0.956	0.955	0.954	0.920	0.997	0.935	0.986	0.960
	0.9	0.894	0.754	0.566	0.787	1.003	0.992	0.975	0.988
$b_1=0.7$	0.7	0.984	0.988	0.987	0.965	0.992	0.997	1.002	1.001
	0.8	0.982	0.960	0.942	0.896	0.994	0.998	0.954	0.967
	0.9	0.788	0.890	0.512	0.695	0.997	0.983	0.992	0.963
$b_1=0.5$	0.7	1.008	0.979	0.990	0.969	1.001	0.997	0.998	0.956
	0.8	0.976	0.985	0.954	0.959	1.004	0.984	0.997	0.998
	0.9	0.931	0.772	0.617	0.371	1.009	0.996	0.987	1.005
$b_2=1.1$	0.7	0.997	1.000	0.988	0.993	0.995	0.998	0.892	0.994
	0.8	0.941	0.991	0.978	0.975	1.001	0.998	1.001	1.002
	0.9	0.966	0.913	0.906	0.951	0.973	0.998	1.000	0.997
$b_2=1.3$	0.7	0.993	1.002	0.995	0.991	1.001	1.001	0.990	0.997
	0.8	0.984	0.994	0.960	0.982	0.996	0.995	0.998	1.001
	0.9	0.959	0.931	0.921	0.732	1.000	1.002	0.999	0.998
$b_2=1.5$	0.7	1.013	0.998	0.998	0.999	1.005	0.998	0.994	1.000
	0.8	1.020	0.992	0.968	0.989	1.002	0.997	1.001	0.999
	0.9	1.030	0.882	0.961	0.953	1.002	0.996	1.001	1.003

Table 3: MSFE for WLL estimator relative to the post-break estimator. TS data with a break in homoscedastic variance $\sigma^2(x) = 0.1 \cdot \mathbb{1}(t \le T_1) + 0.2 \cdot \mathbb{1}(t > T_1)$, both known (the left panel) and estimated s_0 (the right panel). Forecast horizon $\tau = 1, 2, 3, 4$. Sample size $T = 500$ (the top panel) and $T = 1000$ (the bottom panel) with $M = 1000$ Monte-Carlo iterations.

DGP	s_0	s_0 known				s_0 estimated			
		$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$
$T = 500$									
$b_1 = 0.9$	0.7	0.989	0.974	0.982	0.980	0.997	0.998	0.994	0.996
	0.8	0.910	0.887	0.753	0.923	0.997	0.999	1.000	0.974
	0.9	0.751	0.675	0.721	0.655	0.995	0.990	0.996	0.991
$b_1 = 0.7$	0.7	0.964	0.984	0.889	0.956	0.978	0.989	0.988	0.989
	0.8	0.853	0.797	0.937	0.877	0.995	0.996	0.995	0.993
	0.9	0.876	0.755	0.737	0.695	0.993	0.995	0.980	0.994
$b_1 = 0.5$	0.7	0.995	0.980	0.991	0.985	1.001	0.991	0.996	0.995
	0.8	0.970	0.847	0.587	0.923	0.997	0.993	0.986	0.968
	0.9	0.848	0.656	0.783	0.732	1.004	0.997	0.995	0.999
$b_2 = 1.1$	0.7	0.977	0.993	0.986	0.999	1.002	1.001	0.993	0.996
	0.8	0.984	0.981	0.990	0.991	0.989	0.997	0.997	0.990
	0.9	0.977	0.955	0.964	0.931	0.992	0.995	1.000	0.995
$b_2 = 1.3$	0.7	0.987	0.995	0.994	0.989	1.004	1.005	0.996	0.994
	0.8	0.993	0.992	0.992	0.976	0.997	1.003	0.994	0.986
	0.9	0.969	0.688	0.911	0.920	0.999	1.002	0.997	0.993
$b_2 = 1.5$	0.7	1.008	0.996	0.989	0.999	0.998	0.994	0.998	0.998
	0.8	1.007	0.979	0.986	0.968	1.000	0.999	0.983	0.998
	0.9	0.986	0.924	0.961	0.983	1.000	0.984	0.999	0.997
$T = 1000$									
$b_1 = 0.9$	0.7	0.978	0.981	0.979	0.966	0.991	0.993	0.992	0.996
	0.8	0.974	0.932	0.928	0.891	0.996	0.995	0.988	0.991
	0.9	0.838	0.733	0.775	0.790	0.996	0.998	0.984	0.890
$b_1 = 0.7$	0.7	0.904	0.965	0.971	0.773	0.981	0.981	0.993	0.998
	0.8	0.920	0.912	0.952	0.680	0.999	0.951	0.987	0.983
	0.9	0.859	0.863	0.711	0.520	0.937	0.994	0.992	0.996
$b_1 = 0.5$	0.7	0.909	0.971	1.000	0.994	1.003	0.991	0.995	0.997
	0.8	0.977	0.968	0.949	0.886	1.001	0.954	0.990	0.948
	0.9	0.826	0.664	0.519	0.832	1.000	0.588	0.961	0.985
$b_2 = 1.1$	0.7	0.980	1.002	0.998	0.982	0.997	1.000	1.000	0.997
	0.8	0.980	0.994	0.994	0.999	1.001	1.000	0.993	0.991
	0.9	0.800	0.675	0.912	0.873	0.999	0.999	0.996	1.003
$b_2 = 1.3$	0.7	0.985	0.999	0.999	0.988	0.997	0.994	1.011	0.996
	0.8	0.957	0.994	0.980	0.996	1.001	0.995	0.995	0.966
	0.9	0.951	0.945	0.715	0.937	0.997	0.996	0.994	0.996
$b_2 = 1.5$	0.7	1.015	0.996	0.992	0.995	1.002	1.000	1.001	0.996
	0.8	1.019	0.994	0.985	0.939	1.000	0.994	0.999	0.998
	0.9	0.986	0.929	0.946	0.957	1.002	1.001	0.998	0.999

