Optimal forecasting under parameter instability[†,††,†††]

Yu Bai[a]

[a]*Monash University*

Abstract

This paper considers the problem of using local estimator in a forecasting model which is affected by parameter instability. We first show that local estimator is consistent under various types of parameter instability. Then, we analyse the choices of weighting function and tuning parameter associated with the local estimator. We prove the asymptotic optimality of the tuning parameter selection procedure and provide analytical criterion on the choice of weighting function. The theoretical results are examined through an extensive Monte Carlo study and four empirical applications on forecasting inflation, growth and inflation shocks, house price changes and bond returns.

Keywords: parameter instability, local estimator, tuning parameter, weighting function

JEL classification: C14, C51, C53

[†]Please click HERE for the latest version.

[††]This version: October 2023.

[†††]I am in debt to Shuping Shi for her continued support and guidance. I also thank Chaohua Dong, Bin Peng and Farshid Vahid for useful comments. Financial support from ARC Discovery Projects DP210100476 is greatly acknowledged.

1. Introduction

Many important economic decisions are based on a forecasting model that is known to be affected by parameter instability. It is now widely recognized that parameter instability is a crucial source of forecast failure. For instance, full sample parameter estimation might be inconsistent under parameter instability, results in poor out-of-sample (OOS) forecasting performance. The empirical evidence has also been well documented, see, for instance, equity premium forecasting (Welch and Goyal (2008)), volatility forecasting (Inoue et al. (2021)) and macroeconomic forecasting (Stock and Watson (1996)).

Motivated by concerns of parameter instability, forecasters often want to make predictions using the most recent data. They may do this by using a window of recent data, which is the so-called "rolling window" forecast scheme. As rolling window estimator is a special case of the local estimator when a flat weighting function is used (Inoue et al. (2017)), forecaster may have alternative choices of weighting functions and need to select the tuning parameter. This paper aims to address three issues associated with the local estimator in an out-of-sample forecasting context.

First, while the local estimator is quite popular in the applied work, it remains unclear what types of parameter instability are allowable to achieve consistency. We first show that under a general condition on the amount of time variation in model parameters, the local estimator is consistent. This covers a broad range of parameter instability considered in the literature, which include local structure break, smooth structural change (Robinson (1989), Cai (2007)) and realization of bounded persistent stochastic processes (Giraitis et al. (2014), Dendramis et al. (2021)). The consistency rate depends on the amount of local time variation and estimation becomes more precise when the amount of these variations are small.

The second and third issues are related to out-of-sample forecasting. We show that, from an end-of-sample risk reduction perspective, minimizing the end-of-sample risk is equivalent to minimize the regret risk (Hirano and Wright (2017)), which depends on the weighting function and tuning parameter associated with the local estimator. These are the two inputs forecaster has to choose. Tuning parameter determines the effective number of observations used in the local estimator and simplifies to window size when an indicator weighting function is used. We propose method to select the tuning parameter by directly minimizing the regret risk. The procedure is similar to the one proposed in Inoue et al. (2017) for rolling window selection, but we show that the asymptotic optimality holds when a generic weighting function is used for local estimation and a general loss function is used for forecast evaluation. The optimality does require stronger condition on parameter instability, but as we point out, this covers all cases considered in the literature.

Finally, we provide analyses on the choice of the weighting functions, which has been less addressed in the literature. Our analyses are based on the limiting behavior of the regret risk, which reflects the usual bias-variance trade-off. We show that, when estimation variance from the local estimator dominates, the regret risk converges in distribution to a zero-mean random variable. In this case, both the convergence rate and the term related to the weighting function do not depend on the types of parameter instability.

When estimation bias dominates, regret risk converges in probability to a non-zero constant. In this case, both the convergence rate and the term related to the weighting function are related to the property of parameter instability, making the choice more involved. However, this still provides some guidance on the implementation of tuning parameter selection procedure.

The theoretical analyses are examined through an extensive Monte Carlo study using a linear predictive regression model. We find that, the local estimator performs well under various types of parameter instability, but the estimation quality does depend on the amount of local time variation. In terms of forecasting performance, we find that our tuning parameter selection procedure works pretty well. In general, using all but downweighting the data is preferred.

We present four empirical applications on forecasting inflation, growth and inflation shocks, house prices and bond returns. These applications illustrate a variety of environments (target, validity of assumptions, forms of regret risk, estimation methods, loss functions). We would like to examine whether using local estimator with optimal tuning parameter selection improves forecast accuracy compared to well-known benchmark forecasts. In addition, we are more concerned on whether choice of weighting function matters in these applications.

Our empirical results are promising. Using local estimator with optimal tuning parameter selection procedure generally delivers gains. Choice of weighting function does have an impact on forecasting performance, but using a flat weighting function is generally outperformed by alternative weighting functions. We briefly summarize our main findings below.

(i) In the first application, we consider inflation forecasting for the United States (U.S.) and Canada. We find that a simple autoregressive distributed lag model performs quite well for Canada and could still achieve gains if certain predictor is used for U.S., such as the growth rate of civilian employment. In terms of weighting functions, we find that using all but downweighting the data is preferred.

(ii) In the second application, we consider the use of predictive quantile regression to forecast how specific features of the macroeconomic shock distributions respond to systemic risk. We find that TED spread (differences between 3-month LIBOR rate and 3-month treasury bill rate) is a useful predictor for left tail information about growth shocks, as well as inflation shocks for both left and right tail information. We again find that using all but downweighting the data is preferred.

(iii) We examine the international predictability of real house price changes in the third application. We find that, mixed frequency data sampling regression model (MIDAS) performs better particularly when the specification is based on valuation ratio. In terms of weighting functions, using only the recent data and downweighting them is preferred.

(iv) We consider international excess bond return predictability in the fourth application. We find that gains are quite substantial and significant in many cases compared to benchmark forecasts from

principal components of the global yield curve. We find that using only the recent data and down-weighting them is preferred.

The rest of the paper is organized as follows. Section 2 presents the setup and local estimation. Section 3 discusses selection of the tuning parameter and choice of weighting function from an end-of-sample risk reduction perspective. Section 4 provides Monte Carlo study on the theoretical analyses. Section 5 presents our empirical applications, and Section 5 concludes. Data descriptions are provided in the Appendix. Technical assumptions, auxiliary results lemmas and the proofs of the main theorems are provided in the supplementary material.

NOTATION: $\|\cdot\|$ is the Euclidean norm. $|\cdot|$ denotes the associated norm when $\cdot$ is one dimensional. $f^{(i)}(x) = \frac{d^i f(x)}{dx^i}$ denotes the ith derivative of function $f(\cdot)$ with respect to x. $x_n = O_p(y_n)$ states that the vector of random variables x_n is at most of order y_n in probability, and $x_n = o_p(y_n)$ is of smaller order than y_n in probability. $x_n \asymp y_n$ states that $x_n/y_n = O_p(1)$. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution. $E_T[\cdot] = E[\cdot|\mathcal{F}_T]$ is the conditional expectation operator, where $\mathcal{F}_T$ is the information set available at time T.

2. Estimation under parameter instability

Let $\{y_t\}_t$ be the scalar variable of interest and $\{X_t\}_t$ be a $s \times 1$ vector of predictors (which may include lags of y_t). We wish to forecast $y_{T+h} (1 \leqslant h < \infty)$, given the knowledge of X_T[1]. The forecast $\hat{y}_{T+h|T}$ is created using a rule: $\hat{y}_{T+h|T}(\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^{\bar{k}}$ is a $k \times 1$-dimensional model parameters. The model parameters are estimated via M-estimation minimizing

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \ell_t(\theta), \tag{1}$$

where $\ell_t(\theta) = L(y_{t+h}, \hat{y}_{t+h|t}(\theta))$ is some in-sample loss function.

Example 1. *Consider the linear predictive regression model:*

$$y_{t+h} = X_t'\theta + \varepsilon_{t+h}, \quad t = 1, 2, \cdots, T-h,$$

where $\{\varepsilon_{t+h}\}$ is a disturbance term. Then, OLS estimator is equivalent to (1) *when $\ell_t(\cdot)$ is the mean squared error loss: $\ell_t(\theta) = (y_{t+h} - X_t'\theta)^2$.*

It is well known that parameter instability plagues commonly used forecasting models and predictive content is unstable over time (Rossi (2013)). To handle the instability issues and remain agnostic on the

[1]We only consider direct forecast when $h > 1$.

types of parameter time variation, we assume that the time-varying parameters are modeled as the function of scaled time point $\tau_t = t/T \in (0, 1]$

$$\theta_t = \theta(\tau_t), \quad \theta(\cdot) : (0, 1] \longrightarrow \Theta. \tag{2}$$

As explained in Robinson (1989), the requirement that time-varying parameter is a function of scaled time point is essential to derive the consistency of the nonparametric estimator, since the amount of local information on which an estimator depends has to increase suitably with sample size T.

Since the forecasts and their evaluations are based on $\theta(1)$[2], we consider a local estimator for $\theta(1)$ defined by

$$\hat{\theta}_{K,b,T} = \arg\min_{\theta\in\Theta} \frac{1}{Tb} \sum_{t=1}^{T} k_{tT}\ell_t(\theta), \tag{3}$$

where $k_{tT} = K((t-T)/(Tb))$, $K(\cdot)$ is a weighting function, and $b = b_T > 0$ is a tuning parameter satisfying $b \to 0$, $Tb \to \infty$ as $T \to \infty$. Different specifications of $K(\cdot)$ lead to different types of forecasting schemes. If $k_{tT} = 1$ for all t, we are back to the non-local estimation as in (1). If $K(u) = \mathbb{1}_{\{-1<u<0\}}$, we are in the rolling forecast scheme with window size $\lfloor Tb \rfloor$ (Giacomini and Rossi (2009)).

Local estimator like (3) is widely used in the out-of-sample forecasting context under parameter instability. Typically, it (mainly rolling window estimator) is used under the case when parameters are assumed to have break points, possibly at unknown dates (Pesaran and Timmermann (2007)). In this case, (2) is a piecewise continuous function on $(0, 1]$. It can also be used when (2) is twice continuously differentiable on $(0, 1]$ (Inoue et al. (2017)). What are the minimum requirements on (2) to achieve consistency of (3)? For the ℓth elements in (2), consider the condition

$$|\theta_\ell(t/T) - \theta_\ell(s/T)| \leqslant c_\ell \Big(\frac{|t-s|}{T}\Big)^\gamma, \quad t, s = 1, 2, \cdots, T, \tag{4}$$

for some $0 < \gamma \leqslant 1$ and c_ℓ is a positive bounded constant. This is similar to a Hölder continuous condition. The amount of local time variation vanishes asymptotically as $T \to \infty$. Chen and Hong (2016) derives the consistency of estimator like (3) for GARCH models, but consistency rate is not provided. The case when $\theta(t/T)$ is twice continuous differentiable on $(0, 1]$ (Robinson (1989), Cai (2007)) also satisfies (4). Giraitis et al. (2014) show that (3) can handle realization of persistent bounded stochastic processes. For example, realization of bounded random walk process $\overline{\theta}_{\ell,t} = \frac{1}{\sqrt{T}} v_t$, where $\Delta v_t \overset{i.i.d.}{\sim} (0, 1)$ satisfies (4) with $\gamma = 1/2$[3]. Similar condition is also used in Li and Müller (2009) for unstable generalized method of

[2]The target y_{T+1} depends on parameter $\theta(1 + 1/T)$, which is different from $\theta(1)$. However, under Assumption A4(i), the local time variation is asymptotically negligible, we shall treat $\theta(1)$ as the parameter related to the target.

[3]As shown in Dendramis et al. (2021), such process satisfies the condition $\big|\overline{\theta}_{\ell,t} - \overline{\theta}_{\ell,s}\big| \leqslant \xi_{\ell,ts} \Big(\frac{|t-s|}{T}\Big)^\gamma$, where $\xi_{\ell,ts}$ has a thin-tailed distribution: $\mathbb{P}\Big(\big|\xi_{\ell,ts}\big| > \omega\Big) \leqslant \exp\Big(-c_0|\omega|^\alpha\Big)$, $\omega > 0$, for some $c_0 > 0$, $\alpha > 0$, which does not depend on ℓ, t, s and T. Then,

moments models. They consider both the case of realization of bounded random walk process and local one-time break of the form: $\theta_{\ell,t} = a_T \mathbb{1}_{\{t/T>e\}}$, where $e \in (0,1]$ and $a_T = o(1)$ as $T \to \infty$.

In Lemma B1, we show that, under (4) and other regularity conditions, $\hat{\theta}_{K,b,T}$ is consistent: $\hat{\theta}_{K,b,T} \xrightarrow{p} \theta(1)$. In addition,

$$\left\|\hat{\theta}_{K,b,T} - \theta(1)\right\| = O_p((Tb)^{-1/2} + b^{\gamma}),$$

where $b \to 0$, $Tb \to \infty$ as $T \to \infty$. The consistence rate is inversely related to γ. Relatively large γ implies that local changes are small, the estimation bias vanishes at a faster rate. When γ is too small (large local changes), consistency rate gets distorted.

3. Out-of-sample forecasting

To implement the local estimator (3), a forecaster faces a concrete decision problem as she has to choose weighting function K and tuning parameter b. In order to understand the implications of selecting K and b, we will analyze the end-of-sample risk $E_T(\ell_{T+1}(\hat{\theta}_{K,b,T}))$. Under Assumption A1(i), a second-order Taylor series expansion around the true $\theta(1)$ gives (ignoring the smaller order terms)[4]:

$$\ell_{T+1}(\hat{\theta}_{K,b,T}) \approx \ell_{T+1}(\theta(1)) + \frac{\partial \ell_{T+1}(\theta(1))}{\partial \theta'}(\hat{\theta}_{K,b,T} - \theta(1)) + \frac{1}{2}(\hat{\theta}_{K,b,T} - \theta(1))' \frac{\partial^2 \ell_{T+1}(\overline{\theta}(1))}{\partial \theta \partial \theta'}(\hat{\theta}_{K,b,T} - \theta(1)), \quad (5)$$

where $\overline{\theta}(1)$ lies between $\hat{\theta}_{K,b,T}$ and $\theta(1)$. Taking conditional expectations on both sides we then find

$$E_T(\ell_{T+1}(\hat{\theta}_{K,b,T})) \approx \underbrace{E_T(\ell_{T+1}(\theta(1)))}_{R_T^1} + \underbrace{E_T\Big(\frac{\partial \ell_{T+1}(\theta(1))}{\partial \theta'}\Big)(\hat{\theta}_{K,b,T} - \theta(1))}_{R_T^2}$$
$$+ \frac{1}{2}(\hat{\theta}_{K,b,T} - \theta(1))' E_T\Big(\frac{\partial^2 \ell_{T+1}(\overline{\theta}(1))}{\partial \theta \partial \theta'}\Big)(\hat{\theta}_{K,b,T} - \theta(1)). \quad (6)$$

We see that the end-of-sample risk can be decomposed into three components. The component R_T^1 is related to the future risk, which has nothing to do with parameter estimation. Following Hirano and Wright (2017), we define the regret risk when using the parameter estimates $\hat{\theta}_{K,b,T}$:

$$R_T(K,b) = E_T\Big(\frac{\partial \ell_{T+1}(\theta(1))}{\partial \theta'}\Big)(\hat{\theta}_{K,b,T} - \theta(1)) + \frac{1}{2}(\hat{\theta}_{K,b,T} - \theta(1))' E_T\Big(\frac{\partial^2 \ell_{T+1}(\overline{\theta}(1))}{\partial \theta \partial \theta'}\Big)(\hat{\theta}_{K,b,T} - \theta(1)). \quad (7)$$

we could always find a generic c_ℓ such that realization of the process satisfies the condition (4).

[4]Following Granger (1969) and Weiss (1996), we use the same loss function for parameter estimation and out-of-sample forecasting (OOS) evaluation. However, in some applications, there may be gains from using an alternative loss function for estimation, see Hansen and Dumitrescu (2022).

Under Assumption A1(ii)[5], the regret risk defined in (7) simplifies to (ignoring the constant 1/2):

$$R_T(K,b) = (\hat{\theta}_{K,b,T} - \theta(1))' \, E_T\Big(\frac{\partial^2 \ell_{T+1}(\overline{\theta}(1))}{\partial\theta\partial\theta'}\Big) (\hat{\theta}_{K,\bar{b},T} - \theta(1)). \tag{8}$$

Thus, minimizing the end-of-sample risk is equivalent to minimize $R_T(K,b)$.

3.1. Selection of the tuning parameter b

Since (3) is a nonparametric estimator, it is well known that the tuning parameter b is essential in risk reduction (trade-off between reducing bias and variance). We now introduce our tuning parameter selection procedure. Suppose that the weighting function K is chosen. Write $\hat{\theta}_{\overline{K},b,T} = \hat{\theta}_{b,T}$ and $\omega_T(\overline{\theta}(1)) = \frac{\partial^2 \ell_{T+1}(\overline{\theta}(1))}{\partial\theta\partial\theta'}$. We consider to choose b by simply minimizing the regret risk (8) over the choice set I_T:

$$\hat{b} := \underset{b\in I_T}{\arg\min} \, (\hat{\theta}_{b,T} - \theta(1))' \, \omega_T(\overline{\theta}(1)) \, (\hat{\theta}_{b,T} - \theta(1)). \tag{9}$$

Notice that, the cardinality of the set $|I_T|$ must shrink to zero as $T \to \infty$, since the consistency of $\hat{\theta}_{b,T}$ requires $b \to 0$.

We first drive the rate of the optimal tuning parameter implied by (9), which is characterised in the following theorem:

Theorem 1. *Under Assumptions A1, A2, A3(i) and A4, the optimal tuning parameter $\hat{b}$ obtained by minimizing* (9) *is of order $T^{-\frac{1}{2\gamma+1}}$ in probability for some $0 < \gamma \leqslant 1$.*

Theorem 1 shows that, the optimal tuning parameter $\hat{b}$ should be equal to $cT^{-\frac{1}{2\gamma+1}}$, as T goes to infinite, for some finite constant $0 < c < \infty$. This implies that the effective number of observations $\lfloor Tb \rfloor$ is inversely related to γ: when the local time variation in $\theta(t/T)$ is large, the effective number of observations $\lfloor Tb \rfloor$ should also be lower.

Since $\hat{\theta}_{b,T}$ is consistent (Lemma B1), we have $\overline{\theta}(1) \xrightarrow{p} \theta(1)$. Following Inoue et al. (2017), we consider to use the local linear estimator to approximate the unknown $\theta(1)$. The local linear estimate proceeds as follows. Under Assumption A3(ii), $\theta(t/T)$ is twice continuously differentiable, a second-order Taylor expansion of $\theta(t/T)$ around 1 gives

$$\theta(t/T) \approx \beta_1 + \beta_2 \, (\frac{t-T}{T}) + \frac{\beta_3}{2} \, (\frac{t-T}{T})^2, \tag{10}$$

where $\beta_1 = \theta(1), \beta_2 = \theta^{(1)}(1)$ and $\beta_3 = \theta^{(2)}(c)$, where c lies between 1 and t/T. Let $D(u) = [1, u]$, $u = \frac{t}{T} - 1$

[5]We assume that the score of the loss $\left\{\frac{\partial \ell_{t+1}(\theta(t/T))}{\partial\theta'}\right\}_t$ is a martingale difference sequence (M.D.S.). For the model considered in Example 1, this implies that the error term is M.D.S, so a sufficient number of lags of y_t have to be included in the model to clear the serial correlation in the error term.

and $\beta = (\beta_1', \beta_2')'$. The local linear estimator is defined by the minimizer of

$$\min_{\beta} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \ell_t\Big(D(u)\beta\Big), \tag{11}$$

where the weights $\tilde{k}_{tT} = \tilde{K}\left(\frac{t-T}{T\tilde{b}}\right)$ are computed with bandwidth parameter $\tilde{b}$ such that $\tilde{b} \to 0$ and $T\tilde{b} \to \infty$ as $T \to \infty$.

Let $\tilde{\theta}_T$ be the collection of the first $k \times 1$ elements of the minimizer of (11), we then replace the unknowns in (9) with the local linear estimator $\tilde{\theta}_T$, which leads to a feasible selection criteria:

$$\hat{b} := \underset{b \in I_T}{\arg\min} \, (\hat{\theta}_{b,T} - \tilde{\theta}_T)' \, \omega_T(\tilde{\theta}_T) \, (\hat{\theta}_{b,T} - \tilde{\theta}_T). \tag{12}$$

The asymptotic optimality of the feasible selection procedure (12) is formally stated in the next theorem.

Theorem 2. *Under Assumptions A1-A5, choosing* $\hat{b}$ *by* (12) *is asymptotically optimal in the sense that*

$$(\hat{\theta}_{b,T} - \tilde{\theta}_T)' \, \omega_T(\tilde{\theta}_T) \, (\hat{\theta}_{b,T} - \tilde{\theta}_T) \asymp \inf_{b \in I_T} (\hat{\theta}_{b,T} - \theta(1))' \, \omega_T(\theta(1)) \, (\hat{\theta}_{b,T} - \theta(1))$$

where $\tilde{\theta}_T$ *is the local linear estimator from* (11) *with bandwidth* $\tilde{b}$*.*

Theorem 2 provides an extension to the ones in Inoue et al. (2017) by showing that the asymptotic optimality holds for a generic weighting function when using (3) and a general loss function for forecast evaluation. The asymptotic optimality implies that $\hat{b}$ chosen from (12) yields the same forecasts obtained from the true optimal tuning parameter by minimizing the infeasible objective function in (9). The key to establish this result is to use the fact that the asymptotic bias from local linear estimator vanishes at a faster rate than local estimator in (3). As in Assumption A5, the requirements for two tuning parameters involved: b and $\tilde{b}$ are rather intuitive. We could first let $T\tilde{b}^5 \to 0$ to obtain the best possible convergence rate for $\tilde{\theta}_T$. Then, the remaining conditions hold when b goes to zero at a faster rate than $\tilde{b}$.

Remark 1. *The condition for* $\theta(t/T)$ *imposed on Theorem 2 is stronger than Theorem 1 since it requires that* $\theta(t/T)$ *is twice continuously differentiable. However, this condition is not that restrictive as it covers particular the ones considered in Giraitis et al. (2014) and Dendramis et al. (2021), where* (3) *is used to estimate a path of the stochastic time-varying coefficient. To see this, suppose that* $\theta(t/T)$ *is a realization of a bounded random walk process:* $\frac{1}{\sqrt{T}}\xi_t$*, where* $\Delta\xi_t = v_t \overset{i.i.d.}{\sim} (0, 1)$*. Simple algebra gives* $\theta(t/T) = \sqrt{\frac{t}{T}} \frac{1}{\sqrt{t}}\xi_t$*. We know that* $\frac{1}{\sqrt{t}}\xi_t = O_p(1)$*, this implies that* $\theta(t/T) = C_t \sqrt{\frac{t}{T}}$*, where* C_t *is a positive bounded constant.*

3.2. Implications on the choice of K

Another input forecaster has to choose is the weighting function K. Consider the following three candidate choices of $K(u)$:

$$K_1(u) = \mathbb{1}_{\{-1<u<0\}}, \quad K_2(u) = \frac{2}{\sqrt{2\pi}} \exp\big(-\frac{u^2}{2}\big)\mathbb{1}_{\{u<0\}}, \quad K_3(u) = \frac{3}{2}(1-u^2)\mathbb{1}_{\{-1<u<0\}}. \tag{13}$$

$K_1(u)$ leads to a rolling window estimator with window size $\lfloor Tb \rfloor$. $K_2(u)$ imposes an exponential-type downweighting scheme and $K_3(u)$ implies a hyperbolic type downweighting scheme. Although $K_1(u)$ used to dominate in the applied work, there has been a growing interest in using other weighting functions. For instance, $K_2(u)$ has been used in macroeconomic forecasting context (Kapetanios et al. (2019) and Dendramis et al. (2020)). $K_3(u)$ is recommended in equity premium forecasts as in Farmer et al. (2022). A graphical illustration of the weighting functions in (13) is provided in Figure 1. For $K_1(u)$ and $K_3(u)$, the tuning parameter b determines the number of observation used in (3). For $K_2(u)$, b determines how fast the weights decay.

As shown in Lemma B1, (3) obeys the following expansion:

$$\hat{\theta}_{K,b,T} - \theta(1) = -H_{1,T}^{-1} S_{1,T} - H_{1,T}^{-1} S_{2,T},$$

where

$$H_{1,T} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}, \quad S_{1,T} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT} \frac{\partial \ell_t(\theta(t/T))}{\partial\theta}, \quad S_{2,T} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\bar{\theta}(1))}{\partial\theta\partial\theta'}\,(\theta(1)-\theta(t/T)),$$

and $\bar{\theta}(1)$ lies between $\hat{\theta}_{K,b,T}$ and $\theta(1)$. Since $\|H_{1,T}\| = O_p(1)$, $\|S_{1,T}\| = O_p((Tb)^{-1/2})$, $\|S_{2,T}\| = O_p(b^{\gamma})$, the limiting distribution of the estimator depends on the relative order of $S_{1,T}$ and $S_{2,T}$, which affects the regret risk $R_T(K,b)$. The limiting behavior of $R_T(K,b)$ is summarized in the next Theorem.

Theorem 3. *Suppose that Assumptions A1, A2, A3(i), A4 hold with $b \to 0$ and $Tb \to \infty$. Then, it holds that*

(i) If $T^{1/2}b^{1/2+\gamma} \to 0$, we have

$$Tb \cdot R_T(K,b) \xrightarrow{d} \phi_{0,K}\Sigma^{1/2}(1)Z'\omega_T(\theta(1))Z\Sigma^{1/2}(1),$$

where $\phi_{0,K} = \int_C K^2(u)du$, $Z \sim \mathcal{N}(0, I_k)$ and $\Sigma(1)$ is defined as in Lemma B1;

(ii) If $T^{1/2}b^{1/2+\gamma} \to \infty$, we have

$$b^{-2\gamma} \cdot R_T(K,b) \xrightarrow{p} \mu_{\gamma,K}^2 C'\omega_T(\theta(1))C,$$

where $\mu_{\gamma,K} = \int u^\gamma K(u)du$ *and* C *is a collection of Hölder constant given in Assumption A3(i);*

(iii) If $T^{1/2}b^{1/2} \asymp b^{-\gamma}$*, we have*

$$Tb \cdot \left(R_T(K,b) + b^{2\gamma}\mu_{\gamma,K}^2 C'\omega_T(\theta(1))C\right) \xrightarrow{d} \phi_{0,K}\Sigma^{1/2}(1)Z'\omega_T(\theta(1))Z\Sigma^{1/2}(1),$$

where $\mu_{\gamma,K}$*,* C *and* $\phi_{0,K}$ *are defined as in (i) and (ii).*

The limiting risk reflects the usual bias-variance trade-off. Consider first when $T^{1/2}b^{1/2+\gamma} \to 0$. In this case, the bias introduced by (3) vanishes asymptotically. The rescaled regret risk converges in distribution to a random variable centered at 0[6]. As estimation variance dominates and $\phi_{0,K}$ affects it, we clearly prefer a weighting function which has smallest $\phi_{0,K}$. For the weighting functions considered in (13), we have $\phi_{0,K_1} = 1$, $\phi_{0,K_2} \approx 0.5642$ and $\phi_{0,K_3} = 1.2$. In this case, there is a clear winner. $K_2(u)$ is preferred: all data should be used and downweighted.

If $T^{1/2}b^{1/2+\gamma} \to \infty$, estimation bias from (3) dominates and regret risk converges to a non-stochastic term which is not 0. Since we do not know γ, b may be set improperly so we are in case (ii) described in Theorem 3. In this case, $\mu_{\gamma,K} = \int u^\gamma K(u)du$ plays a role so we clearly want to choose a weighting function which has smallest $\mu_{\gamma,K}$. Consider the case when $\gamma = 1$. We have $\mu_{1,K_1}^2 = 0.25$, $\mu_{1,K_2}^2 \approx 0.637$ and $\mu_{1,K_3}^2 = 0.141$. Then, we may expect that, if the bias term dominates, $K_3(u)$ would be preferred: only recent data should be used and downweighted. In the third case when $(Tb)^{1/2}$ and $b^{-\gamma}$ diverge at the same rate, the rescaled regret risk converges in distribution, but estimation bias is still present. This implies that both $\phi_{0,K}$ and $\mu_{\gamma,K}$ matter. However, as we do not know γ, the criterion is more involved in Cases (ii) and (iii) since it depends on the properties of parameter time variation.

Remark 2. *Using a similar expansion as in* (8)*, Oh and Patton (2021) analyze the OOS forecast accuracy from both the local and non-local estimator. They argue that,* (8) *is dominated by the estimation error in the local estimator, and b determines the usual bias-variance tradeoff. Yet, the analysis above show that under certain condition on b,* (8) *is asymptotically dominated by the estimation variance in the local estimator.*

3.3. Treatment of γ

What remains is the issue that γ is unknown so Theorem 1 is still not applicable. In both Monte Carlo study and empirical applications, we fix $\gamma = 1$ and thus $b = cT^{-1/3}$. c is selected by minimizing the regret risk (12). This implies that we may fall into case (ii) in Theorem 3, but the regret risk converges to a non-zero constant at the slowest possible rate. As effective number of observations are of order $T^{2/3}$, a relatively larger $\lfloor Tb \rfloor$ would also be useful to reduce the finite sample variability of the regret risk.

[6]Notice that when $\omega_T(\theta(1))$ is idempotent matrix, the asymptotic distribution has a more elegant expression, since $Z'\omega_T(\theta(1))Z \sim \chi^2$, where the degree of freedom is given by trace$(\omega_T(\theta(1)))$.

4. Monte Carlo experiments

We now turn to the Monte Carlo experiments of our analysis in section 2. The DGPs are based on a bivariate VAR(1) as in Pesaran and Timmermann (2007) and Inoue et al. (2017):

$$\begin{bmatrix} y_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} a_t & b_t \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} + \begin{bmatrix} \varepsilon^y_{t+1} \\ \varepsilon^x_{t+1} \end{bmatrix}, \tag{14}$$

where the error terms $(\varepsilon^y_{t+1}, \varepsilon^x_{t+1})'$ are generated from $\mathcal{N}(0, I_2)$.

In DGP 1, the parameters are constant over time: $a_t = 0.9$ and $b_t = 1$ for all t. In DGPs 2-4, we have a one-time local break in these two parameters: $a_t = 0.9 - \frac{1}{T^{0.2}} \mathbb{1}(t \geqslant \pi T + 1)$, $b_t = 1 + \frac{1}{T^{0.2}} \mathbb{1}(t \geqslant \pi T + 1)$, where $\pi = 0.25, 0.5, 0.75$, respectively. DGPs 5-12 use the smooth time-varying parameters (Assumption A4(ii) is satisfied). In DGP 5, we set $a_t = 0.9 - 0.4(t/T)$ and $b_t = 1 + (t/T)$. In DGP 6, we set $a_t = 0.9 - 0.4(t/T)^2$ and $b_t = 1 + (t/T)^2$. In DGP 7, we set $a_t = 0.9 - 0.4 \exp(-3.5t/T)$ and $b_t = 1 + \exp(-16(t/T - 0.5)^2)$. In DGPs 8-12, we consider various degree of smoothness in time-varying parameters. We first generate $v_{it} = (1 - L)^{1-d} \epsilon_{it}$, where $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1/(100^2))$. Then, we generate ξ_{it} from the random walk model: $\Delta \xi_{it} = v_{it}$. Finally, we set $a_t = 0.95 \frac{\xi_{1t}}{\max_{1 \leqslant j \leqslant t} |\xi_{1j}|}$ and $b_t = \xi_{2t} / \sqrt{T}$. We consider $d = 0.51, 0.75, 1, 1.25, 1.49$ for DGPs 7-11 respectively, which correspond to the setting where $\gamma = 0.01, 0.25, 0.5, 0.75, 0.99$.

We consider the following predictive regression model:

$$y_{t+1} = X_t' \theta_t + \varepsilon_{t+1},$$

where $X_t = (y_t, x_t)'$ and $\theta_t = (a_t, b_t)'$. The model parameters are estimated by the local least square (LS):

$$\hat{\theta}_{K,b,T} = \Big(\sum_{t=1}^{T-1} k_{tT} X_t X_t' \Big)^{-1} \Big(\sum_{t=1}^{T-1} k_{tT} X_t y_{t+1} \Big),$$

where the weights $k_{tT} = K\Big(\frac{t-T}{Tb}\Big)$ are computed from a weighting function $K(u)$ with tuning parameter b. We consider three different choices of weighting functions as discussed in section 2.2:

$$K_1(u) = \mathbb{1}_{\{-1<u<0\}}, \quad K_2(u) = \frac{2}{\sqrt{2\pi}} \exp\big(-\frac{u^2}{2}\big) \mathbb{1}_{\{u<0\}}, \quad K_3(u) = \frac{3}{2}(1 - u^2) \mathbb{1}_{\{-1<u<0\}}. \tag{15}$$

Of course, when $k_{tT} = 1$ for all t, we are back to the non-local full sample LS estimates.

The forecasts are evaluated by the mean squared forecast error (MSFE) loss, and the regret risk becomes

$$R(K, b) = (\hat{\theta}_{K,b,T} - \theta(1))' \Big(X_T X_T' \Big) (\hat{\theta}_{K,b,T} - \theta(1)). \tag{16}$$

Notice that, in this case, Assumption A1(ii) implies that the forecast error is a M.D.S.: $E[\varepsilon_{t+1} | \mathcal{F}_t] = 0$. The true parameters $\theta(1)$ in $R(K, b)$ are approximated by the local linear estimator, which are the first $k \times 1$

elements of the following $(\tilde{\theta}_T)$:

$$\hat{\beta}_T = \Big(\sum_{t=1}^{T-1} \tilde{k}_{tT} Z_t Z_t'\Big)^{-1} \Big(\sum_{t=1}^{T-1} \tilde{k}_{tT} Z_t y_{t+1}\Big),$$

where $Z_t = [X_t', X_t'(\frac{t-T}{T})]'$, $\tilde{k}_{tT} = K\Big(\frac{t-T}{T\tilde{b}}\Big)$ are computed from a weighting function $K(u)$ with tuning parameter $\tilde{b}$. We use the same weighting function used for $\hat{\theta}_{K,b,T}$ to compute $\tilde{\theta}_T$ and $\tilde{b}$ is selected by the rule-of-thumb method: $\tilde{b} = 1.06T^{-1/5}$. For b, we set $b = cT^{-1/3}$ and select c by minimizing $R(K, b)$ using a course grid of width 0.05 from 1 to 5.

We first evaluate the performance of the local LS estimator under various types of time variation as in DGPs 2-12. The number of Monte Carlo simulations is set to 5,000 and four different sample sizes are considered $T = 100, 200, 300, 400$. Table 1 reports the average of mean absolute deviations (MADs): $\frac{1}{M}\sum_{m=1}^{M}\big|\hat{\theta}^m_{K,b,T} - \theta^m(1)\big|$ for both $a(1)$ (upper panel) and $b(1)$ (bottom panel). For each DGP, we consider all the weighting functions in (15) and the tuning parameter b is selected by minimizing (16). Overall, the performances are quite satisfactory as MADs decrease as sample size increases. It is clear from DGPs 8-12 that MADs also decrease as γ increases, indicating that estimation gets harder when changes are more frequent. Finally, in terms of choices of weighting function, we see that using exponential type weighting function $(K_2(u))$ is generally preferred for $b(1)$, but in terms of $a(1)$ parabolic type weighting function $(K_3(u))$ is better for DGPs 4 and 9-12.

We then evaluate the performance of the out-of-sample prediction of y_{T+1} over 5000 Monte Carlo simulations. The benchmark forecasts are generated from $\hat{\theta}_{T+1|T} = X_T' \hat{theta}_T$, where $\hat{\theta}_T$ is obtained from non-local LS estimator (when all the weights are equal to 1: $k_{tT} = 1$). For comparison, we also consider rolling window forecasts with rolling window size equal to 40. Table 2 report the ratios of the RMSFEs (square root of MSFEs) relative to the benchmark forecasts:

$$\frac{\sqrt{\sum_{m=1}^{M}(y^m_{T+1} - \hat{y}^m_{T+1|T})^2}}{\sqrt{\sum_{m=1}^{M}(y^m_{T+1} - \tilde{y}^m_{T+1|T})^2}},$$

where $\tilde{y}^m_{T+1|T}$ is the benchmark forecast and $\hat{y}^m_{T+1|T}$ are the forecasts from using local estimators. If the ratio of RMSFEs is less than 1, the forecasts generated from local estimator is more accurate than the ones from non-local estimator. Entries shaded in gray indicate the best performing model.

Table 2 summarizes the results. First, all of the forecasts generated from local estimator are more accurate than the ones from non-local estimator when there is time variation in model parameters. One exception is DGP 8 when $T = 100$. In this case, the time variation is too rough and estimation may not be precise when sample size is small. Of course, when the parameters are constant over time (DGP 1), the non-local estimator is more efficient and forecasts are more accurate. Interesting, for DGP4, when

the break occurs close to the time when forecasts are made, either using fixed rolling forecasts or optimal selection with $K_3(u)$ is the best, indicating that not using all data is useful. This is also true for DGP3 when sample size is small ($T = 100$). When the time variation is a quadratic function of t/T, using $K_1(u)$ is always the best. For all the other cases (34 out of 44), using exponential type weighting function ($K_2(u)$) is better than others, which implies that it is more likely that using all data delivers more benefits.

5. Empirical applications

We present four empirical applications. First, we consider forecasting inflation in the United States (U.S.) and Canada. Unobserved component model with stochastic volatility (UCSV) originally proposed by Stock and Watson (2007) is a well-known benchmark for forecasting U.S. inflation. We would like to investigate whether local estimation of a simple distributed lag model would lead to accuracy gains compared to this benchmark. In addition, as the UCSV model is proposed for U.S. data, we also want to examine whether same conclusion holds for Canada.

Our second application considers forecasting macroeconomic shocks using systematic risk measures. Recent Covid-19 crisis has shifted interests to the tails of macroeconomic variables rather than the means (Carriero et al. (2022) and Clark et al. (2023)). We use a predictive quantile regression model to forecast how specific features of the macroeconomic shock distribution respond to systemic risk. Compared to Giglio et al. (2016), we investigate not only the left quantiles of industrial production shock, but also inflation shocks in both left and right quantiles. In addition, we also examine whether using local estimator could improve tail forecast accuracy.

Our third application focuses on forecasting real house price changes for five industrialized countries: U.S., Canada, France, Germany and Australia. Ghysels et al. (2013) use several model specifications to examine in-sample predictability of real estate price for U.S., but out-of-sample (OOS) forecasting is only conducted for real estate investment trusts (REITs). From a macroeconomic perspective, residential real estate sector is a key driver of economic growth and plays a dominant role in business cycle variation (Leamer (2007)). Thus, it is of great interests to investigate the OOS forecasting performance for real house price changes. In addition, estimation method is also different from the previous two applications, as nonlinear models are used.

We consider bond return predictability in our fourth application. Treasury bonds play an important role in many investors' portfolios. Recent literature (Gargano et al. (2019) and Borup et al. (2023)) has documented evidence of time variation in OOS bond return predictability for one-month (nonoverlapping) return data. We would like to examine there is time variation in 12-month overlapping return predictability. Overlapping return data have been used in many existing studies such as Cochrane and Piazzesi (2005) and Liu and Wu (2021). In addition, as explained in Bauer and Hamilton (2018), overlapping data may introduce temporal dependence in the forecasting errors. Assumption A1(iii) may fail and we would like to see how our methods perform in this case.

Across these applications, for the local estimation method, we consider three different choices of weighting functions as in (15):

$$K_1(u) = \mathbb{1}_{\{-1<u<0\}}, \quad K_2(u) = \frac{2}{\sqrt{2\pi}} \exp\big(-\frac{u^2}{2}\big)\mathbb{1}_{\{u<0\}}, \quad K_3(u) = \frac{3}{2}(1-u^2)\mathbb{1}_{\{-1<u<0\}}.$$

It is again worth mentioning that all data are used when $K_2(u)$ is used, but certain data points are discarded when either $K_1(u)$ and $K_3(u)$ is used. The tuning parameter b controls the rate of how fast the weights decay for $K_2(u)$. For $K_1(u)$ and $K_3(u)$, b determines how many observations are used for local estimation.

Except for the second application, the forecast evaluations are all based on the MSE loss function. Then, the regret loss becomes

$$R(K,b) = (\hat{\theta}_{K,b,T} - \theta(1))' \Big(X_T(\theta(1))X_T'(\theta(1))\Big)(\hat{\theta}_{K,b,T} - \theta(1)). \tag{17}$$

Note that the weighting matrix $X_T(\delta_2(1))X_T'(\delta_2(1))$ may depend on the parameters. Since quantile regression model is used in the second application, the forecast evaluations are based on the check loss function $\ell_t^\tau(\theta) = \varepsilon_{t+1}(\theta)(\tau - \mathbb{1}_{\{\varepsilon_{t+1}(\theta)<0\}})$, where $\varepsilon_{t+1}(\theta) = y_{t+1} - X_t'\theta_\tau$ and τ is the quantile of interest. Then, the regret loss becomes

$$R(K,b) = (\hat{\theta}_{K,b,T} - \theta(1))' E_T\Big(\frac{\partial \ell_{T+1}^2(\theta)}{\partial\theta\partial\theta'}\Big)(\hat{\theta}_{K,b,T} - \theta(1)). \tag{18}$$

As the check loss function is not differentiable, we consider the smoothed quantile regression framework in Fernandes et al. (2021). We use the in-sample counterpart to approximate $E_T\Big(\frac{\partial \ell_{t+1}^2(\theta)}{\partial\theta\partial\theta'}\Big)$. More details are provided in Section 5.2.

Once the weighting function is chosen, we set $b = cT^{-1/3}$ and use $\hat{c}$ to obtain our final forecast by minimizing $R(K,b)$ as defined in (17). When data are sampled quarterly (Sec. 5.1 and Sec. 5.3), we let c range from 1 to 5 with a course grid of width 0.05. When data is are sampled monthly (Sec. 5.2 and Sec. 5.4), we let c range from 1 to 7 with a course grid of width 0.1. We replace the unknown $\theta(1)$ with the corresponding local linear estimator as in (11). We use the same weighting function for local linear estimator and use a rule-of-thumb method for $\tilde{b}$ by setting $\tilde{b} = 1.06T^{-1/5}$.

Apart from the local estimator using weighting functions in (15) with optimal tuning parameter selection, we also consider two alternative estimators. The first one is the non-local estimator when all the weights in (3) are set to 1, so the parameter instability is completely ignored. The second one is the standard rolling window estimator with fixed window size. Following Stock and Watson (2003), when data is sampled quarterly, the window size is set to 40 (10 years of quarterly data). When data is monthly, as common in finance applications (Farmer et al. (2022)), the window size is set to 60 (5 years of monthly data).

The overall forecasting performance are presented in Tables 3-11. The benchmark forecasts are different across applications and are stated in the following subsections. The entries related to the benchmark

forecasts are the losses in levels. For all other entries, they are the ratios of losses relative to the benchmark forecasts. Values below 1 indicate that the corresponding specification performs better than the benchmark. Entries shaded in light cran are the ones perform better than the benchmark in each specification and gray shaded entries are the best performing specification.