Table 4: MSFE for WLL estimator relative to the post-break estimator. TS data with a break in heteroscedastic variance $\sigma^2(x) = (0.1 + |x|)\mathbb{1}(t \leq T_1) + (0.2 + |x|)\mathbb{1}(t > T_1)$, both known (the left panel) and estimated s_0 (the right panel). Forecast horizon $\tau = 1, 2, 3, 4$. Sample size $T = 500$ (the top panel) and $T = 1000$ (the bottom panel) with $M = 1000$ Monte-Carlo iterations.

DGP	s_0	s_0 known				s_0 estimated			
		$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$
$T=500$									
$b_1=0.9$	0.7	0.984	0.960	0.938	0.938	1.003	0.996	0.991	0.996
	0.8	0.896	0.863	0.844	0.826	0.977	0.965	0.974	0.961
	0.9	0.757	0.743	0.792	0.683	1.000	0.974	0.968	0.979
$b_1=0.7$	0.7	1.002	0.987	0.972	0.933	1.010	0.998	0.987	0.979
	0.8	0.957	0.882	0.873	0.907	0.987	0.992	0.982	1.000
	0.9	0.858	0.819	0.787	0.729	1.006	0.923	0.982	0.985
$b_1=0.5$	0.7	1.039	1.032	1.023	1.005	1.003	1.002	0.937	0.984
	0.8	1.061	1.032	1.042	0.979	1.008	1.008	1.008	1.006
	0.9	1.005	0.988	0.967	0.742	1.003	0.999	0.973	0.998
$b_2=1.1$	0.7	1.005	0.999	1.007	0.993	0.988	0.986	0.989	0.999
	0.8	0.976	0.987	0.987	0.962	0.998	0.997	0.997	0.998
	0.9	0.902	0.825	0.915	0.880	1.002	1.001	0.997	0.999
$b_2=1.3$	0.7	1.021	1.010	1.016	1.005	0.995	0.996	0.992	0.990
	0.8	1.012	1.005	1.002	0.983	1.003	0.994	1.001	1.003
	0.9	0.951	0.916	0.907	0.809	1.006	1.006	1.000	1.003
$b_2=1.5$	0.7	1.044	1.047	1.051	1.018	1.008	1.008	1.017	1.008
	0.8	1.056	1.043	1.031	0.999	1.005	1.005	1.005	1.004
	0.9	1.048	1.027	1.002	0.978	1.004	1.005	1.002	0.999
$T=1000$									
$b_1=0.9$	0.7	0.975	0.880	0.909	0.985	0.986	0.995	0.982	1.001
	0.8	0.968	0.964	0.913	0.972	1.006	0.963	0.950	0.991
	0.9	0.925	0.966	0.887	0.854	1.008	1.001	0.997	1.001
$b_1=0.7$	0.7	1.012	0.955	1.000	1.024	0.975	1.004	1.003	0.998
	0.8	1.007	0.984	0.934	0.976	1.000	0.988	0.992	0.954
	0.9	0.994	0.989	0.971	0.900	0.974	1.016	0.928	1.011
$b_1=0.5$	0.7	0.991	0.980	0.980	0.981	0.994	0.994	0.991	0.976
	0.8	0.993	0.986	0.992	0.954	0.974	0.995	0.995	0.989
	0.9	0.973	0.968	0.979	0.955	0.979	0.981	0.983	0.993
$b_2=1.1$	0.7	0.998	1.013	0.978	0.991	1.000	1.002	0.997	1.000
	0.8	0.974	0.937	0.982	0.978	0.998	0.997	0.882	0.989
	0.9	0.754	0.902	0.959	0.825	0.917	0.961	0.980	0.996
$b_2=1.3$	0.7	1.012	1.022	1.017	1.000	0.993	0.952	0.969	0.987
	0.8	1.011	0.988	1.011	1.027	0.978	0.999	0.979	0.986
	0.9	1.010	0.910	0.989	1.003	0.989	0.998	0.985	0.978
$b_2=1.5$	0.7	0.994	1.000	0.992	0.996	0.999	0.999	0.995	0.996
	0.8	1.004	0.995	0.997	0.988	0.996	0.995	0.993	0.991
	0.9	1.002	1.017	1.006	0.991	0.976	1.003	1.003	0.992

Table 5: MSFE for WLL estimator relative to the post-break estimator. AR data with a break in homoscedastic variance $\sigma^2(x) = 0.1 \cdot \mathbb{1}(t \leq T_1) + 0.2 \cdot \mathbb{1}(t > T_1)$, both known (the left panel) and estimated s_0 (the right panel). Forecast horizon $\tau = 1, 2, 3, 4$. Sample size $T = 500$ (the top panel) and $T = 1000$ (the bottom panel) with $M = 1000$ Monte-Carlo iterations.

DGP	s_0	s_0 known				s_0 estimated			
		$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$
$T=500$									
$b_1=0.9$	0.7	0.993	0.982	0.984	0.995	0.921	0.973	0.990	0.990
	0.8	0.984	0.973	0.919	0.927	1.009	0.650	0.966	0.966
	0.9	0.937	0.817	0.718	0.878	1.001	0.986	0.994	0.934
$b_1=0.7$	0.7	0.997	0.981	0.996	0.986	0.927	0.970	0.916	0.989
	0.8	0.974	0.982	0.964	0.877	1.003	1.003	0.984	0.984
	0.9	0.705	0.863	0.690	0.789	0.963	0.946	0.992	0.986
$b_1=0.5$	0.7	1.079	1.023	1.034	0.927	0.857	0.806	0.971	0.980
	0.8	1.092	0.678	0.733	0.836	0.946	0.975	0.965	0.905
	0.9	0.908	0.481	0.817	0.493	0.851	0.971	0.516	1.020
$b_2=1.1$	0.7	0.997	1.016	1.008	1.015	0.991	0.987	0.959	1.012
	0.8	0.991	1.023	1.013	0.989	0.968	0.994	0.981	0.992
	0.9	1.005	0.963	0.914	0.974	0.996	0.996	0.968	1.012
$b_2=1.3$	0.7	1.011	1.012	1.010	1.009	1.004	1.001	1.003	1.004
	0.8	1.024	1.026	1.025	0.968	0.990	0.992	0.999	0.991
	0.9	0.993	0.972	0.848	0.944	1.003	1.000	0.998	1.003
$b_2=1.5$	0.7	1.064	1.065	0.999	0.885	1.030	1.011	0.994	0.987
	0.8	1.086	1.048	1.001	1.008	0.942	0.843	1.002	0.570
	0.9	1.075	1.066	0.696	0.824	1.015	1.008	0.899	0.932
$T=1000$									
$b_1=0.9$	0.7	0.994	1.000	1.000	0.994	1.000	1.000	1.000	0.986
	0.8	0.998	0.990	0.994	0.996	1.000	1.000	0.999	1.000
	0.9	0.998	0.973	0.923	0.959	1.000	0.988	1.000	0.997
$b_1=0.7$	0.7	0.992	0.997	1.000	1.000	0.998	1.036	1.586	0.997
	0.8	0.999	0.970	0.996	0.932	1.000	0.998	0.966	0.984
	0.9	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
$b_1=0.5$	0.7	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
	0.8	1.000	1.000	0.998	0.916	1.000	1.000	1.000	1.000
	0.9	0.999	0.999	0.898	0.876	0.999	0.996	0.994	0.979
$b_2=1.1$	0.7	1.000	0.999	0.995	0.999	1.000	1.000	0.999	0.999
	0.8	0.998	1.000	1.000	0.999	1.000	1.000	1.000	0.998
	0.9	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000
$b_2=1.3$	0.7	1.000	1.000	1.000	0.996	1.000	0.999	1.000	0.999
	0.8	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	0.9	1.000	1.000	0.998	0.996	1.000	1.000	1.000	1.000
$b_2=1.5$	0.7	1.000	0.999	0.999	1.000	1.000	1.000	1.000	1.000
	0.8	1.000	0.999	0.996	0.962	1.000	1.000	0.999	0.998
	0.9	0.992	0.999	1.000	0.998	1.000	1.000	1.000	1.000

Table 6: MSFE for WLL estimator relative to the post-break estimator. AR data with a break in heteroscedastic variance $\sigma^2(x) = (0.1 + |x|)\mathbb{1}(t \leq T_1) + (0.2 + |x|)\mathbb{1}(t > T_1)$, both known (the left panel) and estimated s_0 (the right panel). Forecast horizon $\tau = 1, 2, 3, 4$. Sample size $T = 500$ (the top panel) and $T = 1000$ (the bottom panel) with $M = 1000$ Monte-Carlo iterations.

our weighted local linear estimator successfully enhances the forecasting power using the pre-break sample. Furthermore, when comparing different break sizes, we notice that the relative MSFEs tend to increase as b_1 decreases or b_2 increases. This implies that as the break size increases, the forecast performance based on our weighted local linear estimator gradually becomes similar to the post-break estimator. Additionally, when comparing between the cases of known and unknown s_0, we observe that the relative MSFEs tend to be higher in the case of unknown s_0. This is because the estimated s_0 introduces estimation risk to our WLL estimator, leading to relatively poorer forecast performance, as indicated by relatively higher MSFEs. This tendency is observed for both sample sizes $T = 500, 1000$. Furthermore, when considering the heteroscedastic case, Table 2 demonstrates higher relative MSFEs compared to the homoscedastic case. Nevertheless, our proposed estimator still outperforms the post-break estimator. Table 3, 4, 5 and 6 report the relative MSFEs for TS and AR data with breaks in both homoscedastic and heteroscedastic variance. These tables show similar results to the IID case.