Finally, to provide a rough gauge of whether differences in accuracy are significantly different, we apply the Diebold and Mariano (1995) (DM) test for equal forecast accuracy with fixed smoothing asymptotics as in Coroneo and Iacone (2020) , which is shown to deliver predictive accuracy tests that are correctly sized even when the number of out-of-sample observations are small.

5.1. Forecasting inflation

The target variable is the annualized inflation rate: $y_t = 400 \ln(Q_t/Q_{t-1})$, where Q_t is the implicit price deflator of the gross domestic product (GDP). The benchmark forecasts are obtained from the UCSV model:

$$
\begin{aligned}
y_t &= \tau_t + \varepsilon_t^y, \quad \varepsilon_t^y \sim N(0, e^{h_t}), \\
\tau_t &= \tau_{t-1} + \varepsilon_t^\tau, \quad \varepsilon_t^\tau \sim N(0, e^{g_t}), \\
h_t &= h_{t-1} + \varepsilon_t^h, \quad \varepsilon_t^h \sim N(0, \omega_h^2), \\
g_t &= g_{t-1} + \varepsilon_t^g, \quad \varepsilon_t^g \sim N(0, \omega_g^2),
\end{aligned}
$$

with initial conditions τ_0, h_0 and g_0 as unknown parameters. We assume Normal priors for all model parameters: $\omega_h \sim N(0, 0.2^2)$, $\omega_g \sim N(0, 0.2^2)$, $h_0 \sim N(0, 10)$, $g_0 \sim N(0, 10)$, and $\tau_0 \sim N(0, 10)$. The model is estimated using Bayesian methods in non-centered parameterization and then transform back to the centered parameterization to perform predictive simulation. Estimation details can be found in Appendix B in Chan (2018).

Alternatives forecasts are obtained from ARDL(p, q) model:

$$
y_{t+1} = c_t + \sum_{\ell_1=0}^{p-1} \rho_{\ell_1} y_{t-\ell_1} + \sum_{\ell_2=0}^{q-1} \beta_{\ell_2} X_{t-\ell_2} + u_{t+1}, \tag{19}
$$

where X_t is a scalar predictor. We set $p = 4$ when non-local estimator is used and $p = 1$ for local estimator (fewer lags when parameter instability is taken into consideration). q is set to 1 in all cases.

The data set is collected at a quarterly frequency, with a sample period of 1962Q3-2023Q1. The choices of predictors follow closely from Inoue et al. (2017) and consist of asset prices, measures of real economy activity, price indices and monetary measures. For U.S., both the target variables and predictors are taken from Fred-QD (McCracken et al. (2021)). The data for Canada are taken from FRED and Stevanovic et al. (2021). Tables 13-14 lists mnemonics (in the associated database), data description and

data transformation for all the predictors. The initial estimation sample runs from 1962Q3 to 1984Q4 and the first available forecast is for 1985Q1.

To avoid the complication due to Covid-19, we first consider results from the evaluation period 1985Q1-2019Q4. Tables 3-4 report the overall forecasting performance. Let us start by commenting the results obtained for U.S. Overall, UCSV is indeed a tough benchmark as there are many cases when UCSV performs better than the alternative specifications. However, there are also some cases when local estimator with optimal tuning parameter selection improves forecast accuracy compared to the benchmark. The best result is achieved when the changes of employment numbers (CE16OV) is used as a predictor. It already provides benefits when CE16OV is added to an AR(4) model, and using $K_2(u)$ with optimal tuning parameter selection gives additional benefits. $K_2(u)$ is also useful when stock return (S&P 500), volatility (VXOCLSx), some measures of real economic activity (consumption, investment and industrial production) and monetary measures are used.

Moving to the results obtained for Canada, some different patterns clearly emerge. Using $K_2(u)$ with optimal tuning parameter selection always improves forecast accuracy compared to the benchmark UCSV model. The best result is archived when stock return (TSX_CLO) is used as the predictor. Interesting, when using non-local estimator for the ARDL(1,1) model, they are always outperformed by the benchmark. This clearly shows that ignoring issues related to parameter instability can be detrimental to forecast accuracy.

Interestingly, results from U.S. show that changes of employment is a useful predictor, indicating the traditional backward looking Phillips curve type of forecasting model is useful. However, the results do not hold when we use the unemployment directly as the predictor. To get a better understanding on the source of gains, we evaluate the models' forecasting performance over time by plotting the associated cumulative sum of MSFEs over time for these two predictors in Figure 2. The solid line shows the results obtained from using non-local estimator and the dashed line shows the results from local estimator with optimal tuning parameter selection when $K_2(u)$ is used as the weighting function. When changes of employment is used, the gains from using local estimator is initially negative. Non-local estimator performs better than the local estimator before 2010, but it is outperformed by the local estimator afterwards. When unemployment rate is used, using non-local estimator improves forecast accuracy compared to the UCSV benchmark during the period 1990-1996. However, the gains become negative afterwards. Using local estimator is not that useful as the gains are always negative.

Another interesting findings are when monthly inflation (changes of CPI) is used, it performs better for Canada but not for U.S. We again plot the associated cumulative sum of MSFEs over time in Figure 3. The solid line shows the results obtained from using non-local estimator and the dashed line shows the results from local estimator with optimal tuning parameter selection when $K_2(u)$ is used as the weighting function. For U.S., even though the gains remain negative for the entire evaluation period, the forecasting performances are relatively stable over time. However, the performances for Canada change marked after 2010. The gains from using local estimator mostly come from that period. When non-local estimator is

used, gains are positive from 2004 to 2008, but quickly becomes negative during 2009-2010.

In Appendix D, we report the results from the evaluation period 2020Q1-2023Q1. Tables D.1-D.2 presents the results for U.S. and Canada, respectively. Overall, as clearly shown in the first row of each table, RMSFEs from UCSV model are larger compared to the pre-Covid, which implies that it is difficult to get precise inflation forecasts in this turbulence time. For U.S., non-local estimator works better in this period, and in general, it performs better than the benchmark model. There are still cases in which local estimator with optimal tuning parameter selection works better, particularly when unemployment rate is used as the predictor. However, they are generally outperformed by the AR(4) model, except when term spreads are used as the predictor. For Canada, $K_2(u)$ with optimal tuning parameter selection still improves forecast accuracy when predictors are from the category of asset prices and price indices. The best results are obtained when unemployment rate is used as the predictor, indicating that backward looking Phillips curve forecasting model still works relatively well.

5.2. Forecasting growth and inflation shocks

Let y_{t+1} be the macroeconomic shock whose conditional quantile we wish to forecast based on systemic risk measures. The τth conditional quantile of y_{t+1} is affine function of observable x_t:

$$\mathbb{Q}_\tau(y_{t+1}|x_t) = \alpha_\tau + \beta_\tau x_t, \tag{20}$$

where x_t is a systematic risk measure. $\mathbb{Q}_\tau(y_{t+1}|x_t) := \inf\{y : F(y|x_t) \geqslant \tau\}$, $F(\cdot|x_t)$ is the conditional c.d.f. of Y_{t+1} given $X_t = x_t$, with density $f(\cdot|x_t)$. As in Koenker and Bassett (1978), quantile regression estimator $\theta_\tau = (\alpha_\tau, \beta_\tau)'$ minimizes the sample analog of the check loss based on the empirical distribution, namely,

$$L_T(\theta;\tau) = \frac{1}{T}\sum_{t=2}^{T} \ell_t^\tau(\theta), \tag{21}$$

where $\ell_t^\tau(\theta) = \varepsilon_{t+1}(\theta)[\tau - \mathbb{1}_{\varepsilon_{t+1}(\theta)<0}]$ and $\varepsilon_{t+1}(\theta) = y_{t+1} - \alpha_\tau - \beta_\tau x_t$. The objective function for the local estimator in (3) and local linear estimator in (11) can be similarly defined.

As the check loss is not differentiable, expansion like (5) cannot be directly used. We follow Fernandes et al. (2021) to apply kernel smoothing to the empirical objective function[7] in (21), yielding

$$\tilde{L}_T(\theta;\tau) = \frac{1}{T}\sum_{t=2}^{T} \tilde{\ell}_t^\tau(\theta), \tag{22}$$

where $\tilde{\ell}_t^\tau(\theta) = (\ell^\tau(\theta) * K_{b^*}^*)(\varepsilon_{t+1}(\theta)) = \int_{-\infty}^{\infty} \ell^\tau(v;\theta) K_h^*(v - \varepsilon_{t+1}(\theta))dv$ and $*$ is the convolution operator. $K_{b^*}^*(\cdot)$

[7] We use the kernel smoothing objective function to obtain the implementable regret loss in this application. Estimation is still based on the empirical distribution as in (21), as we find that it is numerically more stable and delivers better forecasting performance.

is another weighting function with a tuning parameter $b^* > 0$. It can be shown that the second order derivative of (22) with respect to θ is

$$\tilde{L}_T^{(2)}(\theta;\tau) = \frac{1}{T}\sum_{t=2}^{T} X_t X_t' K_{b^*}^*(-\varepsilon_{t+1}(\theta)), \tag{23}$$

where $X_t = [1\ x_t]'$. The local version of (23) can be defined accordingly, which will be used to approximate the weighting matrix $E_T\left(\frac{\partial \ell_{T+1}^2(\theta)}{\partial\theta\partial\theta'}\right)$ defined in (18), which leads to feasible regret loss:

$$R(K,b) = (\hat{\theta}_{K,b,T} - \theta(1))'\, \tilde{L}_T^{(2)}(\theta;\tau)\, (\hat{\theta}_{K,b,T} - \theta(1)).$$

What remains is to choose K^* and b^*. We use an exponential-type weighting function $K^*(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ and set $b^* = \{(2+\log T)/T\}^{2/5}$ (He et al. (2021)).

We consider both growth and inflation shocks for the United States. As in Giglio et al. (2016), growth shocks are measured as the residuals from an AR(p) regression based on the percentage changes of industrial production (IP) index. Inflation shocks are obtained similarly as the residuals from an AR(p) regression based on the percentage changes of consumer price index (CPI). Both IP and CPI data are taken from FRED. The systematic risk measures we consider include CatFin (Allen et al. (2012)), default spread, TED spread, term spread, slope factor of the yield curve, VIX and stock return. The details of variable descriptions and data sources are provided in Table 15. Due to the availability of CatFin, our sample period starts in 1973M1 and ends in 2022M12. The initial estimation sample runs from 1973M1 to 1989M12 and the first available forecast is for 1990M1.

As in Giglio et al. (2016), the macroeconomic shock series are constructed carefully to preserve the out-of-sample nature. This means that the forecast of a macroeconomic shock at time $T+1$ is constructed using only information from the estimation sample up to time T. We fit IP growth and inflation series using an AR(13) model and estimate it using Bayesian methods with natural conjugate prior. The forecast residual at time $T+1$ is constructed based on these estimates. The benchmark forecasts are based on the historical unconditional quantile. We consider the left tails (5th, 10th and 15th percentiles) for IP growth shock and both left tails (5th, 10th and 15th percentiles) and right tails (85th, 90th and 95th percentiles) for inflation shocks.

Table 5 reports the results for IP growth shocks forecasts. The only systemic risk measure achieves gains across all three percentiles and weighting functions are TED spread. Using local estimator provides additional benefits. $K_2(u)$ delivers best results for 5th and 10th percentiles but $K_1(u)$ performs better in the 15th percentile. Most systemic risk measures improve forecast accuracy at the 5th percentile and generally performs better than the local estimator, but gains from measures such as term spread and slope factor get lost at the 10th and 15th percentiles. Default spread with non-local estimator is the best at 15th percentile, but performance from $K_1(u)$ with tuning parameter selection is quite close.

Table 6 reports the results for inflation shocks forecasts. Let us start by commenting results from the left tail. First, except for term spread at 5th percentile, all systemic risk measures with tuning parameter selection deliver benefits in all cases. Second, TED spread with fixed rolling window forecasts is the best at 5th percentile, but stock return with $K_2(u)$ is the best at both 10th and 15th percentiles. Third, TED spread is very useful as it always improves forecast accuracy no matter what types of weighing function is used. Moving to the right tail, we see that $K_2(u)$ is still useful at both 5th and 10th percentile as it is generally beneficial for all systemic risk measures. TED spread is again very helpful at these two percentiles and local estimator performs better than the non-local ones. Results are slight different at the 15th percentile, but local estimator still provides gains if Catfin or Term spread is used.

5.3. Forecasting house prices changes

Let $y_t = 100 \ln(P_t/P_{t-1})$ be the growth rate of real house prices (P_t). Following Ghysels et al. (2013), our baseline specifications involve predictability based on both serial correlation (using lag) and valuation ratio (using price-to-rent ratio):

$$y_{t+1} = \alpha + \beta y_t + \varepsilon_{t+1}, \tag{24}$$

$$y_{t+1} = \alpha + \beta hp_t + \varepsilon_{t+1}, \tag{25}$$

where $hp_t = \ln(H_t) - \ln(P_t)$, H_t is the net of all operating expenses of a property and P_t denotes its current price.

Since there is ample evidence that economic variables are associated with future house prices variations (Campbell et al. (2009)), we would like to know which variable would be useful to improve forecast accuracy, which lead to the following model specification:

$$y_{t+1} = \alpha + \beta y_t + \gamma X_t + \varepsilon_{t+1}, \tag{26}$$

$$y_{t+1} = \alpha + \beta hp_t + \gamma X_t + \varepsilon_{t+1}. \tag{27}$$

Finally, as economic variables are often available at a higher frequency than y_t and hp_t, we also consider the following mixed data sampling (MIDAS) regression model specification:

$$y_{t+1} = \alpha + \beta_1 y_t + \beta_2 B(L^{1/m}; \delta) x_t^{(m)} + \varepsilon_{t+1}, \tag{28}$$

$$y_{t+1} = \alpha + \beta_1 hp_t + \beta_2 B(L^{1/m}; \delta) x_t^{(m)} + \varepsilon_{t+1}, \tag{29}$$

where $B(L^{1/m}; \delta) = \sum_{k=1}^{K} b(k; \delta) L^{(k-1)/m}$, $L^{s/m} x_t^{(m)} = x_{t-s/m}^{(m)}$ and $B(1; \delta) = \sum_{k=1}^{K} b(k; \delta) = 1$. Here t indexes the basic time unit (quarters, in our case), m is the higher sampling frequency ($m = 3$ when x is monthly and y is quarterly), and, as shown, $L^{1/m}$ operates at this higher frequency. As explained in Andreou et al. (2010), when X_t is available at a higher frequency, using higher frequency data directly may be

more efficient. Thus, it is of great interests to examine whether MIDAS specification helps to improve forecasting performance.

The data for P_t and hp_t are taken from the International Housing Observatory Database. This is a publicly available database which covers a number of countries. The detailed source of description and methodology can be found in Mack et al. (2011). Due to the data availability of hp_t, the countries we consider include 4 of the G7 countries (United States, Canada, France and Germany) as well as Australia. Economic variables we consider include stock return, treasury-bill rate, term spread (measured by the difference between long term, government bond yield and treasury-bill rate), inflation (measured by changes of CPI) and growth (measured by changes of industrial production), which are the ones considered in Ghysels et al. (2013)[8]. Table 16 provides data sources, detailed descriptions and data transformations for the predictors. Our sample starts from 1975Q3 and ends at 2022Q4 (1975M7 – 2022M12 for monthly predictors). The initial estimation sample runs from 1975Q3 to 1995Q4 and the first available forecast is for 1996Q1.

Model parameters in specifications (24)-(25) and (26)-(27) can be estimated by LS method. For the MIDAS specification (28)-(29), since y_t and hp_t are available at quarterly frequency and economic variable is sampled at monthly frequency, we have $m = 3$. For the parametrization of $b(k;\delta)$. We use Beta polynomials, which is based on the Beta probability density function which involves two parameters $\delta = (\delta_1, \delta_2)'$:

$$b(k;\delta_1,\delta_2) = \frac{f(\frac{k}{K};\delta_1,\delta_2)}{\sum_{k=1}^{K} f(\frac{k}{K};\delta_1,\delta_2)},$$

where $f(x;\delta_1,\delta_2) = \frac{x^{\delta_1-1}(1-x)^{\delta_2-1}\Gamma(\delta_1+\delta_2)}{\Gamma(\delta_1)\Gamma(\delta_2)}$ and $\Gamma(\delta_1) = \int_0^\infty e^{-x}x^{\delta_1-1}dx$. Following Ghysels and Qian (2019), we set $\delta_1 = 1$ and restrict δ_2 to be larger than 1 to ensure a downward sloping weighting scheme. Then, the model parameters are estimated by nonlinear LS method as in Andreou et al. (2010). Finally, for the number of high frequency lags, we set $K = 3$.

Table 7 reports the overall forecasting performance based on the specification with lag ((24), (26) and (28)). Table 8 reports the overall forecasting performance based on the specification with valuation ratio ((25), (27) and (29)). The upper panel in each country presents the results obtained from (26) and the bottom panel is for the results obtained from MIDAS specification (28).

Let us start by commenting the results obtained from Table 7. There are several findings. First, all the alternatives are outperformed by the benchmark in the United States, indicating that the specification with lag is a tough benchmark. This is also true for France, as there are only two cases in which alternatives perform better. Second, for Canada, fixed rolling window forecast with specification (26) and treasury bill as a predictor has the overall best performance. Finally, we see that local estimator with optimal tuning

[8]Since monthly CPI and INDPRO are not available in Australia, these two variables are not investigated with specification (28)-(29).

parameter selection is useful for Germany and Australia. For Germany, using specification (26) with local estimator (either $K_1(u)$ and $K_3(u)$) always improves compared to the benchmark. $K_2(u)$ is a better choice for Australia as it always delivers gains. Overall, the forecasting performance based on the specification with lag is rather heterogenous across countries.

Moving to the results in Table 8, some different stories are evident. First, there are more entries that the numbers are below 1, indicating that local estimators are more likely to perform better than the non-local benchmark. Second, for all countries, using MIDAS specification with treasury bill rate always improves forecast accuracy, and in general not using all data ($K_1(u)$ and $K_3(u)$) are better (except Germany). $K_1(u)$ also in general deliver gains compared to the benchmark, but Canada is an exception, except for the case when treasury bill rate is used as the predictor. Finally, $K_3(u)$ is always the best for United States and France. Overall, the parameter instability is more pervasive in the specification with valuation ratio, and using MIDAS specification with treasury bill rate as an additional predictor is preferred.

To provide a better understanding of the source of the gains, we plot the cumulative sums of MSFEs differences (relative to the benchmark (24) or (25)) over the evaluation sample. The results from (28)-(29) are in Figure 4 and the results from (26)-(27) are plotted Figure 5. Based on the findings in Table 8, we only consider the case when treasury bill rate is used as the predictor and $K_3(u)$ is used as the weighting function. For all countries, the gains in the specification with valuation ratio are generally positive over the evaluation sample. For Germany, the gains markedly increase afterwards for the specification with valuation ratio. The gains in the specification with lag are also present in United States and Canada (with MIDAS) before 2008, but the predictability gets lost afterwards. Finally, patterns from (26)-(27) are roughly similar. For Canada, the gains with lag specification are less evident (compared to MIDAS) in the initial evaluation sample, and the overall better performance is mostly from the COVID period (after 2019).

5.4. Bond return predictability

As in Cochrane and Piazzesi (2005), we use the following notation for the (log) yield of an n-year bond:

$$y_t^{(n)} = -\frac{1}{n} p_t^{(n)},$$

where $p_t^{(n)}$ is the log price of the n-year zero-coupon bond at time t. The holding-period return of buying an n-year bond at time t and selling it as an $(n-1)$-year bond at time $t+12$ is

$$r_{t+12}^{(n)} = p_{t+12}^{(n-1)} - p_t^{(n)}.$$

The excess return is

$$rx_{t+12}^{(n)} = r_{t+12}^{(n)} - y_t^{(1)},$$

where $y_t^{(1)}$ is the one-year risk-free rate.

We consider three different specifications to assess whether the excess bond returns $rx_{t+12}^{(n)}$ are predictable:

(i) Fama-Bliss (FB) univariate

$$rx_{t+12}^{(n)} = \alpha + \beta fs_t^{(n)} + \varepsilon_{t+12};$$

(ii) Cochrane-Piazzesi (CP) univariate

$$rx_{t+12}^{(n)} = \alpha + \beta CP_t + \varepsilon_{t+12};$$

(iii) Fama-Bliss and Cochrane-Piazzesi predictors

$$rx_{t+12}^{(n)} = \alpha + \beta_1 fs_t^{(n)} + \beta_2 CP_t + \varepsilon_{t+12}.$$

The Fama-Bliss (FB) forward spreads are given by

$$fs_t^{(n)} = f_t^{(n)} - y_t^{(1)} = p_t^{(n-1)} - p_t^{(n)} - y_t^{(1)}.$$

The Cochrane-Piazzesi (CP) factor is constructed as the linear combination of forward rates:

$$CP_t = \hat{\gamma}' \boldsymbol{f}_t,$$

where $\boldsymbol{f}_t = (y_t^{(1)}, f_t^{(2)}, f_t^{(3)}, f_t^{(4)}, f_t^{(5)})'$. The coefficient vector $\hat{\gamma}$ is estimated from a predictive regression of $\frac{1}{4}\sum_{n=2}^{5} rx_{t+12}^{(n)}$ on $[1 \ \boldsymbol{f}_t']'$.