As demonstrated in Tables 1 to 4, the majority of forecast horizons (τ =1, 2, 3, and 4) exhibit a relative mean squared forecast error that is less than or close to 1. However, there are a few exceptions where the relative MSFE is slightly higher than 1, indicating that our weighted local linear estimator performs at least as well as the post-break estimator. This is particularly evident when the break magnitude is large and/or when s_0 is estimated. Despite these exceptions, our simulation results consistently indicate that our WLL estimator enhances forecasting performance compared to the conventional post-break estimator.

4 An Empirical Example

In our empirical application, we present a forecasting model of GDP growth using yield curve as a predictor. The yield curve in this case is defined as the difference between interest rates ("term spread") on long and short maturity debt, e.g. government debt. In practice, the literature uses either the difference between long-term government bond rate and three-month government bill rate, or instead, the long bond rate minus the overnight rate (e.g. the federal funds rate in the United States). Stock and Watson (1989) and Estrella and Hardouvelis (1991) showed in their empirical studies that a positive slope of the yield curve is associated with future increases in real economic activity six or seven quarters ahead. The marginal predictive power of the yield curve was interpreted as evidence that market participants were able to forecast economic expansions or contractions six or seven quarters in advance. This finding led to the argument that the term structure could be an indicator of monetary policy stance. The reader is referred to Stock and Watson (2003) for a survey of literature. The economic reasoning behind the use of term spread variable for output growth prediction lies in its predictive power on the effectiveness of monetary policy. For example, monetary tightening will lead to a short-term interest rate that is high relative to the long-term rate. These high short-term rate will in turn cause an economic slowdown (Bernanke and Blinder, 1992). However, recent evidence suggests that the forecasting relationship between output growth and yield curve may be subject to structural breaks; see, for instance, Stock and Watson (1999), Giacomini and Rossi (2006), Estrella et al. (2003), Schrimpf and Wang (2010), and references therein. The presence of structural breaks in turn leads to an instability of the model coefficients, which calls its usefulness for forecasting into question. Thus, structural break tests are usually done before forecast is made. Also, the relationship between output growth and yield curve seems to be nonlinear; see, for instance, Figure 1. In this section, therefore, we apply our proposed forecasting method to investigate whether it can improve the forecasts of GDP growth using the slope of the yield curve as the predictor via a nonparametric forecasting technique.

4.1 Econometric Modeling

The forecast is based on the following nonparametric regression model

$$Y_{t+\tau} = m(s_t) + u_t$$

for $\tau = 1$, 2, 3 and 4, where $Y_{t+\tau} = 100 \ln(P_{t+\tau}/P_t)$, P_t is the level of real GDP at time t, and $s_t = r_t^L - r_t^S$ is the slope of the yield curve, defined as the difference between the long term interest rate, r_t^L, and the short term interest rate, r_t^S.

The model specification we employ here represents a nonparametric alternative to the commonly used linear models found in the literature, as referenced in previous works such as Pesaran et al. (2013), Estrella and Hardouvelis (1991), Estrella and Mishkin (1997), and references therein. Our nonparametric modeling approach is motivated by the common observation that real-world data frequently display intricate, nonlinear relationships between the dependent variables and the regressors. For instance, when examining data encompassing two industrialized economies: France and Italy, spanning the time period from 1979Q1 to 2019Q4, as depicted in Figure 1, it shows clearly that there exists a potentially nonlinear relationship between the variables Y_t and s_t.

The advantage of using our nonparametric specification is that we let the data reveal which functional relationship exists between both variables, instead of imposing a linear one. Similar to our simulation studies in Section 3, the Gaussian kernel is used for estimating the mean function $\widehat{m}(\cdot)$ and the bandwidth h and the weight are chosen by the multifold cross-validation in (11). We evaluate the forecasts for horizons $\tau = 1$, 2, 3, and 4 quarters

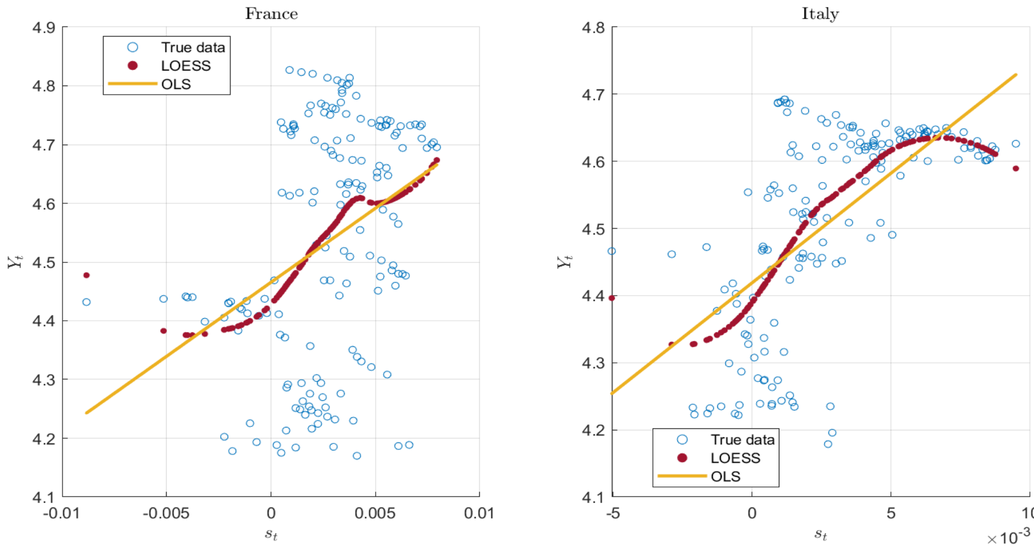

Figure 1: Scatterplot of term spread s_t vs. real GDP growth rate Y_t (blue dots) and locally weighted scatterplot smoothing (LOESS) fitted data (red dots). The yellow solid line represents the Ordinary Least Squares (OLS) regression line $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 s_t$. The sample period spans from 1979Q1 to 2019Q4. The left panel represents France, while the right panel represents Italy.

using Diebold-Mariano test proposed in Diebold and Mariano (1995). Let $e_{i,t} = Y_t - \widehat{Y}_{i,t}$ and $e_{j,t} = Y_t - \widehat{Y}_{j,t}$ be the forecast errors for method i and j, respectively, and choose the loss differential $d_t = e_{i,t}^2 - e_{j,t}^2$. Then, the Diebold-Mariano test is defined as follows

$$\text{DM} = \frac{\bar{d}}{\sqrt{\sum_{\tau=-(T-1)}^{(T-1)} \widehat{\Upsilon}_d(\tau)/T}} \approx \mathcal{N}(0,1),$$

where $\bar{d} = \sum_{t=1}^{T} d_t/T$ is the sample mean of the loss differential, or simply $\text{MSFE}_i - \text{MSFE}_j$, and $\widehat{\Upsilon}_d$ is the associated sample auto-covariance, calculated as follows

$$\widehat{\Upsilon}_d(\tau) = \frac{1}{T} \sum_{t=|\tau|+1}^{T} (d_t - \bar{d})(d_{t-|\tau|} - \bar{d}).$$

To assess the statistical significance of the improved predictive performance achieved by method j, we conduct a hypothesis test comparing it to method i, where method i serves as the benchmark estimator. The null hypothesis (H_0) asserts that there is no significant difference in mean squared forecasting error (MSFE) between the two methods, specifically

$H_0 : \text{MSFE}_i = \text{MSFE}_j$. In contrast, the alternative hypothesis (H_a) posits that method j outperforms method i, i.e., $H_a : \text{MSFE}_i > \text{MSFE}_j$.

In this context, method i represents the benchmark estimator, while method j corresponds to our proposed weighted local linear estimator. Our study considers two benchmark estimators: a nonparametric post-break estimator that employs the most recent 12 quarters of data leading up to the forecasting point and the weighted generalized least squares estimator. The latter, proposed by Lee et al. (2022), is a notable linear estimator designed for forecasting under structural breaks.

This framework allows us to rigorously assess and validate the enhanced predictive ability of our proposed method against established benchmarks.

4.2 Data

Our quarterly GDP and interest rate data are sourced from the GVAR toolbox; see, for example, Mohaddes and Raissi (2020) for details. For our analysis, we focus on two industrialized economies: France and Italy. Our dataset spans from 1979Q1 to 2019Q4, encompassing a total of $T = 164$ observations. In our approach, we adopt a recursive out-of-sample forecasting method. We segment the sample of observations into two segments: the initial T observations serve as the in-sample estimation period, while the remaining observations constitute the pseudo out-of-sample evaluation period. Forecasts are generated recursively throughout the out-of-sample period, using only the available information at each forecast point. As we extend the estimation window, we reevaluate the break date using the method described in Section 2.5. We initiate this scheme by generating our forecast using data up to 2006Q4. Subsequently, we present our findings for the period 2007Q1-2019Q1. This time period represents the era following the collapse of the sub-prime mortgage market.

Note that our interest rate calculations follow this formula: $r_t^S = 0.25(1 + R_t^S/100)$ and $r_t^L = 0.25(1 + R_t^L/100)$. Here, R_t^S and R_t^L denote the short-term and long-term nominal interest rates per annum, expressed as percentages, respectively.

4.3 Empirical Results

We conduct a thorough comparison of forecasting performance between our novel weighted local linear estimator and two benchmark estimators, as detailed in Table 7. To gain insights

Benchmark	Postbreak estimator							
Country	$\tau = 1$		$\tau = 2$		$\tau = 3$		$\tau = 4$	
	DM-test	p-value	DM-test	p-value	DM-test	p-value	DM-test	p-value
France	11.464	0.000	4.805	0.000	3.466	0.000	2.887	0.002
Italy	12.155	0.000	4.866	0.000	3.495	0.000	2.894	0.002
Benchmark	Weighted general least squares estimator							
Country	$\tau = 1$		$\tau = 2$		$\tau = 3$		$\tau = 4$	
	DM-test	p-value	DM-test	p-value	DM-test	p-value	DM-test	p-value
France	2.031	0.021	2.920	0.002	1.330	0.092	0.360	0.360
Italy	3.398	0.000	2.843	0.002	2.226	0.013	2.006	0.022