We study excess bond return predictability in four bond markets: United States, Canada, United Kingdom and Japan, which are among the largest bond markets in the world. The yield data for United States are taken from Liu and Wu (2021). The yield data for Canada and UK are obtained from Bank of Canada and Bank of England, respectively. The yield data for Japan are collected from Ministry of Finance Japan. Since the holding period we consider is up to 5 years, all yield data are collected up to 5 years maturity. Due to data availability in different countries, our sample period runs from 1986M1 to 2022M12. A more detailed description of the data is provided in Table 17. The initial estimation sample runs from 1986M1 to 1999M12 and the first available forecast is for 2000M1. The benchmark forecasts are obtained from the three principal components (PC) of the global yield curve (by stacking all yield data together from four markets).

Tables 9-11 report the forecasting results for all four bond markets. In each market, we consider four different maturities: 2 years, 3 years, 4 years and 5 years. The entries for benchmark PCs of yield curve forecasts are the RMSFEs in levels and all other entries are ratios of RMSFEs relative to the benchmark. Overall, the results are very promising, particularly when $K_3(u)$ is used with optimal tuning parameter

selection, as it provides sizable and sometimes significant improvement over the benchmark forecasts in all cases. For Japan, $K_2(u)$ leads to better forecasting performance in some cases, but the results are very close to the ones obtained from $K_3(u)$. Forecasts from all local estimators provide gains for Canada, but choice of weighting functions generally matters for other bond markets. Finally, standard predictive regression by using non-local least square estimator deteriorates forecasting performance for some markets, particularly for Canada. This again shows that parameter instability matters and ignoring it may lead to forecast failure.

6. Conclusion

Parameter instability is pervasive in forecasting models, and local estimator is often used in the presence of parameter instability. In this paper, we first provide conditions on the parameter instability to achieve consistency of the local estimator. It is shown that local estimator can handle a broad range of parameter instability considered in the literature, which includes local structural break, smooth structural change and realization of persistent and bounded stochastic process. The consistency rate depends on the amount of local variation and we obtain faster rate when these variations are small.

We then move on to the analysis of OOS forecasting. We focus on the end-of-sample risk and show that under certain conditions, minimizing the end-of-sample risk is equivalent to minimize the regret risk, which depends on the choices of tuning parameter and weighting function. We propose method to select tuning parameter by directly minimizing the regret risk. This is similar to Inoue et al. (2017), but we show that asymptotic optimality holds when a generic weighting function is used for estimation and a general loss function is used for forecast evaluation. We also provide analyses on the choice of weighting function, which has been less addressed in the literature. Our analyses are based on the limiting behavior of the regret risk, which reflects the usual bias-variance trade-off. When the estimation variance dominates, the criteria to select the weighting function is quite simple. When estimation bias dominates, the criteria is more involved as it depends on the property of parameter time variation. However, it still provides guidance on the implementation of our tuning parameter selection procedure.

Our theoretical analyses are evaluated through an extensive Monte Carlo study with linear predictive regression model. We find that local estimation performs well under various form of parameter instability. Our tuning parameter selection procedure is also useful in forecasting. In general, using all data and downweighting them is preferred.

We present four empirical applications. Our methods are quite useful and they generally improves forecast accuracy. Weighting functions do matter in the forecasting performance. While we find using all data and downweighting them is preferred in the application of forecasting inflation and growth (inflation) shocks, using only recent data and downweighting them is more useful in forecasting bond returns and real house price changes.

References

Abadir, K.M., Magnus, J.R., 2005. Matrix algebra. volume 1. Cambridge University Press.

Allen, L., Bali, T.G., Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns? The Review of Financial Studies 25, 3000–3036.

Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. Journal of Econometrics 158, 246–261.

Bauer, M.D., Hamilton, J.D., 2018. Robust bond risk premia. The Review of Financial Studies 31, 399–448.

Borup, D., Eriksen, J.N., Kjær, M.M., Thyrsgaard, M., 2023. Predicting bond return predictability. Management Science .

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163–188.

Campbell, S.D., Davis, M.A., Gallin, J., Martin, R.F., 2009. What moves housing markets: A variance decomposition of the rent–price ratio. Journal of Urban Economics 66, 90–102.

Carriero, A., Clark, T.E., Marcellino, M., 2022. Nowcasting tail risk to economic activity at a weekly frequency. Journal of Applied Econometrics 37, 843–866.

Chan, J.C., 2018. Specification tests for time-varying parameter models with stochastic volatility. Econometric Reviews 37, 807–823.

Chen, B., Hong, Y., 2016. Detecting for smooth structural changes in garch models. Econometric Theory 32, 740–791.

Clark, T.E., Huber, F., Koop, G., Marcellino, M., Pfarrhofer, M., 2023. Tail forecasting with multivariate bayesian additive regression trees. International Economic Review 64, 979–1022.

Cochrane, J.H., Piazzesi, M., 2005. Bond risk premia. American economic review 95, 138–160.

Coroneo, L., Iacone, F., 2020. Comparing predictive accuracy in small samples using fixed-smoothing asymptotics. Journal of Applied Econometrics 35, 391–409.

Dendramis, Y., Giraitis, L., Kapetanios, G., 2021. Estimation of time-varying covariance matrices for large datasets. Econometric Theory 37, 1100–1134.

Dendramis, Y., Kapetanios, G., Marcellino, M., 2020. A similarity-based approach for macroeconomic forecasting. Journal of the Royal Statistical Society Series A: Statistics in Society 183, 801–827.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics , 253–263.

Farmer, L., Schmidt, L., Timmermann, A., 2022. Pockets of predictability. Journal of Finance, forthcoming .

Fernandes, M., Guerre, E., Horta, E., 2021. Smoothing quantile regressions. Journal of Business & Economic Statistics 39, 338–357.

Gargano, A., Pettenuzzo, D., Timmermann, A., 2019. Bond return predictability: Economic value and links to the macroeconomy. Management Science 65, 508–540.

Ghysels, E., Plazzi, A., Valkanov, R., Torous, W., 2013. Forecasting real estate prices. Handbook of economic forecasting 2, 509–580.

Ghysels, E., Qian, H., 2019. Estimating midas regressions via ols with polynomial parameter profiling. Econometrics and statistics 9, 1–16.

Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. The Review of Economic Studies 76, 669–705.

Giglio, S., Kelly, B., Pruitt, S., 2016. Systemic risk and the macroeconomy: An empirical evaluation. Journal of Financial Economics 119, 457–471.

Giraitis, L., Kapetanios, G., Yates, T., 2014. Inference on stochastic time-varying coefficient models. Journal of Econometrics 179, 46–65.

Granger, C.W., 1969. Prediction with a generalized cost of error function. Journal of the Operational Research Society 20, 199–207.

Hansen, P.R., Dumitrescu, E.I., 2022. How should parameter estimation be tailored to the objective? Journal of Econometrics 230, 535–558.

He, X., Pan, X., Tan, K.M., Zhou, W.X., 2021. Smoothed quantile regression with large-scale inference. Journal of Econometrics .

Hirano, K., Wright, J.H., 2017. Forecasting with model uncertainty: Representations and risk reduction. Econometrica 85, 617–643.

Inoue, A., Jin, L., Pelletier, D., 2021. Local-linear estimation of time-varying-parameter garch models and associated risk measures. Journal of Financial Econometrics 19, 202–234.

Inoue, A., Jin, L., Rossi, B., 2017. Rolling window selection for out-of-sample forecasting with time-varying parameters. Journal of econometrics 196, 55–67.

Kapetanios, G., Marcellino, M., Venditti, F., 2019. Large time-varying parameter vars: A nonparametric approach. Journal of Applied Econometrics 34, 1027–1049.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica , 33–50.

Kristensen, D., Lee, Y.J., 2023. Local polynomial estimation of time-varying parameters in nonlinear models. arXiv preprint arXiv:1904.05209 .

Leamer, E.E., 2007. Housing is the business cycle.

Li, H., Müller, U.K., 2009. Valid inference in partially unstable generalized method of moments models. The Review of Economic Studies 76, 343–365.

Liu, Y., Wu, J.C., 2021. Reconstructing the yield curve. Journal of Financial Economics 142, 1395–1425.

Mack, A., Martínez-García, E., et al., 2011. A cross-country quarterly database of real house prices: a methodological note. Globalization and Monetary Policy Institute Working Paper 99.

McCracken, M.W., Ng, S., et al., 2021. Fred-qd: A quarterly database for macroeconomic research. Federal Reserve Bank of St. Louis Review 103, 1–44.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handbook of econometrics 4, 2111–2245.

Oh, D.H., Patton, A.J., 2021. Better the devil you know: Improved forecasts from imperfect models .

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

Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. Springer.

Rossi, B., 2013. Advances in forecasting under instability, in: Handbook of economic forecasting. Elsevier. volume 2, pp. 1203–1324.

Stevanovic, D., Surprenant, S., Leroux, M., Fortin-Gagnon, O., 2021. Large canadian database for macroeconomic analysis (lcdma) - vintages.

Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business & Economic Statistics 14, 11–30.

Stock, J.H., Watson, M.W., 2003. Forecasting output and inflation: The role of asset prices. Journal of economic literature 41, 788–829.

Stock, J.H., Watson, M.W., 2007. Why has us inflation become harder to forecast? Journal of Money, Credit and banking 39, 3–33.

Weiss, A.A., 1996. Estimating time series models using the relevant cost function. Journal of Applied Econometrics 11, 539–560.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.

Figures and Tables

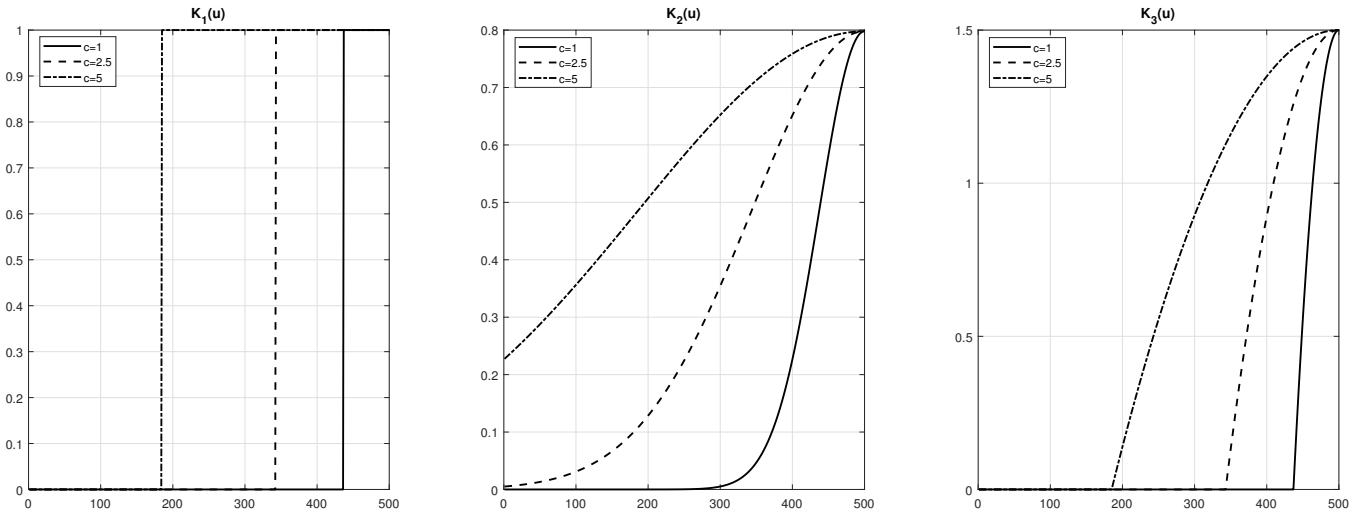

Figure 1: Shape of the weighting scheme with $T = 500$, $b = cT^{-1/3}$ with c equal to 1,2.5 and 5.

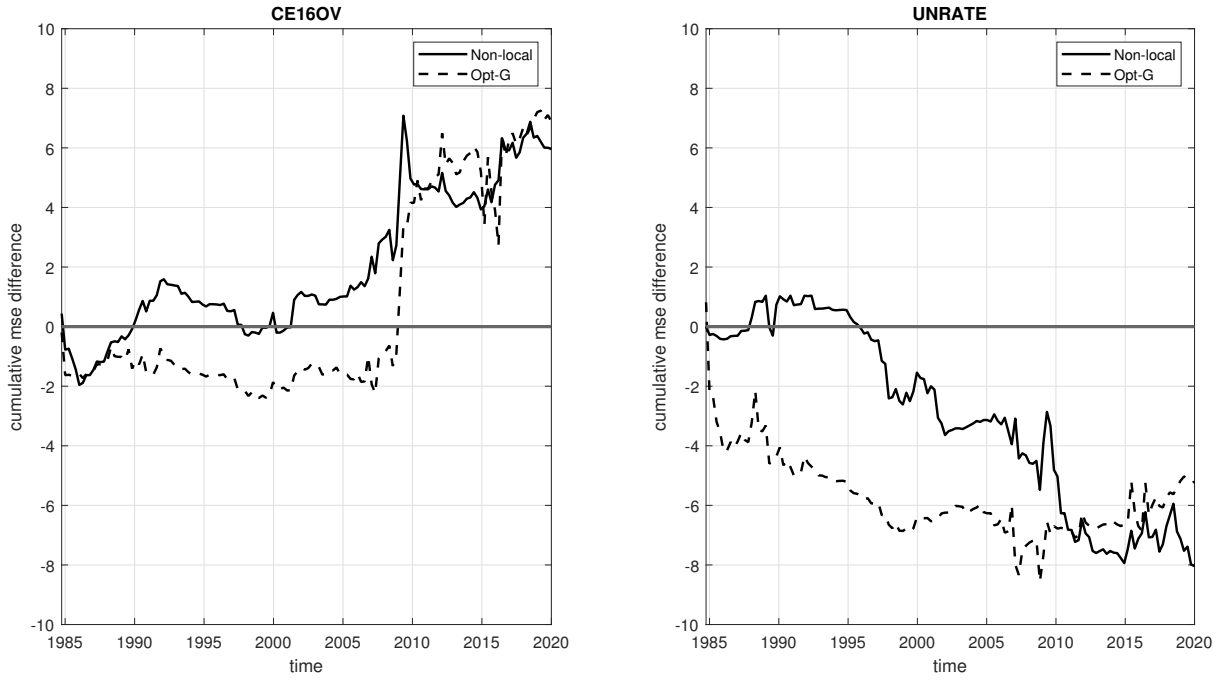

Figure 2: The figure presents cumulative sums (taken over time) of MSFEs differences (relative to the UCSV benchmark) for inflation forecasts. The predictors are CE16OV (changes of civilian employment) and UNRATE (unemployment rate). Non-local: full sample LS estimation from ARDL(1,1); Opt-G: optimal tuning parameter selection with $K_2(u)$.

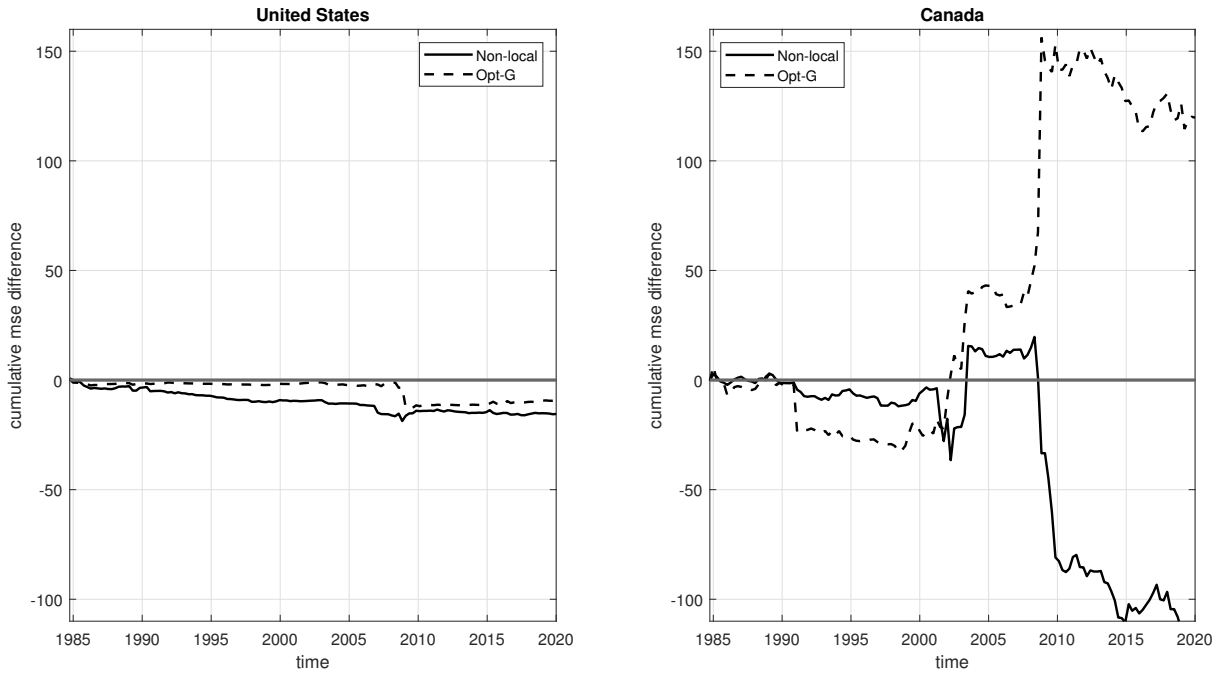

Figure 3: The figure presents cumulative sums (taken over time) of MSFEs differences (relative to the UCSV benchmark) for inflation forecasts when changes of CPI is used as the predictor. The left panels shows the results obtained for U.S. and the right panel shows the results for Canada. Non-local: full sample LS estimation from ARDL(1,1); Opt-G: optimal tuning parameter selection with $K_2(u)$.

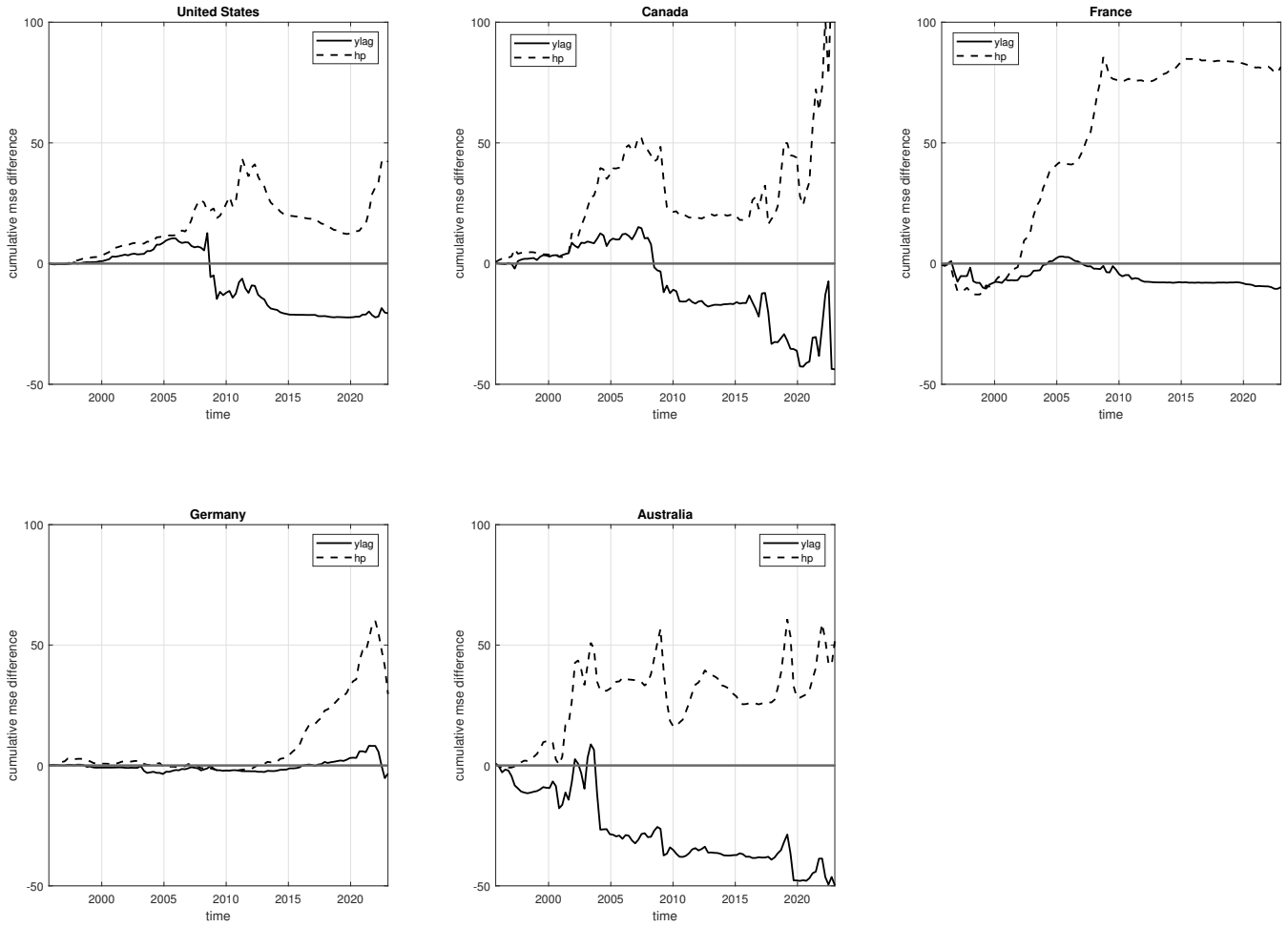

Figure 4: The figure presents cumulative sums (taken over time) of MSFEs differences from MIDAS model (relative to the non-local benchmark without any economic variable) for real house price growth forecasts. The predictor is 3-month treasury bill rate and weighting function used is $K_3(u)$. The solid line shows results obtained from the specification with lag and dashed line shows results obtained from the specification with valuation ratio.