Table 7: Diebold-Mariano test statistics and their p-values for GDP-growth forecast in two selected countries for different out-of-sample forecasting periods. The null hypothesis H_0 : $\text{MSFE}_i = \text{MSFE}_j$ vs. the alternative hypothesis $H_a : \text{MSFE}_i > \text{MSFE}_j$, where i represents the benchmark and weighted general least squares estimator, and j stands for the weighted local linear estimator. Significant p-values indicate an outperformance of the latter over the former estimator.

into the precision of our break date estimation, we examine the estimated change points, denoted as $\hat{s}_T$, which serve as the foundation for identifying the break date, denoted as $T_1 = \lfloor \hat{s}_T T \rfloor$. Figure 2 displays a wide range of values, typically between 0 and 0.5, signifying that our nonparametric break date estimator successfully detects the change point relatively early within the data samples.

Furthermore, we scrutinize the estimated weights, denoted as $\hat{\gamma}$, assigned to the pre-break samples by our proposed weighted local linear estimator. The observed values fall within the range of 0 to 0.11 are presented in Figure 3, suggesting that the pre-break observations do

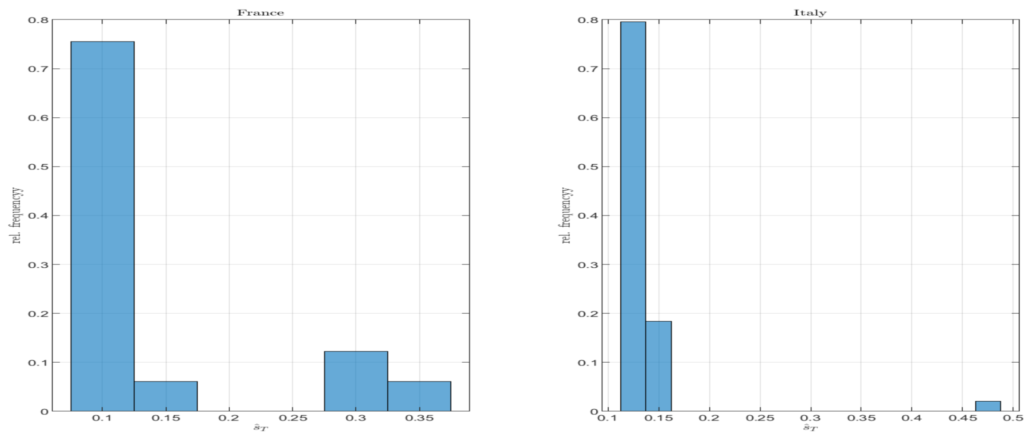

Figure 2: Estimated rescaled change points $\hat{s}_T$.

not significantly contribute to, or marginally enhance, post-break forecasting improvements when utilizing the estimated break date.

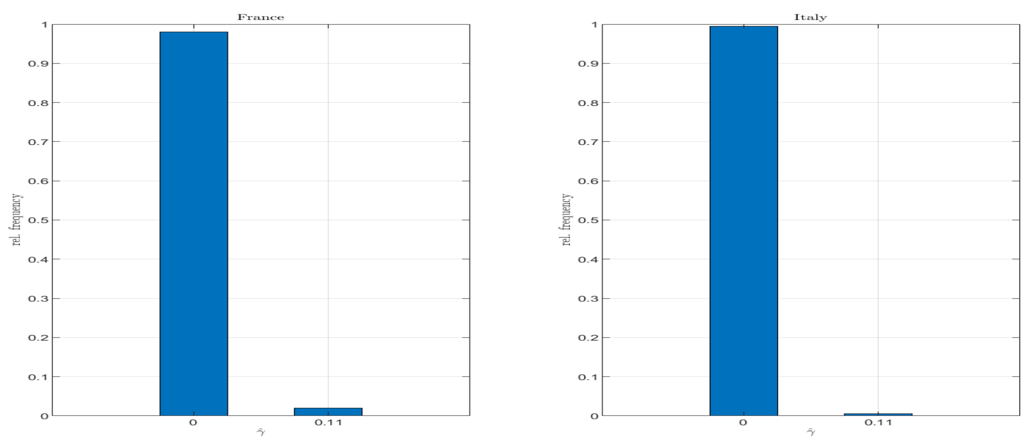

Figure 3: Estimated weights for the pre-break samples $\hat{\gamma}$.

Our evaluation extends to performing Diebold-Mariano (DM) tests for the time period 2007Q1-2019Q1, as presented in Table 7. Across these analyses, we consistently observe substantial DM-test statistics accompanied by highly significant p-values when the post-break estimator is served as the benchmark. Interestingly, for a lower step-ahead forecast ($\tau = 1$ or 2), our proposed method outperforms the weighted least squares approach due to its ability to capture local properties. However, the scenario is different when using the weighted least squares estimator, as insignificant p-values are identified for a higher step-ahead forecast ($\tau = 3$ or 4) for France. This means that for higher step-ahead forecasts, both our proposed method and the weighted least squares approach perform comparably.

This intriguing result suggests that, within the context of our forecasting exercise, our nonparametric forecasting model remains accurate in many instances, even when assuming a single break. It underscores the potential added value of our novel weighted local linear estimator, especially in situations where nonlinear patterns or the presence of multiple structural breaks could potentially impact forecasting performance.

5 Conclusions

When forecasting time series data, structural breaks can present a significant challenge. Existing literature has proposed several methods to handle structural breaks, but they tend to be (semi-)parametric in nature. Typically, these methods incorporate information from the pre-break period by assigning weights between 0 and 1 to the relevant observations. Building on this idea, our paper proposes a similar nonparametric estimator. Our proposed weighted local linear estimator has been shown in previous studies to outperform the usual post-break estimator in parametric cases. However, our study only considers a single break and a single covariate as a predictor, which could be problematic in more complex situations, such as longer time series data with multiple breaks or with missing relevant covariates. In real-world applications, where the break date is unknown, accurate estimation of the break date is essential. To address this issue, future research could explore robust nonparametric methods for identifying multiple breaks in time series data, and extend these methods to the multivariate setting. Such efforts would help to further improve the accuracy and reliability of time series forecasting in the presence of structural breaks.

Disclosure Statement

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere, and, also, we declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bernanke, B. S. and A. S. Blinder (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review 82*(4), 901–921.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated error. *Journal of Econometrics 136*(1), 163–188.

Cai, Z., J. Fan, and Q. Yao (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association 95*(451), 941–956.

Cai, Z. and E. Masry (2000). Nonparametric estimation of additive nonlinear arx time series: Local linear fitting and projections. *Econometric Theory 16*(4), 465–501.

Cai, Z. and R. C. Tiwari (2000). Application of a local linear autoregressive model to bod time series. *Envirometrics 11*(3), 341–350.

Clements, M. P. and D. F. Hendry (2006). Forecasting with breaks, In G. Elliott, C. Granger and A. Timmermann (eds.). Volume 1 of *Handbook of Economic Forecasting*, pp. 605–657. Elsevier.

Clements, M. P. and D. F. Hendry (2011). Forecasting from misspecified Models in the presence of unanticipated location shifts, In, M.P. Clements and D. F. Hendry (eds.). The Oxford Handbook of Economic Forecasting, pp. 271–314. Oxford University Press.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Estrella, A. and G. A. Hardouvelis (1991). The term structure as a predictor of real economic activity. *Journal of Finance 46*(2), 555–576.

Estrella, A. and F. S. Mishkin (1997). The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank. *European Economic Review 41*(7), 1375–1401.

Estrella, A., A. R. Rodrigues, and S. Schich (2003). How stable is the predictive power of the yield curve? Evidence from Germany and the United States. *Review of Economics and Statistics 85*(3), 629–644.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modeling and Its Applications*. Chapman & Hall.

Giacomini, R. and B. Rossi (2006). How stable is the forecasting performance of the yield curve for output growth? *Oxford Bulletin of Economics and Statistics 68*(1), 783–795.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica 75*(4), 1175–1189.

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics 146*(2), 342–350.

Hansen, B. E. and J. S. Racine (2012). Jackknife model averaging. *Journal of Econometrics 167*(1), 38–46.

Hjort, N. L. and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association 98*(464), 879–899.

Lee, T.-H., S. Parsaeian, and A. Ullah (2022). Forecasting under structural breaks using improved weighted estimation. *Oxford Bulletin of Economics and Statistics 84*(6), 1485–1501.

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

Mohaddes, K. and M. Raissi (2020). Compilation, revision and updating of the global var (gvar) database, 1979q2-2019q4. *Mimeo, Judge Business School, University of Cambridge*.

Mohr, M. and L. Selk (2020). Estimating change points in nonparametric time series regression models. *Statistical Papers 61*, 1437–1463.

Pesaran, M. and A. Pick (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics 29*(2), 307–318.

Pesaran, M. H., A. Pick, and M. Pranovich (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics 177*(2), 134–152.

Pesaran, M. H. and A. Timmermann (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics 129*(1-2), 183–217.

Pesaran, M. H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics 137*(1), 134–161.

Rossi, B. (2013). Advances in forecasting under instability, In G. Elliott, C. Granger and A. Timmermann (eds.). Volume 2 of *Handbook of Economic Forecasting*, pp. 1203–1324. Elsevier.

Schrimpf, A. and Q. Wang (2010). A reappraisal of the leading indicator properties of the yield curve under structural instability. *International Journal of Forecasting 26*(4), 836–857.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association 88*(422), 486–494.

Stock, J. H. and M. W. Watson (1989). *New Indexes of Coincident and Leading Economic Indicators*. NBER Macroeconomics Annual 1989, Volume 4. MIT Press.

Stock, J. H. and M. W. Watson (1999). Business cycle fluctuations in U.S. macroeconomic time series, In Taylor, R. and Woodford, M. (eds.). Volume 1 of *Handbook of Macroeconomics*, pp. 3–64. Elsevier.

Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

Sun, Y., Y. Hong, T.-H. Lee, S. Wang, and X. Zhang (2021). Time-varying model averaging. *Journal of Econometrics 222*(2), 974–992.