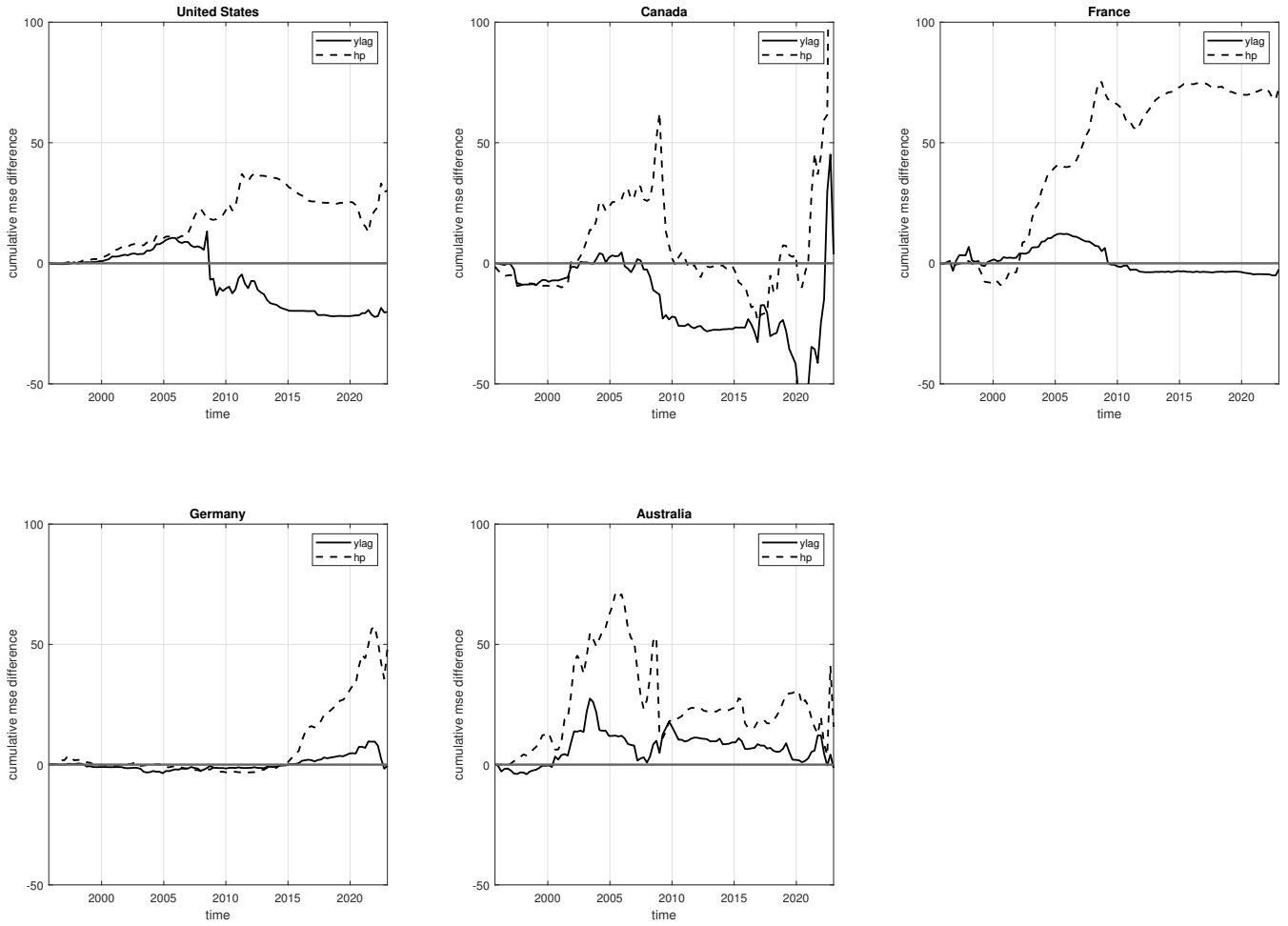

Figure 5: The figure presents cumulative sums (taken over time) of MSFEs differences from standard predictive regression model (relative to the non-local benchmark without any economic variable) for real house price growth forecasts. The predictor is 3-month treasury bill rate and weighting function used is $K_3(u)$. The solid line shows results obtained from the specification with lag and dashed line shows results obtained from the specification with valuation ratio.

Table 1: Small sample properties of the local estimator: Average MADs

DGP	Opt-R	Opt-G	Opt-E	Opt-R	Opt-G	Opt-E	Opt-R	Opt-G	Opt-E	Opt-R	Opt-G	Opt-E
						$a(1)$						
	$T = 100$			$T = 200$			$T = 300$			$T = 400$		
						One-time structural break						
2	0.079	0.049	0.077	0.052	0.034	0.046	0.030	0.027	0.034	0.026	0.022	0.029
3	0.090	0.106	0.089	0.064	0.049	0.060	0.051	0.038	0.048	0.041	0.031	0.040
4	0.101	0.259	0.090	0.063	0.199	0.058	0.050	0.161	0.046	0.044	0.134	0.040
						Smooth structural change						
5	0.052	0.045	0.051	0.040	0.036	0.038	0.033	0.031	0.033	0.030	0.028	0.029
6	0.066	0.076	0.060	0.052	0.062	0.047	0.044	0.054	0.039	0.040	0.049	0.036
7	0.021	0.014	0.023	0.015	0.010	0.016	0.012	0.009	0.013	0.011	0.008	0.012
8	0.279	0.261	0.280	0.251	0.240	0.249	0.238	0.227	0.235	0.223	0.217	0.221
9	0.248	0.234	0.245	0.213	0.209	0.207	0.198	0.193	0.192	0.185	0.184	0.177
10	0.206	0.193	0.202	0.168	0.163	0.163	0.153	0.152	0.146	0.141	0.141	0.134
11	0.165	0.151	0.166	0.128	0.125	0.125	0.114	0.111	0.110	0.103	0.103	0.100
12	0.139	0.129	0.141	0.107	0.101	0.106	0.087	0.088	0.086	0.076	0.077	0.075
						$b(1)$						
	$T = 100$			$T = 200$			$T = 300$			$T = 400$		
						One-time structural break						
2	0.186	0.127	0.186	0.134	0.093	0.126	0.093	0.077	0.103	0.081	0.065	0.089
3	0.204	0.214	0.204	0.157	0.119	0.151	0.131	0.100	0.126	0.111	0.085	0.109
4	0.229	0.424	0.209	0.156	0.372	0.149	0.129	0.323	0.122	0.115	0.275	0.109
						Smooth structural change						
5	0.170	0.137	0.175	0.133	0.109	0.135	0.112	0.095	0.115	0.101	0.086	0.103
6	0.199	0.197	0.193	0.159	0.161	0.151	0.134	0.140	0.127	0.123	0.128	0.117
7	0.151	0.128	0.150	0.109	0.088	0.106	0.090	0.074	0.088	0.078	0.064	0.076
8	0.092	0.066	0.103	0.068	0.050	0.075	0.058	0.042	0.064	0.050	0.037	0.055
9	0.093	0.067	0.104	0.068	0.049	0.075	0.056	0.041	0.063	0.051	0.037	0.056
10	0.095	0.067	0.106	0.068	0.050	0.076	0.056	0.041	0.063	0.050	0.036	0.055
11	0.096	0.067	0.106	0.069	0.050	0.078	0.059	0.042	0.065	0.051	0.037	0.057
12	0.096	0.068	0.107	0.071	0.052	0.078	0.060	0.046	0.066	0.057	0.043	0.061

Notes: $R = 40$: rolling window forecast with window size equal to 40; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$.

Table 2: Forecasting performance from simulated dataset

DGP	$R = 40$	Opt-R	Opt-G	Opt-E	$R = 40$	Opt-R	Opt-G	Opt-E	$R = 40$	Opt-R	Opt-G	Opt-E	$R = 40$	Opt-R	Opt-G	Opt-E
	$T = 100$				$T = 200$				$T = 300$				$T = 400$			
Constant coefficient																
1	1.020	1.056	1.019	1.070	1.027	1.035	1.015	1.039	1.033	1.026	1.014	1.032	1.030	1.022	1.009	1.025
One-time structural break																
2	0.757	0.791	0.756	0.790	0.775	0.783	0.763	0.784	0.765	0.762	0.751	0.767	0.782	0.776	0.767	0.778
3	0.648	0.685	0.685	0.690	0.671	0.685	0.666	0.687	0.670	0.675	0.661	0.678	0.666	0.664	0.653	0.666
4	0.864	0.636	0.824	0.627	0.605	0.608	0.765	0.610	0.615	0.611	0.737	0.613	0.642	0.634	0.709	0.636
Smooth structural change																
5	0.931	0.950	0.928	0.956	0.922	0.927	0.913	0.930	0.921	0.920	0.907	0.923	0.918	0.914	0.904	0.916
6	0.821	0.808	0.812	0.814	0.780	0.780	0.784	0.783	0.763	0.760	0.765	0.761	0.764	0.757	0.760	0.760
7	0.842	0.863	0.835	0.865	0.852	0.862	0.836	0.862	0.834	0.829	0.818	0.834	0.843	0.836	0.825	0.838
8	1.006	1.028	1.003	1.034	1.002	1.004	0.991	1.004	1.012	1.007	0.997	1.013	1.001	0.997	0.992	0.997
9	0.988	0.994	0.981	1.001	0.979	0.980	0.970	0.981	0.976	0.976	0.970	0.977	0.973	0.971	0.967	0.971
10	0.966	0.971	0.956	0.972	0.949	0.953	0.946	0.952	0.949	0.948	0.942	0.949	0.952	0.948	0.944	0.948
11	0.959	0.975	0.952	0.980	0.945	0.951	0.937	0.955	0.919	0.919	0.911	0.920	0.912	0.907	0.905	0.908
12	0.901	0.915	0.896	0.923	0.881	0.885	0.876	0.887	0.875	0.873	0.869	0.875	0.853	0.847	0.844	0.849

Notes: $R = 40$: rolling window forecast with window size equal to 40; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$.

Table 3: Forecasting performance for inflation in the United States: 1985Q1-2019Q4

UCSV	0.571				
AR(4)	1.048				
	Non-local	$R = 40$	Opt-R	Opt-G	Opt-E
Asset prices					
FEDFUNDS	1.047	0.992	1.087	1.008	1.109
TB3MS	1.050	0.988	1.089	1.007	1.106
GS10	1.076	1.018	1.103	1.041	1.112
GS10TB3Mx	1.085	1.014	1.062	1.038	1.090
term spread	1.082	1.002	1.070	1.023	1.092
S&P 500	1.097*	1.020	0.994	0.984	1.016
VXOCLSx	1.039	1.033	1.016	0.972	0.964
Real economic activity					
DPIC96	1.011	0.990	0.987	0.966	1.009
GPDIC1	1.043	1.012	1.006	0.981	1.032
INDPRO	0.973	0.996	0.991	0.971	1.012
CE16OV	0.925	0.983	0.971	0.914	0.949
UNRATE	1.101	1.166	1.097	1.065	1.108
LNS14000026	1.114*	1.185	1.104	1.069	1.096
HOUST	1.076	1.038	1.049	1.008	1.063
PERMIT	1.057	1.020	1.028	0.990	1.037
Price indices					
CPIAUCSL	1.194*	1.155	1.156	1.117	1.263
CPIAPPSL	1.237*	1.038	1.027	1.007	1.062
CPIENGSL	1.154*	1.102	1.108	1.068	1.184
PPIACO	1.296*	1.047	1.014	1.023	1.087
PCECTPI	1.183	1.157	1.131	1.064	1.273
Monetary measures					
BOGMBASEREALx	1.209	0.998	1.041	0.993	1.235
M1REAL	1.064	0.995	1.016	1.000	1.051
M2REAL	1.091	1.025	1.031	0.999	1.086

Notes: The description of predictors is detailed in Table 13. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. The row "UCSV" presents the exact RMSFEs of the forecasts from the UCSV model. The row "AR(4)" presents the ratio of the RMSFEs of the forecasts from an AR(4) model relative to the benchmark. In other columns, the numbers are also the ratios of the RMSFEs relative to the benchmark. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 4: Forecasting performance for inflation in Canada: 1985Q1-2019Q4

UCSV	8.474				
AR(4)	1.107				
	Non-local	$R = 40$	Opt-R	Opt-G	Opt-E
Asset prices					
BANK_RATE_L	1.116	0.953	0.968	0.915	1.011
TBILL_3M	1.118	0.955	0.983	0.914	1.023
GOV_AVG_10pY	1.112	0.970	1.026	0.941	1.003
G_AVG_5.10.Bank_rate	1.123	0.960	0.971	0.935	0.988
G_AVG_10p.TBILL_3M	1.118	0.964	0.960	0.929	0.995
TSX_CLO	1.096	0.897	0.929	0.883	0.949
Real economic activity					
REAL_GDP	1.116	0.983	1.000	0.932	1.016
hhold_dispo_income	1.104	0.956	0.970	0.935	0.985
REAL_I	1.112	1.037	1.077	0.964	1.247
CANPROINDQISMEI	1.130	0.993	0.982	0.931	1.008
LFEMTTTTCAQ647S	1.109	1.031	1.006	0.946	1.062
UNEMP_CAN	1.110	0.985	1.019	0.942	1.002
hstart_CAN	1.109	0.945	0.972	0.929	0.991
Price indices					
CPI_ALL_CAN	1.113	0.916	0.908	0.899	0.946
IPPI_CAN	1.194	0.887	0.925	0.885	0.958
C_PRICE	1.099	0.906	0.962	0.905	0.971
Monetary measures					
MBASE1	1.103	0.944	0.958	0.919	1.027
CRED_BUS_cb	1.124	0.997	1.008	0.966	1.040
CRED_HOUS_cb	1.096	0.982	0.973	0.926	0.977

Notes: The description of predictors is detailed in Table 14. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. The row "UCSV" presents the exact RMSFEs of the forecasts from the UCSV model. The row "AR(4)" presents the ratio of the RMSFEs of the forecasts from an AR(4) model relative to the benchmark. In other columns, the numbers are also the ratios of the RMSFEs relative to the benchmark. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 5: Quantile IP growth shock forecasts

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	5th percentile						10th percentile				
UQ	0.122					UQ	0.161				
CatFin	0.973	1.042	0.975	0.980	0.986	CatfFin	1.003	1.028	1.023	1.003	0.988
Default spread	0.942	0.964	0.992	0.969	0.979	Default spread	0.959	1.005	0.990	0.973	0.980
TED spread	0.948	0.990	0.952	0.916*	0.950	TED spread	0.981	0.994	0.973	0.955	0.980
Term spread	0.978	1.061	1.004	1.002	1.022	Term spread	1.005	1.056	1.042	1.014	1.021
Slope factor	0.990	1.082	1.020	1.015	1.014	Slope factor	1.001	1.062	1.049	1.016	1.035
VIX	0.986	1.019	0.940	0.938	0.912	VIX	1.006	1.021	0.998	1.003	1.071
Stock return	1.010	1.083	1.017	0.994	1.010	Stock return	1.000	1.051	1.037	1.001	1.030
	15th percentile										
UQ	0.186										
CatFin	0.998	1.056	1.010	1.008	1.023						
Default spread	0.972	0.979	1.008	0.998	1.005						
TED spread	0.987	0.993	0.977	0.996	1.019						
Term spread	1.010	1.034	1.033	1.018	1.034						
Slope factor	1.005	1.042	1.042	1.014	1.034						
VIX	1.003	1.149*	1.061*	1.004	1.101*						
Stock return	1.003	1.054*	1.024	1.014	1.038						

Notes: See section 5.2 for the definition of the predictors. "UQ": historical unconditional quantile forecasts; $R = 60$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 7 (width 0.1). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the benchmark results (from unconditional historical quantile) are check losses in level and the other entries present the ratios of check losses relative to the benchmark. Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 6: Quantile inflation shock forecasts

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	5th percentile						10th percentile				
UQ	0.039					UQ	0.058				
CatFin	1.043	0.946	0.989	0.929	1.016	CatfFin	1.010	1.043	1.002	0.992	1.010
Default spread	1.053	0.839	0.893	0.851	0.992	Default spread	1.021	0.938	0.978	0.938	0.983
TED spread	1.027	0.834	0.954	0.848	0.910	TED spread	0.987	0.958	0.938*	0.924*	0.959
Term spread	1.016	1.035	0.989	1.011	1.067	Term spread	1.020	1.041	1.003	0.967	1.022
Slope factor	1.034*	1.033	1.016	0.989	0.987	Slope factor	1.019	1.036	1.007	0.960	1.039
VIX	0.996	1.301	1.062	0.975*	1.233	VIX	0.988	1.236*	1.014	0.948	1.057
Stock return	1.036	1.050	1.118	0.956	1.067	Stock return	1.011	0.943	0.928	0.922	0.945
	15th percentile										
UQ	0.070										
CatFin	1.000	1.025	0.993	0.980	1.022						
Default spread	1.022	0.968	0.997	0.991	1.010						
TED spread	0.981	0.963	0.949*	0.950*	0.961						
Term spread	1.020	1.010	1.015	0.963*	1.025						
Slope factor	1.022	1.004	0.999	0.963*	1.014						
VIX	0.996	1.091	1.011	0.986	1.022						
Stock return	1.010	0.948	0.957	0.921*	0.974						
	85th percentile						90th perentile				
UQ	0.068					UQ	0.052				
CatFin	1.007	0.976	0.934	0.961	0.956	CatfFin	1.007	1.011	0.959	0.948	0.975
Default spread	0.960	0.976	0.978	0.986	0.992	Default spread	0.954	1.028	1.022	0.964	1.014
TED spread	0.941	0.937	0.925	0.934	0.916	TED spread	0.947	0.949	0.956	0.937	0.948
Term spread	0.937	0.936	0.937	0.934	0.935	Term spread	0.932	0.924	0.928	0.926	0.932
Slope factor	0.935	0.929	0.943	0.932	0.925	Slope factor	0.927	0.941	0.969	0.924	0.933
VIX	1.001	1.061	1.007	0.956	1.019	VIX	1.003	1.217*	1.050	1.001	1.083
Stock return	1.023*	0.931	0.941	0.949	0.936	Stock return	1.029*	0.989	0.986	0.962	0.995
	95th percentile										
UQ	0.032										
CatFin	0.992	1.096	0.978	0.993	1.019						
Default spread	0.947	1.178	1.171	1.000	1.145						
TED spread	1.014	1.075	1.024	1.036	1.006						
Term spread	0.977	1.031	1.046	1.049	0.962						
Slope factor	0.965	1.040	1.014	1.033	1.031						
VIX	0.980	1.341*	1.262*	1.123	1.082						
Stock return	1.017	1.179	1.076	1.085	1.082						

Notes: See section 5.2 for the definition of the predictors. "UQ": historical unconditional quantile forecasts; $R = 60$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 7 (width 0.1). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the benchmark results (from historical unconditional quantile) are check losses in level and the other entries present the ratios of check losses relative to the benchmark. Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 7: Forecasting performance: specification with lag

	Non-local	$R = 40$	Opt-R	Opt-G	Opt-E		Non-local	$R = 40$	Opt-R	Opt-G	Opt-E
			United States						Canada		
OLS	1.452					OLS	5.804				
stock return	1.019	1.142	1.150	1.050	1.127	stock return	0.985	1.018	1.081	1.015	1.108
treasury bill	1.065	1.115	1.122	1.044	1.129	treasury bill	0.998	0.926	0.959	1.049	0.994
spread	1.014	1.105	1.140	1.030	1.116	spread	1.004	1.086	1.161	1.081	1.176
inflation	1.028	1.130	1.175	1.074	1.170	inflation	0.975	0.975	1.066	1.001	1.119*
growth	1.018	1.208	1.300	1.149	1.351	growth	1.041	1.079	1.158	1.085	1.194
MIDAS						MIDAS					
stock return	1.021	1.118	1.139	1.051	1.199	stock return	1.021	1.123	1.191	1.093	1.245*
treasury bill	1.055	1.133	1.148	1.046	1.131	treasury bill	1.006	0.994	1.027	1.071	1.070
spread	1.038	1.121	1.123	1.050	1.160	spread	1.000	1.068	1.149	1.075	1.174
inflation	1.026	1.158	1.163	1.126	1.182	inflation	0.965	1.062	1.165*	1.061	1.264
growth	1.023	1.133	1.092	1.038	1.078	growth	1.013	1.109	1.215*	1.105	1.211
			France						Germany		
OLS	1.166					OLS	0.848				
stock return	0.956	1.163	1.115	1.029	1.031	stock return	0.993	1.014	0.985	1.014	0.965
treasury bill	1.046	1.193	1.007	1.081	1.020	treasury bill	0.978	0.984	0.951	1.046	0.909
spread	1.005	1.240*	1.150	1.079	1.008	spread	0.978	1.004	0.988	1.079	0.982
inflation	1.012	1.272*	1.185	1.071	1.078	inflation	0.947	0.979	0.950	1.014	0.942
growth	1.012*	1.231*	1.165	1.090	1.064	growth	0.950	1.005	0.931	1.057	0.936
MIDAS						MIDAS					
stock return	1.037	1.227*	1.115	1.089	1.048	stock return	1.034	1.077	1.052	1.025	1.012
treasury bill	1.055	1.203	1.134	1.076	1.078	treasury bill	1.018	1.015	1.045	1.000	1.037
spread	1.001	1.251*	1.260	1.079*	1.078	spread	0.996	0.983	0.988	0.964	0.930
inflation	1.008	1.191*	1.032	1.058	0.956	inflation	0.998	1.076	1.072	0.991	1.093
growth	1.005	1.164	1.138	1.046	1.059	growth	0.982	0.948	0.945	0.943	0.898
			Australia								
OLS	2.573										
stock return	1.006	0.897	0.918	0.871	0.960						
treasury bill	0.992	1.045	0.990	0.965	1.005						
spread	1.001	0.941	0.916	0.952	0.929						
inflation	1.007*	1.088	1.178	0.991	1.130						
growth	0.988	1.021	1.040	0.978	1.057						
MIDAS											
stock return	1.023	1.126*	1.101	1.067	1.166*						
treasury bill	1.005	1.064	1.100	1.002	1.179						
spread	0.998	0.997	1.063	0.978	1.073						