Timmermann, A. (2006). Forecast combinations, In G. Elliott, C. Granger and A. Timmermann (eds.). Volume 1 of *Handbook of Economic Forecasting*, pp. 135–196. Elsevier.

Appendix

We now give some regularity conditions that are sufficient for the consistency and asymptotic normality of the proposed estimators, although they might not be the weakest ones possible. As pointed out by Cai et al. (2000), the conditions list below are standard and they are satisfied for many applications; see, for instance, the paper by Cai et al. (2000) for details. Then, we present the sketch proofs of the asymptotic properties as mentioned in Section 2.3.

Condition A:

(A1) The second order derivatives of both mean functions $m_{(1)}(x)$ and $m_{(2)}(x)$ are continuously differentiable.

(A2) Function $f(x)$ is continuous and positive within the support.

(A3) The condition density of Y_t given X_t is bounded and satisfies the Lipschitz condition.

(A4) The kernel function $K(\cdot)$ is symmetric and has a compact support, say $[-1, 1]$.

(A5) The time series $\{Y_t, \boldsymbol{X}_t\}$ is α-mixing with the coefficient $\alpha(k)$ satisfying $\sum_{k=1}^{\infty} k^{c_0}\alpha^{1-2/\delta_0}(k)$ for some $\delta_0 > 2$ and $c > 1 - 2/\delta_0$.

(A6) Assume that $h \to 0$ and $T\,h \to \infty$.

Condition B:

(B1) Assume that

$$E\left[Y_t^2 + Y_{t+s}^2 \,|\, \boldsymbol{X}_t = \boldsymbol{x}_1, \boldsymbol{X}_{t+s} = \boldsymbol{x}_2\right] \leq M < \infty$$

for any t and all $s \geq 1$, and $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

(B2) Assume that there exists a sequence of positive integers $\{s_T\}$ such that $s_T \to \infty$, $s_T = o((Th)^{1/2})$ and $(T/s_T)^{1/2}\alpha(s_T) \to 0$, as $T \to \infty$.

(B3) There exists $\delta^* > \delta_0$, where δ_0 is given in Assumption A(5) such that $\alpha(k) = O(k^{-\theta})$, where $\theta > \delta_0\delta^*/[2(\delta^* - \delta_0)]$.

(B4) $T^{1/2-\delta_0/4}h^{\delta_0/\delta^*-1/2-\delta_0/4} = O(1)$.

Sketch of Theoretical Proofs

Proof of (8): To establish (8), first, we need to show that under Condition A,

$$S_j(x) \xrightarrow{p} \mu_j f(x). \tag{C.1}$$

Indeed, it is easy to show that $E[S_j(x)] \to \mu_j f(x)$ and $Th\text{Var}(S_j(x)) \to f(x)\nu_j$, by following the same idea as in the proof of Theorem 1 in Cai et al. (2000). Next, it is easy to see that in view of (C.1), the asymptotic bias term of $\hat{\beta}_{0,(1)}(x)$ can be asymptotically expressed as

$$\begin{aligned} B_1(x) &\approx \frac{1}{T_1}\sum_{t=1}^{T_1} K_h(X_t - x)\left\{m_{(1)}(X_t) - m_{(1)}(x) - m'_{(1)}(x)(X_t - x)\right\}/f(x) \\ &\approx \frac{m''_{(1)}(x)}{2}\frac{1}{T_1}\sum_{t=1}^{T_1} K_h(X_t - x)(X_t - x)^2/f(x) \\ &\approx m''_{(1)}(x)\mu_2 h^2/2 \end{aligned}$$

by Taylor expansion and following the same proof of (C.1). Similarly, $B_2(x)$, the asymptotic bias for $\hat{\beta}_{0,(2)}(x)$, can be obtained easily. Therefore, (8) is established.

Proof of (10): To establish (10), first, we show that $C_1 \xrightarrow{d} N\left(0, \sigma^2_{m,1}(x)\right)$ and $C_2 \xrightarrow{d} N\left(0, \sigma^2_{m,2}(x)\right)$. To this end, let $Z_t = K_h(X_t - x)u_t\sqrt{h/T_1}$. Then, $C_1 = \sum_{t=1}^{T_1} Z_t$. By following the same procedures as in the proof of Lemma A.1 in Cai et al. (2000), it is not difficult to show that under Conditions A and B, $\text{Var}(C_1) \to \sigma^2_{m,1}(x)$ as $T_1 \to \infty$. To establish the asymptotic normality of C_1, we employ the small-block and large-block technique — namely, $C_1 = Q_l + Q_s + Q_r$, to show that Q_l, the sum of the large-blocks converges a normal distribution in distribution, Q_s, the sum of the small-blocks, can be ignored in probability, Q_r, the sum of the remainder terms, converges to zero in probability, and the large-blocks are asymptotically independent. Also, we prove that for Q_l, the Lindeberg's condition is satisfied. Then, by the Lindeberg's central limit theorem, the asymptotic normal of C_1 is established. By the same token, we can establish the asymptotic normality for C_2. Finally, by following the same steps as used in proving Lemma A.1 in Cai et al. (2000), it is easy to show that $\text{Cov}(C_1, C_2) \to 0$ as $T \to \infty$. This completes the proof of (10).