Notes: The description of predictors is detailed in Table 16. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the entries present the ratios of MSFEs relative to the non-local benchmark (without any economic variable). Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 8: Forecasting performance: specification with valuation ratio

	Non-local	$R = 40$	Opt-R	Opt-G	Opt-E		Non-local	$R = 40$	Opt-R	Opt-G	Opt-E
			United States						Canada		
OLS	2.246					OLS	7.954				
stock return	1.018	1.145	1.267	1.014	1.007	stock return	0.916	1.015	1.010	0.919	1.015
treasury bill	1.028	1.028	1.162	1.002	0.876	treasury bill	0.968	0.822	0.835	0.865	0.802
spread	1.037	1.070	1.023	1.007	0.883	spread	1.045	1.144	1.089	1.042	1.082
inflation	1.079	1.141	1.196	0.994	0.917	inflation	1.002	1.113	1.077	1.041	1.104
growth	1.004	1.227*	1.375	1.039	1.071	growth	1.063	1.136	1.121	1.031	1.141
MIDAS						MIDAS					
stock return	1.014	1.109	0.970	0.992	0.859	stock return	0.992	1.145	1.018	0.989	1.030
treasury bill	1.014	1.096	0.956	0.967	0.825	treasury bill	0.935	0.842	0.792	0.912	0.799
spread	1.059	1.292	0.959	1.034	0.898	spread	1.000	1.150	1.020	0.996	1.046
inflation	1.004	1.109	0.953	0.926	0.825	inflation	1.017	1.133	1.078	1.058	1.102
growth	1.011	1.206	0.976	1.005	0.887	growth	1.012	1.152	1.004	0.998	1.006
			France						Germany		
OLS	2.202						1.600				
stock return	0.993	1.061	0.809	0.865	0.659	stock return	1.003	0.863	0.718	0.848	0.721
treasury bill	0.958	0.834	0.694	0.809	0.694	treasury bill	0.894	0.673	0.705	0.785	0.721
spread	0.997	1.045	0.818	0.838	0.721	spread	0.927	0.615	0.577	0.722	0.537
inflation	0.984	1.035	0.886	0.888	0.763	inflation	0.971	0.838	0.770	0.841	0.773
growth	1.031	1.007	0.793	0.899	0.666	growth	1.002	0.906	0.855	0.919	0.877
MIDAS						MIDAS					
stock return	1.011	1.108	0.827	0.902	0.745	stock return	1.033	0.939	0.898	0.958	0.824
treasury bill	0.962	0.958	0.737	0.802	0.658	treasury bill	0.912	0.763	0.832	0.816	0.828
spread	0.992	1.131	0.777	0.924	0.638	spread	0.936	0.651	0.633	0.753	0.554
inflation	1.016	1.075	0.781	0.928	0.705	inflation	1.012*	0.925	0.821	0.904	0.794
growth	1.025	1.069	0.804	0.883	0.731	growth	0.982	0.901	0.827	0.918	0.794
			Australia								
OLS	4.802										
stock return	0.986	0.958	0.918	0.926	0.949						
treasury bill	0.945*	0.948	0.932	0.889	0.970						
spread	0.998	0.964	0.847	1.022	0.866						
inflation	0.993	1.122	1.055	1.063	1.147						
growth	1.006*	1.103	1.066	1.065	1.099						
MIDAS											
stock return	1.005	1.078	0.915	0.976	0.948						
treasury bill	0.982	1.068	0.911	0.985	0.901						
spread	1.002	1.115	0.979	1.005	1.007						

Notes: The description of predictors is detailed in Table 16. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the entries present the ratios of MSFEs relative to the non-local benchmark (without any economic variable). Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 9: Out-of-sample forecasting performance on bond returns: United States

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	USA - 2 years						USA - 3 years				
PC-yields	1.592					PC-yields	6.046				
FB	1.047	1.150	1.103	0.958	0.852	FB	0.979	1.038	0.967	0.922	0.743
CP	1.113	1.122	0.949	1.005	0.744	CP	1.106	1.075	0.899	0.965	0.705
FB+CP	1.107	0.964	0.876	0.919	0.652	FB+CP	1.116	0.903	0.780	0.882	0.578*
	USA - 4 years						USA - 5 years				
PC-yields	11.836					PC-yields	18.670				
FB	0.960	0.943	0.884	0.905	0.709	FB	0.941	0.872	0.875	0.900	0.707
CP	1.101	1.037	0.863	0.943	0.708*	CP	1.099	1.025	0.862	0.941	0.738*
FB+CP	1.099	0.778	0.694	0.841	0.518*	FB+CP	1.075	0.751	0.693	0.861	0.535*

Notes: See section 5.4 for the definition of the predictors FB and CP. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 7 (width 0.1). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the benchmark results (from PCs of the yields) are MSFEs in level and the other entries present the ratios of MSFEs relative to the benchmark. Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 10: Out-of-sample forecasting performance on bond returns: Canada

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	Canada - 2 years						Canada - 3 years				
PC-yields	1.171					PC-yields	3.534				
FB	1.011	0.920	0.953	0.826	0.726	FB	1.029	0.868	0.905	0.859	0.706
CP	1.051	0.888	0.908	0.809	0.744	CP	1.094	0.907	0.898	0.852	0.757
FB+CP	1.034	0.861	0.931	0.798	0.687	FB+CP	1.096	0.813	0.852	0.826	0.642
	Canada - 4 years						Canada - 5 years				
PC-yields	6.545					PC-yields	10.133				
FB	1.033	0.860	0.887	0.892	0.707	FB	1.032	0.873	0.899	0.929	0.730
CP	1.129	0.911	0.859	0.882	0.758	CP	1.165	0.931	0.864	0.914	0.781
FB+CP	1.137	0.822	0.847	0.861	0.661	FB+CP	1.149	0.843	0.867	0.895	0.682

Notes: See section 5.4 for the definition of the predictors FB and CP. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 7 (width 0.1). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the benchmark results (from PCs of the yields) are MSFEs in level and the other entries present the ratios of MSFEs relative to the benchmark. Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 11: Out-of-sample forecasting performance on bond returns: UK

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	UK - 2 years						UK - 3 years				
PC-yields	1.415					PC-yields	4.378				
FB	0.807	0.821	0.907	0.790	0.648	FB	0.897	0.897	1.057	0.866	0.769
CP	0.923	0.764	0.704	0.646	0.593	CP	1.041	0.839	0.769	0.729	0.650
FB+CP	0.921	0.688	0.669	0.645	0.514	FB+CP	1.050	0.751	0.745	0.724	0.591
	UK - 4 years						UK - 5 years				
PC-yields	8.224					PC-yields	12.962				
FB	0.949	0.942	1.042	0.897	0.884	FB	0.980	0.983	1.028	0.923	0.936
CP	1.087	0.884	0.811	0.782	0.691*	CP	1.097	0.916	0.850	0.813	0.727
FB+CP	1.092	0.797	0.780	0.770	0.638	FB+CP	1.075	0.835	0.811	0.789	0.669

Notes: See section 5.4 for the definition of the predictors FB and CP. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 7 (width 0.1). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the benchmark results (from PCs of the yields) are MSFEs in level and the other entries present the ratios of MSFEs relative to the benchmark. Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Table 12: Out-of-sample forecasting performance on bond returns: Japan

	Non-local	$R = 60$	Opt-R	Opt-G	Opt-E		Non-local	$R = 60$	Opt-R	Opt-G	Opt-E
	Japan - 2 years						Japan - 3 years				
PC-yields	0.333					PC-yields	1.146				
FB	0.222	0.105*	0.115*	0.099*	0.097*	FB	0.244	0.148*	0.167*	0.145*	0.150*
CP	0.582	0.102*	0.112*	0.093*	0.098*	CP	0.677	0.155*	0.164*	0.140*	0.144*
FB+CP	0.610	0.101*	0.134*	0.091*	0.094*	FB+CP	0.679	0.149*	0.179*	0.140*	0.141*
	Japan - 4 years						Japan - 5 years				
PC-yields	2.517					PC-yields	4.050				
FB	0.246	0.197*	0.186*	0.186*	0.165*	FB	0.291	0.267*	0.243*	0.247*	0.219*
CP	0.817	0.181*	0.182*	0.162*	0.160*	CP	0.902	0.220*	0.223*	0.196*	0.190*
FB+CP	0.772	0.186*	0.189*	0.162*	0.168*	FB+CP	0.871	0.182*	0.187*	0.167*	0.169*

Notes: See section 5.4 for the definition of the predictors FB and CP. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 7 (width 0.1). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. For each specification, the benchmark results (from PCs of the yields) are MSFEs in level and the other entries present the ratios of MSFEs relative to the benchmark. Cyan shading indicates the best performing specification within each variable. Gray shading indicates the overall best performing case. To provide a rough gauge of whether the two forecasts have significantly different accuracy, we use a Diebold-Mariano t-statistic with fixed smoothing asymptotics as in Coroneo and Iacone (2020). Differences in accuracy that are statistically different from zero (using either fixed b-smoothing or fixed m-smoothing asymptotics) are denoted by an asterisk, corresponding to the 5 percent significance level.

Data appendix

Table 13: Data description and variable transformation: USA, Sec. 5.1

Asset prices		
FEDFUNDS	Effective Federal Funds Rate (Percent)	y_t
TB3MS	3-Month Treasury Bill: Secondary Market Rate (Percent)	y_t
GS10	10-Year Treasury Constant Maturity Rate (Percent)	y_t
GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	y_t
termspread	10-Year Treasury Constant Maturity Minus Effective Federal Funds Rate (Percent)	y_t
S&P 500	S&P's Common Stock Price Index: Composite	$100\Delta \ln y_t$
VXOCLSx	CBOE S&P 100 Volatility Index: VXO	y_t
Real economic activity		
DPIC96	Real Disposable Personal Income (Billions of Chained 2012 Dollars)	$100\Delta \ln y_t$
GPDIC1	Real Gross Private Domestic Investment, 3 decimal (Billions of Chained 2012 Dollars)	$100\Delta \ln y_t$
INDPRO	Industrial Production Index (Index 2012=100)	$100\Delta \ln y_t$
CE16OV	Civilian Employment (Thousands of Persons)	$100\Delta \ln y_t$
UNRATE	Civilian Unemployment Rate (Percent)	y_t
LNS14000026	Unemployment Rate - 20 years and over, Women (Percent)	y_t
HOUST	Housing Starts: Total: New Privately Owned Housing Units Started (Thousands of Units)	$100\Delta \ln y_t$
PERMIT	New Private Housing Units Authorized by Building Permits (Thousands of Units)	$100\Delta \ln y_t$
Price indices		
CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items (Index 1982-84=100)	$100\Delta \ln y_t$
CPIAPPSL	Consumer Price Index for All Urban Consumers: Apparel (Index 1982-84=100)	$100\Delta \ln y_t$
CPIENGSL	Consumer Price Index for All Urban Consumers: Energy (Index 1982-84=100)	$100\Delta \ln y_t$
PPIACO	Producer Price Index for All Commodities (Index 1982=100)	$100\Delta \ln y_t$
PCECTPI	Personal Consumption Expenditures: Chain-type Price Index (Index 2012=100)	$100\Delta \ln y_t$
Monetary measures		
BOGMBASEREALx	Monetary Base (Millions of 1982-84 Dollars), deflated by CPI	$100\Delta \ln y_t$
M1REAL	Real M1 Money Stock (Billions of 1982-84 Dollars), deflated by CPI	$100\Delta \ln y_t$
M2REAL	Real M2 Money Stock (Billions of 1982-84 Dollars), deflated by CPI	$100\Delta \ln y_t$

Notes: The data are taken from FRED-QD (McCracken et al. (2021)) and the mnemonics are exactly as in FRED-QD.

Table 14: Data description and variable transformation: Canada, Sec. 5.1

Asset prices		
BANK_RATE_L	Bank rate	y_t
TBILL_3M	Treasury bills (3 months)	y_t
GOV_AVG_10pY	Governmental bonds (average rate) (10+ years)	y_t
G_AVG_5.10.Bank_rate	Government bonds (5-10 years) - Bank rate	y_t
G_AVG_10p.TBILL_3M	Government Bonds (10+ years) - Treasury Bond (3 months)	y_t
TSX_CLO	Toronto Stock Exchange (close)	$100\Delta \ln y_t$
Real economic activity		
REAL_GDP	Real Gross domestic product , chained (2012) dollars	$100\Delta \ln y_t$
hhold_dispo_income	Households disposable income	$100\Delta \ln y_t$
REAL_I	Real Gross fixed capital formation, chained (2012) dollars	$100\Delta \ln y_t$
CANPROINDQISMEI*	Production: Industry: Total Industry Excluding Construction for Canada	$100\Delta \ln y_t$
LFEMTTTTCAQ647S*	Employed Population: Aged 15 and over: All Persons for Canada	$100\Delta \ln y_t$
UNEMP_CAN	Unemployment rate	y_t
hstart_CAN	House Starts (units)	$100\Delta \ln y_t$
Price indices		
CPI_ALL_CAN	Consumption price index (CPI) (all)	$100\Delta \ln y_t$
IPPI_CAN	Industrial production price index (IPPI) (all)	$100\Delta \ln y_t$
C_PRICE	Implicit price index : Final consumption expenditure, 2012 = 100	$100\Delta \ln y_t$
Monetary measures		
MBASE1	Monetary base	$100\Delta \ln y_t$
CRED_BUS_cb	Business loans, Chartered banks only	$100\Delta \ln y_t$
CRED_HOUS_cb	Personal loans, Chartered banks only	$100\Delta \ln y_t$

Notes: The mnemonics with an asterisk indicate that the variables are taken from Federal Reserve Economic Data (FRED). All the other variables are taken from Stevanovic et al. (2021).

Table 15: Data description: Sec. 5.2

Variable	Data source	Variable description
CatFin	Allen et al. (2012)	measure of aggregate systemic risk
Default spread	FRED	difference between yields on BAA and AAA corporate bonds
TED spread	GFD	difference between 3-month LIBOR and 3-month T-bill interest rates
Term spread	GFD	difference between yields on the ten year and the 3-month treasury bond
Slope factor	Liu and Wu (2021)	slope factor of the yield curve (1-120 month)
VIX	FRED	Chicago Board Options Exchange's CBOE volatility index
Stock return	FRED	S&P500 composite index return

Notes: FRED refers to Federal Reserve Economic Data. GFD refers to Global Financial Database.

Table 16: Data description and variable transformation: Sec. 5.3

Variable	Data source	Variable description	Transformation
		United States	
stock return	CRSP	S&P500 value-weighted index return	y_t
treasury bill	FRED	3-month treasury bill: secondary market rate	y_t
spread	GFD	differences between 5-year government bond yield and 3-month treasury bill	y_t
inflation	FRED	changes of Consumer price index for all urban consumers: all items in U.S. city average	$100\Delta \ln(y_t)$
growth	FRED	changes of Industrial production: total index	$100\Delta \ln(y_t)$
		Canada	
stock return	GFD	S&P/TSX-300 total return index	$100\Delta \ln(y_t)$
treasury bill	GFD	3-month treasury bill yield	y_t
spread	GFD	differences between 10-year government bond yield and 3-month treasury bill	y_t
inflation	FRED	changes of Consumer price index: all items: city: total	$100\Delta \ln(y_t)$
growth	FRED	changes of Production: industry: total industry: total industry excluding construction	$100\Delta \ln(y_t)$
		France	
stock return	GFD	CAC all-tradable total return index	$100\Delta \ln(y_t)$
treasury bill	GFD	3-month treasury bill yield	y_t
spread	GFD	differences between 10-year government bond yield and 3-month treasury bill	y_t
inflation	FRED	changes of Consumer price index of all items	$100\Delta \ln(y_t)$
growth	FRED	changes of Production of total industry	$100\Delta \ln(y_t)$
		Germany	
stock return	GFD	CDAX total return index	$100\Delta \ln(y_t)$
treasury bill	GFD	3-month treasury bill yield	y_t
spread	GFD	differences between 5-year government bond yield and 3-month treasury bill	y_t
inflation	FRED	changes of Consumer price index: all items: total	$100\Delta \ln(y_t)$
growth	FRED	changes of Production: industry: total industry: total industry excluding construction	$100\Delta \ln(y_t)$
		Australia	
stock return	GFD	ASX accumulation index-all ordinaries	$100\Delta \ln(y_t)$
treasury bill	GFD	3-month treasury bill yield	y_t
spread	GFD	differences between 10-year government bond yield and 3-month treasury bill	y_t
inflation	FRED	changes of Consumer price index: all items: total	$100\Delta \ln(y_t)$
growth	FRED	changes of Production: industry: total industry: total industry excluding construction	$100\Delta \ln(y_t)$

Notes: CRSP refers to Center for Research in Security Prices. FRED refers to the database maintained by the Federal Reserve Bank of St.Louis. GFD refers to the Global Financial Database. For Australia, CPI and industrial production are only available at a quarterly frequency.

Table 17: Data description: Sec. 5.4

Country	Data source	Sample period	N_i
United States	Liu and Wu (2021)	1961M6-2022M12	60
Canada	Bank of Canada	1986M1-2022M12	20
United Kingdom	Bank of England	1970M1-2022M12	9
Japan	Ministry of Finance	1980M8-2022M12	5

Notes: N_i is the number of variables (different maturities, up to 5 years) available in each country.

Appendix A. Technical assumptions

Assumption A1. *(Loss function)*

(i) The loss function $\ell_t(\theta)$ is measurable and three-times continuously differentiable with respect to θ;

(ii) $\theta_{0,t} = \theta(t/T)$ is a $\bar{k} \times 1$ vector of true time-varying parameters, $\theta_{0,t} \in \Theta$, Θ is compact;

(iii) $\left\{\frac{\partial \ell_{t+1}(\theta_{0,t})}{\partial \theta'}\right\}_t$ is a martingale difference sequence (M.D.S.) with respect to $\mathcal{F}_t$: $E(\frac{\partial \ell_{t+1}(\theta_{0,t})}{\partial \theta'}|\mathcal{F}_t) = 0$, where $\mathcal{F}_t = \sigma(y_s, X_s, s \leqslant t)$;

(iv) $E[\ell_T(\theta)]$ is uniquely minimized at $\theta(1)$ and

$$\sup_{\theta\in\Theta} \left| \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \ell_t(\theta) - E[\ell_T(\theta)] \right| = o_p(1).$$

Assumption A2. *(CLT and WLLN)*

(i) The following central limit theorem (CLT) holds:

$$\frac{1}{\sqrt{Tb}} \sum_{t=1}^{T} k_{tT} \frac{\partial \ell_t(\theta_{0,t})}{\partial \theta'} \xrightarrow{d} \mathcal{N}(0, \phi_{0,K}\Lambda(1)),$$

where $\phi_{0,K} = \int_C K^2(u)du$ and $\Lambda(1) = Var\left(\frac{\partial \ell_T(\theta(1))}{\partial \theta'}\right)$.

(ii) The following weak law of large numbers (WLLN) holds:

$$\sup_{\theta\in\Theta} \left| \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \left(\frac{\partial^2 \ell_t(\theta)}{\partial \theta_{j_1} \partial \theta_{j_2}} - E\Big[\frac{\partial^2 \ell_t(\theta)}{\partial \theta_{j_1} \partial \theta_{j_2}} \Big] \right) \right| = O_p((Tb)^{-1/2}),$$
$$\sup_{\theta\in\Theta} \left| \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \left(\frac{\partial^2 \ell_t(\theta)}{\partial \theta_{j_1} \partial \theta_{j_2}} - E\Big[\frac{\partial^2 \ell_t(\theta)}{\partial \theta_{j_1} \partial \theta_{j_2}} \Big] \right) \Big(\frac{t-T}{T} \Big)^{\delta} \right| = O_p(b^{\delta}),$$

for any $j_1 = 1, 2, \cdots, \bar{k}$, $j_2 = 1, 2, \cdots, \bar{k}$, where b is a tuning parameter such that $b \to 0$ as $T \to \infty$ and $0 < \delta \leqslant 3$;

(iii) Let $\tilde{H}(u) = E\Big[\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'}\Big]$ be defined on $(0, 1]$. The smallest eigenvalues are uniformly bounded away from 0 over $\theta \in \Theta$.

Assumption A3. *$\theta_t = \theta(t/T)$, $\theta(\cdot) : (0, 1] \longrightarrow \Theta$. Let $\theta_\ell(t/T)$ ($\ell = 1, 2, \cdots, \bar{k}$) be the ℓth elements in $\theta(t/T)$.*

(i) $\theta_\ell(t/T)$ *satisfies the following*

$$|\theta_\ell(t/T) - \theta_\ell(s/T)| \leqslant c_\ell \Big(\frac{|t-s|}{T}\Big)^\gamma, \quad t, s = 1, 2, \cdots, T,$$

where $0 < \gamma \leqslant 1$ *and* c_ℓ *is a positive constant satisfying* $\max_\ell |c_\ell| < \infty$.

(ii) $\theta(\cdot)$ *is twice continuously differentiable on* $(0, 1]$.

Assumption A4. $K(u) \geqslant 0$, $u \in \mathbb{R}$ *is a continuous bounded function and* $\int K(u)du = 1$.

Assumption A5. *The tuning parameters* b *and* $\tilde{b}$ *are such that: (i)* $T\tilde{b}^5 \to 0$*; (ii)* $b/\tilde{b} \to 0$*; (iii)* $T^{1/2}\tilde{b}^{1/2}b^\gamma \to \infty$ *for some* $0 < \gamma \leqslant 1$.

Assumption A1(i) impose conditions on the shape of the loss function. Assumption A1(iii) impose martingale difference assumption on the score of the loss. As explained in section 4, for a linear model, this implies that the forecast error $\{\varepsilon_{t+1}\}_t$ is a M.D.S., so sufficient number of lags of y_y has to be included in the model to ensure that the error term ε_t is serially uncorrelated. Assumption A1(ii) and A1(iv) are needed so that Theorem 2.1 in Newey and McFadden (1994) can be applied to ensure consistency of the estimator.

Assumptions A2 is a high level assumption for the results presented in Lemmas B1-B3. While it is straightforward to verify these results under the assumption that $\{\ell_t(\cdot)\}$ is stationary, this may have undesirable features under parameter instability. Recently, Kristensen and Lee (2023) derive these results under the assumption that $\{\ell_t(\cdot)\}$ is locally stationary with additional requirements on $\{\ell_t(\cdot)\}$. We refer the interesting readers to their paper for the details.

Assumption A3 impose conditions on the time-varying parameters. While A3(i) is more general than A4(ii) and is sufficient for the consistency of the local estimator, for the asymptotic optimality of the tuning parameter selection, we do require differentiability. However, as explained in section 3.1, this condition is not restrictive as the cases considered in Giraitis et al. (2014) and Dendramis et al. (2021) are included. Assumption A4 is a standard condition for the weighting function in the literature. Assumption A5 is a condition for the two tuning parameters which again ensures the asymptotic optimality of the tuning parameter selection procedure.

Appendix B. Auxiliary results

Lemma B1. *Suppose that Assumptions A1, A2, A3(i), A4 hold with* $b \to 0$ *and* $Tb \to \infty$. *Then, it holds that*

(i) Consistency: $\hat{\theta}_{K,b,T} \xrightarrow{p} \theta(1)$ *and*

$$\left\|\hat{\theta}_{K,b,T} - \theta(1)\right\| = O_p((Tb)^{-1/2} + b^\gamma),$$

for some $0 < \gamma \leqslant 1$.

(ii) CLT: if $T^{1/2}b^{1/2+\gamma} \to 0$*, we have*

$$\sqrt{Tb}\left(\hat{\theta}_{K,b,T} - \theta(1)\right) \xrightarrow{d} \mathcal{N}(0, \phi_{0,K}\Sigma(1)),$$

where $\Sigma(1) = H^{-1}(1)\Lambda(1)H^{-1}$, $\phi_{0,K} = \int K^2(u)du$, $\Lambda(1) = Var\left(\frac{\partial \ell_T(\theta(1))}{\partial \theta'}\right)$ *and* $H(1) = E\left[\frac{\partial^2 \ell_T(\theta(1))}{\partial \theta \partial \theta'}\right]$.

Proof. The estimator $\hat{\theta}_{K,b,T}$ is defined as the minimizer of the objective function:

$$\hat{\theta}_{K,b,T} = \arg\min_{\theta \in \Theta} \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \ell_t(\theta), \tag{B.1}$$

where $\ell_t(\theta) = \ell(y_t, \hat{y}_{t|t-1}(\theta))$.

We first prove the consistency of the estimator. Let $L_T(\theta) = \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \ell_t(\theta)$ and $L_0(\theta) = E[\ell_T(\theta)]$. In view of Theorem 2.1 in Newey and McFadden (1994), it is sufficient to verify that

(i) $L_0(\theta)$ is uniquely minimized at $\theta(1)$;

(ii) Θ is compact;

(iii) $L_0(\theta)$ is continuous;

(iv) Uniform weak law of large numbers (UWLLN):

$$\sup_{\theta \in \Theta} \left| \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \ell_t(\theta) - E[\ell_T(\theta)] \right| = o_p(1).$$

These conditions are all maintained in Assumption A1, so the consistency follows accordingly.

We now move on to the derivation of the consistency rate and the CLT. By a Taylor series expansion of $\frac{\partial L_T(\hat{\theta}_{K,b,T})}{\partial \theta} = 0$ around the true value $\theta(1)$, we have

$$\frac{\partial L_T(\theta(1))}{\partial \theta} + \frac{\partial^2 L_T(\bar{\theta}(1))}{\partial \theta \partial \theta'}(\hat{\theta}_{K,b,T} - \theta(1)) = 0,$$

where $\overline{\theta}(1)$ lies between $\theta(1)$ and $\hat{\theta}_{K,b,T}$. By rearranging terms, we have

$$\begin{aligned}
\hat{\theta}_{K,b,T} - \theta(1) &= -\Big(\frac{\partial^2 L_T(\overline{\theta}(1))}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{\partial L_T(\theta(1))}{\partial\theta}\Big) \\
&= -\Big(\frac{\partial^2 L_T(\theta(1))}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{\partial L_T(\theta(1))}{\partial\theta}\Big) + \left[\Big(\frac{\partial^2 L_T(\theta(1))}{\partial\theta\partial\theta'}\Big)^{-1} - \Big(\frac{\partial^2 L_T(\overline{\theta}(1))}{\partial\theta\partial\theta'}\Big)^{-1}\right]\frac{\partial L_T(\theta(1))}{\partial\theta} \\
&= -\Big(\frac{\partial^2 L_T(\theta(1))}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{\partial L_T(\theta(1))}{\partial\theta}\Big) + \Big(\frac{\partial^2 L_T(\theta(1))}{\partial\theta\partial\theta'}\Big)^{-1}\left[\frac{\partial^2 L_T(\overline{\theta}(1))}{\partial\theta\partial\theta'} - \frac{\partial^2 L_T(\theta(1))}{\partial\theta\partial\theta'}\right] \\
&\qquad\qquad \times \Big(\frac{\partial^2 L_T(\overline{\theta}(1))}{\partial\theta\partial\theta'}\Big)^{-1}\frac{\partial L_T(\theta(1))}{\partial\theta}, \\
&= -H_{1,T}^{-1}(\theta(1))S_T(\theta(1)) + H_{1,T}^{-1}(\theta(1))\Big[H_{1,T}(\overline{\theta}(1)) - H_{1,T}(\theta(1))\Big]H_{1,T}^{-1}(\overline{\theta}(1))S_T(\theta(1)) \qquad \text{(B.2)}
\end{aligned}$$

where

$$H_{1,T}(\theta) = \frac{\partial^2 L_T(\theta)}{\partial\theta\partial\theta'} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'}, \quad S_T(\theta) = \frac{\partial L_T(\theta)}{\partial\theta} = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial\ell_t(\theta)}{\partial\theta}.$$

We will show that

$$\big\|S_T(\theta(1))\big\| = O_p((Tb)^{-1/2} + b^{\gamma}), \qquad \text{(B.3)}$$

$$\big\|H_{1,T}^{-1}(\theta(1))\big\| = O_p(1), \qquad \text{(B.4)}$$

$$\big\|H_{1,T}(\overline{\theta}(1)) - H_{1,T}(\theta(1))\big\| = o_p(1). \qquad \text{(B.5)}$$

These bounds together with (B.2) implies the consistency rate in B1(i).

Proof of (B.3). We have that

$$\begin{aligned}
S_T(\theta(1)) = \frac{\partial L_T(\theta(1))}{\partial\theta} &= \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial\ell_t(\theta(1))}{\partial\theta} \\
&= \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial\ell_t(\theta_{0,t})}{\partial\theta} + \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2\ell_t(\overline{\theta}(1))}{\partial\theta\partial\theta'}(\theta(1) - \theta_{0,t}) \\
&= S_{1,T} + S_{2,T}(\overline{\theta}(1)),
\end{aligned}$$

where the second line follows from Taylor series expansion and $\overline{\theta}(1)$ lies between $\theta(1)$ and $\theta_{0,t}$. By Assumption A2, we have $\big\|S_{1,T}\big\| = O_p(\frac{1}{\sqrt{Tb}})$. For $S_{2,T}(\overline{\theta}(1))$, let $\overline{\theta}(1) \to \theta(1)$, we have $S_{2,T}(\overline{\theta}(1)) \xrightarrow{p} S_{2,T}(\theta(1))$. It follows from Assumption A3(i) that

$$\big\|S_{2,T}(\theta_{0,t})\big\| \leqslant C\frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big(\frac{|t-T|}{T}\Big)^{\gamma}.$$

Let $\overline{S}_{2,T}(\theta_{0,t}) = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\left(\frac{|t-T|}{T}\right)^{\gamma}$. Since

$$\begin{aligned}
\overline{S}_{2,T}(\theta(1)) &= \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'} - E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big] + E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]\Big]\Big(\frac{|t-T|}{T}\Big)^{\gamma} \\
&= \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'} - E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]\Big]\Big(\frac{|t-T|}{T}\Big)^{\gamma} + \frac{1}{Tb}\sum_{t=1}^{T} k_{tT} E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]\Big(\frac{|t-T|}{T}\Big)^{\gamma} \\
&= \overline{S}^{(1)}_{2,T}(\theta(1)) + \overline{S}^{(2)}_{2,T}(\theta(1)).
\end{aligned}$$

By Assumption A2(ii), we have $\left\|\overline{S}^{(1)}_{2,T}(\theta(1))\right\| = O_p(b^{\gamma})$. For $\overline{S}^{(2)}_{2,T}(\theta(1))$, we have

$$\left\|\overline{S}^{(2)}_{2,T}(\theta(1))\right\| \leqslant C\,\frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\Big(\frac{|t-T|}{T}\Big)^{\gamma} \approx b^{\gamma}\int_{C} u^{\gamma}K(u)du = O(b^{\gamma}),$$

where the approximation follows from Riemann sum approximation of an integral. By triangular inequality, $\left\|S_{2,T}\right\| = O_p(b^{\gamma})$. Then, (B.3) follows again from triangular inequality.

Proof of (B.4). Write

$$\begin{aligned}
H_{1,T}(\theta(1)) &= \frac{1}{Tb}\sum_{t=1}^{T} k_{tT} E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big] + \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\left(\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'} - E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]\right) \\
&= H^{(1)}_{1,T} + H^{(2)}_{1,T} = H^{(1)}_{1,T}(I_k + \tilde{\Delta}_T),
\end{aligned} \tag{B.6}$$

where $\tilde{\Delta}_T = (H^{(1)}_{1,T})^{-1}(H_{1,T} - H^{(1)}_{1,T})$. By Assumption A2(iii), there exists $v > 0$ such that for all $t \geqslant 1$,

$$a' E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]a \geqslant 1/v > 0.$$

Thus, we have, for any $k \times 1$ vector $a = (a_1, \cdots, a_k)'$ such that $\|a\|^2 = 1$

$$\min_{\|a\|=1} a' H^{(1)}_{1,T} a = \min_{\|a\|=1}\Big(\frac{1}{Tb}\sum_{t=1}^{T} k_{tT} a' E\Big[\frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]a\Big) \geqslant \frac{1}{v}\Big(\frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\Big) > 0.$$

This means that the smallest eigenvalue of $H^{(1)}_{1,T}$ is not smaller than $1/v > 0$, which further implies that

$$\left\|(H^{(1)}_{1,T})^{-1}\right\|_{sp} = O_p(1).$$

In addition, by Assumption A2(ii), we have

$$\left\|H_{1,T} - H^{(1)}_{1,T}\right\|_{sp} = o_p(1).$$

Then,

$$\left\|H_{1,T}^{-1}(\theta(1))\right\|_{sp} \leqslant \left\|(H_{1,T}^{(1)})^{-1}\right\|_{sp} (1 - \left\|H_{1,T} - H_{1,T}^{(1)}\right\|_{sp})^{-1} = O_p(1).$$

Proof of (B.5). This follow immediately by the consistency: $\hat{\theta}_{K,b,T} \xrightarrow{p} \theta(1)$.

Back to (B.2), we have

$$\sqrt{Tb}\left(\hat{\theta}_{K,b,T} - \theta(1)\right) = -H_{1,T}^{-1}(\theta(1))\ \sqrt{Tb}(S_{1,T}(\theta(1)) + S_{2,T}(\theta(1)))$$

Since $\left\|\sqrt{Tb}S_{1,T}(\theta(1))\right\| = O_p(1)$, $\left\|\sqrt{Tb}S_{2,T}(\theta(1))\right\| = O_p(T^{1/2}b^{1/2+\gamma})$, under the condition $T^{1/2}b^{1/2+\gamma} \to 0$, the dominating term is the first one, by applying CLT on $\sqrt{Tb}S_{1,T}(\theta(1))$, together with Slutsky's theorem, we obtain

$$\sqrt{Tb}\left(\hat{\theta}_{K,b,T} - \theta(1)\right) \xrightarrow{d} \mathcal{N}(0, \phi_{0,K}\Sigma(1)),$$

where $\Sigma(1) = H^{-1}(1)\Lambda(1)H^{-1}$, $H(1) = E\left[\frac{\partial^2 \ell_T(\theta(1))}{\partial\theta\partial\theta'}\right]$ and $\Lambda(1) = Var\left(\frac{\partial \ell_T(\theta(1))}{\partial\theta'}\right)$. □

Lemma B2. *Suppose that Assumptions A1, A2, A3(ii), A4 hold with $\tilde{b} \to 0$ and $T\tilde{b} \to \infty$. Then, it holds that*

$$\left\|\tilde{\theta}_T - \theta(1)\right\| = O_p((T\tilde{b})^{-1/2} + \tilde{b}^2).$$

Proof. The objective function is given by

$$L_T(\beta) = \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT}\ell_t\Big(D(u)\beta\Big),$$

where $D(u) = [1\ \frac{t-T}{T}]$, $\beta = (\beta_1', \beta_2')'$, $\beta_1 = \theta(1)$ and $\beta_2 = \theta^{(1)}(1)$. Similarly as in (B.2), we have that

$$\begin{pmatrix} \tilde{\theta}_T(1) - \theta(1) \\ \tilde{\theta}_T^{(1)}(1) - \theta'(1) \end{pmatrix} = -\Big(\frac{\partial L_T^2(\beta)}{\partial\beta\partial\beta'}\Big)^{-1} \frac{\partial L_T(\beta)}{\partial\beta} + o_p(1). \tag{B.7}$$

Notice that

$$\frac{\partial L_T^2(\beta)}{\partial\beta\partial\beta'} = \begin{bmatrix} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t\Big(D(u)\beta\Big)}{\partial\beta_1\partial\beta_1'} & \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t\Big(D(u)\beta\Big)}{\partial\beta_1\partial\beta_2'}\Big(\frac{t-T}{T}\Big) \\ \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t\Big(D(u)\beta\Big)}{\partial\beta_2\partial\beta_1'}\Big(\frac{t-T}{T}\Big) & \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t\Big(D(u)\beta\Big)}{\partial\beta_2\partial\beta_2'}\Big(\frac{t-T}{T}\Big)^2 \end{bmatrix}.$$

By Assumption A3(ii), together with property of the inverse of the partitioned matrices (see, Abadir and Magnus (2005)), we have

$$\Big(\frac{\partial L_T^2(\beta(1))}{\partial\beta\partial\beta'}\Big)^{-1} = \begin{bmatrix} O_p(1) & O_p(\tilde{b}) \\ O_p(\tilde{b}) & O_p(\tilde{b}^2) \end{bmatrix}.$$

Next, notice that

$$\frac{\partial L_T(\beta)}{\partial \beta} = \begin{bmatrix} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t\left(D(u)\beta\right)}{\partial \beta_1} \\ \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t\left(D(u)\beta\right)}{\partial \beta_2}\left(\frac{t-T}{T}\right) \end{bmatrix}. \tag{B.8}$$

By Assumption A3(ii), we have that

$$\theta(t/T) \approx \theta(1) + \theta'(1)\Big(\frac{t-T}{T}\Big) + \frac{\theta''(1)}{2}\Big(\frac{t-T}{T}\Big)^2.$$

Then, for the first $k \times 1$ elements in (B.8), Taylor expansion around $\theta(t/T)$ gives

$$\begin{aligned} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t(D(u)\beta)}{\partial \beta_1} &= \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t(\theta_{0,t})}{\partial \beta_1} + \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t(\overline{\theta}(1))}{\partial \beta_1 \partial \beta_1'} \Big(\theta(1) + \theta^{(1)}(1)\Big(\frac{t-T}{T}\Big) - \theta(t/T)\Big) \\ &= \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t(\theta_{0,t})}{\partial \beta_1} + \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t(\overline{\theta}(1))}{\partial \beta_1 \partial \beta_1'} \frac{\theta^{(2)}(1)}{2} \Big(\frac{t-T}{T}\Big)^2, \end{aligned}$$

where $\overline{\theta}(1)$ lies between $\theta(t/T)$ and $\theta(1) + \theta^{(1)}(1)\Big(\frac{t-T}{T}\Big)$. Similarly, we have

$$\frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t(D(u)\beta)}{\partial \beta_2}\Big(\frac{t-T}{T}\Big) = \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t(\theta_{0,t})}{\partial \beta_2}\Big(\frac{t-T}{T}\Big) + \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t(\overline{\theta}(1))}{\partial \beta_2 \partial \beta_2'} \frac{\theta^{(2)}(1)}{2} \Big(\frac{t-T}{T}\Big)^3.$$

Now, back to (B.7), we have

$$\begin{pmatrix} \tilde{\theta}_T(1) - \theta(1) \\ \tilde{\theta}_T'(1) - \theta^{(1)}(1) \end{pmatrix} = -\Big(\frac{\partial L_T^2(\beta)}{\partial \beta \partial \beta'}\Big)^{-1} \underbrace{\begin{bmatrix} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t\left(\theta_{0,t}\right)}{\partial \beta_1} \\ \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial \ell_t\left(\theta_{0,t}\right)}{\partial \beta_2}\left(\frac{t-T}{T}\right) \end{bmatrix}}_{Q_{1,T}} - \Big(\frac{\partial L_T^2(\beta(1))}{\partial \beta \partial \beta'}\Big)^{-1} \underbrace{\begin{bmatrix} \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t\left(\overline{\theta}(1)\right)}{\partial \beta_1 \partial \beta_1'} \frac{\theta^{(2)}(1)}{2}\left(\frac{t-T}{T}\right)^2 \\ \frac{1}{T\tilde{b}} \sum_{t=1}^{T} \tilde{k}_{tT} \frac{\partial^2 \ell_t\left(\overline{\theta}(1)\right)}{\partial \beta_2 \partial \beta_2'} \frac{\theta^{(2)}(1)}{2}\left(\frac{t-T}{T}\right)^3 \end{bmatrix}}_{Q_{2,T}}.$$

By Assumption A2, we have

$$Q_{1,T} = \begin{bmatrix} O_p((T\tilde{b})^{-1/2}) \\ O_p((T\tilde{b})^{-1/2}\tilde{b}) \end{bmatrix}, \quad Q_{2,T} = \begin{bmatrix} O_p(\tilde{b}^2) \\ O_p(\tilde{b}^3) \end{bmatrix}.$$

Therefore, we obtain the consistency rate for $\tilde{\theta}_T$:

$$\left\|\tilde{\theta}_T - \theta(1)\right\| = O_p((T\tilde{b})^{-1/2} + \tilde{b}^2).$$

□

Lemma B3. *Suppose that Assumptions A1, A2, A3, A4 hold with $b \to 0$ and $Tb \to \infty$. Then, for some*

$0 < \delta < \frac{1}{2}$ *and* $0 < \gamma \leqslant 1$*, it holds that*

$$\sup_{b \in I_T} \left\| \hat{\theta}_{\overline{K},b,T} - \theta_T \right\| = O_p(r_{T,b,\delta,\gamma}), \tag{B.9}$$

where $r_{T,b,\delta,\gamma} = T^{-1/2} b^{-1/2+\delta} + b^{\gamma}$.

Proof. Write $\hat{\theta}_{\overline{K},b,T} = \hat{\theta}_{b,T}$. As in (B.2), the estimator can be decomposed as

$$\begin{aligned} \hat{\theta}_{b,T} - \theta_T &= -H_{1,T} S_T + o_p(1) \\ &= -H_{1,T}(S_{1,T} + S_{2,T}) + o_p(1), \end{aligned} \tag{B.10}$$

where

$$\begin{aligned} H_{1,T} &= \left(\frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\theta(1))}{\partial \theta \partial \theta'} \right)^{-1}, \\ S_{1,T} &= \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial \ell_t(\theta(t/T))}{\partial \theta}, \quad S_{2,T} = \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\overline{\theta}(1))}{\partial \theta \partial \theta'} (\theta(1) - \theta(t/T)), \end{aligned}$$

and $\overline{\theta}(1)$ lies between $\theta(1)$ and $\theta(t/T)$. We will show that

$$\sup_{b \in I_T} \left\| \frac{1}{T^{1/2} b^{1/2-\delta}} \sum_{t=1}^{T} k_{tT} \frac{\partial \ell_t(\theta(1))}{\partial \theta} \right\| = O_p(1), \quad \text{for } 0 < \delta < 1/2, \tag{B.11}$$

$$\sup_{b \in I_T} \left\| \left(\frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\theta(1))}{\partial \theta \partial \theta'} \right)^{-1} \right\| = O_p(1), \tag{B.12}$$

$$\sup_{b \in I_T} \left\| \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\overline{\theta}(1))}{\partial \theta \partial \theta'} (\theta(1) - \theta(t/T)) \right\| = O_p(b^{\gamma}) \text{ for } 0 < \gamma \leqslant 1 \tag{B.13}$$

These bounds together with (B.10) prove (B.9).

Proof of (B.11). By Boole's inequality and Chebyshev's inequality, we have, for any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{b\in I_T}\left\|\frac{1}{T^{1/2}b^{1/2-\delta}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_t(\theta(1))}{\partial\theta}\right\| > \varepsilon\right) &\leqslant \sum_{b\in I_T}\mathbb{P}\left(\left\|\frac{1}{T^{1/2}b^{1/2-\delta}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_t(\theta(1))}{\partial\theta}\right\| > \varepsilon\right)\\
&\leqslant |I_T|\times\sup_{b\in I_T}\mathbb{P}\left(\left\|\frac{1}{T^{1/2}b^{1/2-\delta}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_t(\theta(1))}{\partial\theta}\right\| > \varepsilon\right)\\
&\leqslant |I_T|\times\sup_{b\in I_T}\frac{\left\|\mathrm{Var}\left(\frac{1}{T^{1/2}b^{1/2-\delta}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_t(\theta(1))}{\partial\theta}\right)\right\|}{\varepsilon^2}\\
&\leqslant |I_T|\times\sup_{b\in I_T}\frac{C}{b^{2\delta}\varepsilon^2} = O(1).
\end{aligned}$$

For the fourth inequality, observe that by Assumption A2(i), we have

$$\left\|\frac{1}{\sqrt{Tb}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_t(\theta(t/T))}{\partial\theta'}\right\| = O_p(1).$$

By Taylor expansion of the above sum around $\theta(1)$, we have $\left\|\frac{1}{\sqrt{Tb}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_t(\theta(1))}{\partial\theta'}\right\| = O_p(1)$. For the last equality, since consistency requires $b\to 0$, we can let the cardinality of the set $|I_T|$ to be of order $b^{2\delta}$ to obtain the results.

Proof of (B.12). As in (B.4), write

$$\begin{aligned}
\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'} &= \frac{1}{Tb}\sum_{t=1}^{T}k_{tT}E\Big[\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big] + \frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\left(\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'} - E\Big[\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]\right)\\
&= H_{1,T}^{(1)} + H_{1,T}^{(2)} = H_{1,T}^{(1)}(I_k + \tilde{\Delta}_T),
\end{aligned} \tag{B.14}$$

where $\tilde{\Delta}_T = (H_{1,T}^{(1)})^{-1}(H_{1,T} - H_{1,T}^{(1)})$. First, (B.4) holds uniformly over b:

$$\sup_{b\in I_T}\left\|(H_{1,T}^{(1)})^{-1}\right\|_{sp} = O_p(1). \tag{B.15}$$

For $\tilde{\Delta}_T$, let $\tilde{\Delta}_t = \frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'} - E\Big[\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]$. Then, for any $\varepsilon > 0$, , by Boole's inequality and Chebyshev's inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{b\in I_T}\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\| > \varepsilon\right) &\leqslant \sum_{b\in I_T}\mathbb{P}\left(\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\| > \varepsilon\right)\\
&\leqslant |I_T|\times\sup_{b\in I_T}\mathbb{P}\left(\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\| > \varepsilon\right).
\end{aligned}$$

Similarly as in the proof of (B.11), we have

$$\sup_{b\in I_T}\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\| = O_p(T^{-1/2}b^{-1/2+\delta}), \tag{B.16}$$

for some $0<\delta<1/2$. Since the cardinality of the set $|I_T|$ must be $o(1)$, we have $\sup_{b\in I_T}\left\|\tilde{\Delta}_T\right\|_{sp} = o_p(1)$. To sum up, we continue from (B.6):

$$\sup_{b\in I_T}\left\|H_{1,T}^{-1}\right\|_{sp} \leqslant \underbrace{\sup_{b\in I_T}\left\|(H_{1,T}^{(1)})^{-1}\right\|_{sp}}_{O_p(1)\text{ by (B.15)}}\Big(1-\underbrace{\sup_{b\in I_T}\left\|\tilde{\Delta}_T\right\|_{sp}}_{o_p(1)\text{ by (B.16)}}\Big)^{-1} = O_p(1).$$

Proof of (B.13). Let $\overline{\theta}(1)\to\theta(1)$ then $S_{2,T}(\overline{\theta}(1))\xrightarrow{p}S_{2,T}(\theta(1))$. Write

$$\begin{aligned}S_{2,T}(\theta(1)) &= \frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\Big(\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}-E\Big[\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big]\Big)(\theta(1)-\theta(t/T))+\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}E\Big[\frac{\partial^2\ell_t(\theta(1))}{\partial\theta\partial\theta'}\Big](\theta(1)-\theta(t/T))\\ &= S_{2,T}^{(1)}(\theta(1))+S_{2,T}^{(2)}(\theta(1)).\end{aligned}$$

For $S_{2,T}^{(1)}(\theta(1))$, again, similarly as in (B.11), we have

$$\begin{aligned}\mathbb{P}\left(\sup_{b\in I_T}\left\|S_{2,T}^{(1)}\right\|>\varepsilon\right) &\leqslant \sum_{b\in I_T}\mathbb{P}\left(\left\|S_{2,T}^{(1)}\right\|>\varepsilon\right)\\ &\leqslant |I_T|\times\sup_{b\in I_T}\mathbb{P}\left(\left\|S_{2,T}^{(1)}\right\|>\varepsilon\right)=O(b^{\gamma+\delta}),\end{aligned}$$

for some $0<\delta<1/2$. Moving to $S_{2,T}^{(2)}(\theta(1))$, notice that

$$\left\|S_{2,T}^{(2)}(\theta(1))\right\| \leqslant C\Big(\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\big(\frac{|t-T|}{T}\big)^{\gamma}\Big)\approx b^{\gamma}\int_C u^{\gamma}K(u)du = O(b^{\gamma}),$$

which holds uniformly over b. Thus, we have

$$\sup_{b\in I_T}\left\|S_{2,T}(\theta(1))\right\| \leqslant \sup_{b\in I_T}\left\|S_{2,T}^{(1)}(\theta(1))\right\|+\sup_{b\in I_T}\left\|S_{2,T}^{(2)}(\theta(1))\right\| = O_p(b^{\gamma}).$$

□

Lemma B4. *Define*

$$L(b) = (\hat{\theta}_{b,T} - \theta(1))' \, \omega_T(\theta(1)) \, (\hat{\theta}_{b,T} - \theta(1)),$$
$$A(b) = (\hat{\theta}_{b,T} - \tilde{\theta}_T)' \, \omega_T(\tilde{\theta}_T) \, (\hat{\theta}_{b,T} - \tilde{\theta}_T),$$

where $\hat{\theta}_{b,T} = \hat{\theta}_{\overline{K},b,T}$ *and* $\omega_T(\theta) = E_T\Big(\frac{\partial^2 \ell_{T+1}(\theta)}{\partial\theta\partial\theta'}\Big)$. *Suppose that Assumptions A1-A5 hold, we have*

$$\sup_{b \in I_T} \left| \frac{L(b) - A(b)}{L(b)} \right| = o_p(1). \tag{B.17}$$

Proof. Let us first expand $A(b)$:

$$\begin{aligned}
A(b) &= (\hat{\theta}_{b,T} - \tilde{\theta}_T)' \, \omega_T(\tilde{\theta}_T) \, (\hat{\theta}_{b,T} - \tilde{\theta}_T) \\
&= (\hat{\theta}_{b,T} - \theta(1) + \theta(1) + \tilde{\theta}_T)' \left(\omega_T(\theta(1)) + \underbrace{\Big[\frac{\partial \omega_T(\theta(1))}{\partial \theta_1}(\tilde{\theta}_T - \theta(1)) \; \ldots \; \frac{\partial \omega_T(\theta(1))}{\partial \theta_p}(\tilde{\theta}_T - \theta(1)) \Big]}_{\tilde{\omega}_T(\theta(1))_{p \times p}} \right) \\
&\qquad \times (\hat{\theta}_{b,T} - \theta(1) + \theta(1) + \tilde{\theta}_T) \\
&= L(b) - 2(\hat{\theta}_{b,T} - \theta(1))' \omega_T(\theta(1))(\tilde{\theta}_T - \theta(1)) + (\tilde{\theta}_T - \theta(1))' \omega_T(\theta(1))(\tilde{\theta}_T - \theta(1)) \\
&\quad + (\hat{\theta}_{b,T} - \theta(1))' \tilde{\omega}_T(\theta(1))(\hat{\theta}_{b,T} - \theta(1)) - 2(\hat{\theta}_{b,T} - \theta(1))' \tilde{\omega}_T(\theta(1))(\tilde{\theta}_T - \theta(1)) + (\tilde{\theta}_T - \theta(1))' \tilde{\omega}(1)(\theta(1))(\tilde{\theta}_T - \theta(1)) \\
&= L(b) - 2D_1(b) + D_1' + D_2(b) - 2D_3(b) + D_2',
\end{aligned}$$

where

$$\begin{aligned}
&D_1(b) = (\hat{\theta}_{b,T} - \theta(1))' \omega_T(\theta(1))(\tilde{\theta}_T - \theta(1)), \quad D_1' = (\tilde{\theta}_T - \theta(1))' \omega_T(\theta(1))(\tilde{\theta}_T - \theta(1)), \\
&D_2(b) = (\hat{\theta}_{b,T} - \theta(1))' \tilde{\omega}_T(\theta(1))(\hat{\theta}_{b,T} - \theta(1)), \quad D_3(b) = (\hat{\theta}_{b,T} - \theta(1))' \tilde{\omega}_T(\theta(1))(\tilde{\theta}_T - \theta(1)), \\
&D_2' = (\tilde{\theta}_T - \theta(1))' \tilde{\omega}_T(\theta(1))(\tilde{\theta}_T - \theta(1)).
\end{aligned}$$

Then, we have

$$\frac{L(b) - A(b)}{L(b)} = \frac{2D_1(b)}{L(b)} - \frac{D_1'}{L(b)} - \frac{D_2(b)}{L(b)} + \frac{D_3(b)}{L(b)} - \frac{D_2'}{L(b)}.$$

By Lemma B2 and Assumption A6(i), we have

$$\left\| \tilde{\theta}_T - \theta(1) \right\| = O_p((T\tilde{b})^{-1/2}). \tag{B.18}$$

We will show that

$$\sup_{b\in I_T}\left|\frac{D_1(b)}{L(b)}\right| = o_p(1),\ \ \sup_{b\in I_T}\left|\frac{D_2(b)}{L(b)}\right| = o_p(1),\ \ \sup_{b\in I_T}\left|\frac{D_3(b)}{L(b)}\right| = o_p(1), \tag{B.19}$$

$$\sup_{b\in I_T}\left|\frac{D_1'}{L(b)}\right| = o_p(1),\ \ \sup_{b\in I_T}\left|\frac{D_2'}{L(b)}\right| = o_p(1). \tag{B.20}$$

These bounds together with triangular inequality imply (B.17).

Proof of (B.19). First, by Lemma B3, we have

$$\sup_{b\in I_T}|L(b)| \leqslant \sup_{b\in I_T}\left\|\hat{\theta}_{b,T}-\theta(1)\right\|\,\|\omega_T(\theta(1))\|_{sp}\,\sup_{b\in I_T}\left\|\hat{\theta}_{b,T}-\theta(1)\right\| = O_p(r^2_{T,b,\delta,\gamma}), \tag{B.21}$$

for some $0<\delta<1/2$ and $0<\gamma\leqslant 1$. Write $\tilde{r}_{T,\tilde{b}} = (T\tilde{b})^{-1/2}$, we also have

$$\begin{aligned}
\sup_{b\in I_T}|D_1(b)| &\leqslant \sup_{b\in I_T}\left\|\hat{\theta}_{b,T}-\theta(1)\right\|\,\|\omega_T(\theta(1))\|_{sp}\,\left\|\tilde{\theta}_T-\theta(1)\right\| = O_p(r_{T,b,\delta,\gamma}\tilde{r}_{T,\tilde{b}}),\\
\sup_{b\in I_T}|D_2(b)| &\leqslant \sup_{b\in I_T}\left\|\hat{\theta}_{b,T}-\theta(1)\right\|\,\|\tilde{\omega}_T(\theta(1))\|_{sp}\,\sup_{b\in I_T}\left\|\hat{\theta}_{b,T}-\theta(1)\right\| = O_p(r^2_{T,b,\delta,\gamma}\tilde{r}_{T,\tilde{b}}),\\
\sup_{b\in I_T}|D_3(b)| &\leqslant \sup_{b\in I_T}\left\|\hat{\theta}_{b,T}-\theta(1)\right\|\,\|\tilde{\omega}_T(\theta(1))\|_{sp}\,\left\|\tilde{\theta}_T-\theta(1)\right\| = O_p(r_{T,b,\delta,\gamma}\tilde{r}^2_{T,\tilde{b}}).
\end{aligned}$$

These bounds imply that

$$\sup_{b\in I_T}\left|\frac{D_1(b)}{L(b)}\right| = O_p\Big(\frac{\tilde{r}_{T,\tilde{b}}}{r_{T,b,\delta,\gamma}}\Big) = o_p(1),$$

where $\frac{\tilde{r}_{T,\tilde{b}}}{r_{T,b,\delta,\gamma}}\to 0$ is guaranteed by Assumption A6. Similarly, we have

$$\sup_{b\in I_T}\left|\frac{D_2(b)}{L(b)}\right| = O_p(\tilde{r}_{T,\tilde{b}}) = o_p(1),$$

as $T\tilde{b}\to\infty$. Finally, we have

$$\sup_{b\in I_T}\left|\frac{D_3(b)}{L(b)}\right| = O_p\Big(\frac{\tilde{r}^2_{T,\tilde{b}}}{r_{T,b,\delta,\gamma}}\Big) = o_p(1),$$

where $\frac{\tilde{r}^2_{T,\tilde{b}}}{r_{T,b,\delta,\gamma}}\to 0$ is again guaranteed by Assumption A(6)).

Proof of (B.20). First, it is straightforward to show that

$$\left|D_1'\right| = O_p(\tilde{r}^2_{T,\tilde{b}}),\ \ \left|D_2'\right| = O_p(\tilde{r}^3_{T,\tilde{b}}).$$

Together with (B.21) and following the same reasoning above, we have

$$\sup_{b\in I_T}\left|\frac{D_1'}{L(b)}\right| = O_p\Big(\frac{\tilde{r}^2_{T,\tilde{b}}}{r^2_{T,b,\delta,\gamma}}\Big) = o_p(1), \quad \sup_{b\in I_T}\left|\frac{D_2'}{L(b)}\right| = O_p\Big(\frac{\tilde{r}^3_{T,\tilde{b}}}{r^2_{T,b,\delta,\gamma}}\Big) = o_p(1).$$

□

Appendix C. Proofs of the theorems

Appendix C.1. Proof of Theorem 1

Write $\hat{\theta}_{\overline{K},b,T} = \hat{\theta}_{b,T}$ and $\omega_T(\theta_T) = E_T\Big(\frac{\partial^2 \ell_{T+1}(\theta(1))}{\partial\theta\partial\theta'}\Big)$. It follows from Lemma B1 that, the infeasible objective function can be written as

$$(\hat{\theta}_{b,T} - \theta(1))'\,\omega_T(\theta(1))\,(\hat{\theta}_{b,T} - \theta(1)) = r_{T,b}q_T,$$

where q_T is a scalar $O_p(1)$ random variable and $r_{T,b,\gamma} = (Tb)^{-1/2} + b^\gamma$ for some $0 < \gamma \leqslant 1$. The first-order condition of $r_{T,b,\gamma}$ with respect to b gives $\hat{b} = O_p(T^{-\frac{1}{2\gamma+1}})$. Since the second order derivative of $r_{T,b,\gamma}$ is always positive, the optimal bandwidth minimize the objective function.

Appendix C.2. Proof of Theorem 2

Write $\hat{\theta}_{\overline{K},b,T} = \hat{\theta}_{b,T}$ and $\omega_T(\theta(1)) = E_T\Big(\frac{\partial^2 \ell_{T+1}(\theta(1))}{\partial\theta\partial\theta'}\Big)$. Let

$$\hat{b} := \arg\min_{b\in I_T} (\hat{\theta}_{b,T} - \tilde{\theta}(1))'\,\omega_T(\tilde{\theta}(1))\,(\hat{\theta}_{b,T} - \tilde{\theta}(1))$$

be the bandwidth selected according to the feasible criterion. As in the proof of Lemma B4, the decomposition of $A(b)$ implies that

$$A(\hat{b}) = L(\hat{b}) - 2D_1(\hat{b}) + D_1' + D_2(\hat{b}) - 2D_3(\hat{b}) + D_2'.$$

Then, we have

$$\begin{aligned}\frac{A(\hat{b})}{\inf_{b\in I_T} L(b)} &= \frac{L(\hat{b})}{\inf_{b\in I_T} L(b)} - \frac{2D_1(\hat{b})}{\inf_{b\in I_T} L(b)} + \frac{D_2(\hat{b})}{\inf_{b\in I_T} L(b)} - \frac{2D_3(\hat{b})}{\inf_{b\in I_T} L(b)} + \frac{D_1'}{\inf_{b\in I_T} L(b)} + \frac{D_2'}{\inf_{b\in I_T} L(b)} \\ &= I_1(\hat{b}) + I_2(\hat{b}) + I_3(\hat{b}) + I_4(\hat{b}) + I_5 + I_6.\end{aligned}$$

Following (B.19) and (B.20), we have

$$I_2(\hat{b}) = o_p(1), \quad I_3(\hat{b}) = o_p(1), \quad I_4(\hat{b}) = o_p(1), \quad I_5 = o_p(1), \quad I_6 = o_p(1).$$

What remains is to show that

$$I_1(\hat{b}) \xrightarrow{p} 1,$$

which is equivalent to verify that, for any $b, b' \in I_T$,

$$\sup_{b,b' \in I_T} \left| \frac{L(b) - L(b') - (A(b) - A(b'))}{L(b) + L(b')} \right| \xrightarrow{p} 0.$$

This follows immediately from Lemma B4:

$$\sup_{b,b' \in I_T} \left| \frac{L(b) - L(b') - (A(b) - A(b'))}{L(b) + L(b')} \right| \leqslant \sup_{b \in I_T} \left| \frac{L(b) - A(b)}{L(b)} \right| + \sup_{b' \in I_T} \left| \frac{L(b') - A(b')}{L(b')} \right| = o_p(1).$$

Appendix C.3. Proof of Theorem 3

In Lemma B1, we show that the local estimator obeys the following expansion:

$$\hat{\theta}_{K,b,T} - \theta(1) = -H_{1,T}^{-1} S_{1,T} - H_{1,T}^{-1} S_{2,T},$$

where

$$H_{1,T} = \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\theta(1))}{\partial\theta\partial\theta'}, \quad S_{1,T} = \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial \ell_t(\theta(t/T))}{\partial\theta}, \quad S_{2,T} = \frac{1}{Tb} \sum_{t=1}^{T} k_{tT} \frac{\partial^2 \ell_t(\overline{\theta}(1))}{\partial\theta\partial\theta'} (\theta(1) - \theta(t/T)),$$

and $\overline{\theta}(1)$ lies between $\hat{\theta}_{K,b,T}$ and $\theta(1)$. Since $\left\|H_{1,T}\right\| = O_p(1)$, $\left\|S_{1,T}\right\| = O_p((Tb)^{-1/2})$, $\left\|S_{2,T}\right\| = O_p(b^{\gamma})$, when $T^{1/2} b^{1/2+\gamma} \to 0$, the dominating term is $S_{1,T}$. CLT in LemmaB1(ii) holds. Theorem 3(ii) follows immediately from continuous mapping theorem.

When $T^{1/2} b^{1/2+\gamma} \to \infty$, the dominating term is $S_{2,T}$. By similar analysis as in the proof of Lemma B1, we have

$$b^{-\gamma} S_{2,T} \xrightarrow{p} \mu_{\gamma,K} E\Big[\frac{\partial^2 \ell_T(\theta(1))}{\partial\theta\partial\theta'}\Big] C,$$

where C is a collection of constant given in Assumption A3(i). Theorem 3(ii) follows again from continuous mapping theorem. 3(iii) follows immediately by combining the results obtained in (i) and (ii).

Appendix D. Additional tables and figures

Table D.1: Forecasting performance for inflation in the United States: 2020Q1-2023Q1

UCSV	5.351				
AR(4)	0.855				
	Non-local	$R = 40$	Opt-R	Opt-G	Opt-E
Asset prices					
FEDFUNDS	0.878	1.150	1.092	1.084	1.078
TB3MS	0.874	1.163	1.091	1.099	1.091
GS10	0.864	1.079	1.072	1.062	1.039
GS10TB3Mx	0.851	1.163	1.102	1.081	1.115
term spread	0.850	1.168	1.111	1.092	1.128
S&P 500	0.859	1.262	1.263	1.075	1.243
VXOCLSx	0.856	1.255	1.242	1.170	1.297
Real economic activity					
DPIC96	0.863	1.276	1.206	1.109	1.228
GPDIC1	0.888	1.477	1.451	1.280	1.694
INDPRO	1.090	1.651	1.697	1.453	1.929
CE16OV	2.658	6.570	7.465	4.542	10.500
UNRATE	0.934	1.090	0.917	1.023	0.906
LNS14000026	0.922	1.056	0.954	0.996	0.900
HOUST	0.865	1.099	1.110	1.089	1.107
PERMIT	0.853	1.240	1.188	1.114	1.208
Price indices					
CPIAUCSL	0.885	1.221	1.178	1.088	1.189
CPIAPPSL	1.051	1.199	1.050	1.108	1.149
CPIENGSL	0.871	1.172	1.139	1.080	1.124
PPIACO	0.881	1.301	1.286	1.109	1.256
PCECTPI	0.863	1.235	1.147	1.106	1.214
Monetary measures					
BOGMBASEREALx	0.916	1.045	1.126	1.197	1.100
M1REAL	2.208	2.517	7.956	6.290	7.311
M2REAL	0.968	1.647	1.547	1.304	1.716

Notes: The description of predictors is detailed in Table 13. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. The row "UCSV" presents the exact RMSFEs of the forecasts from the UCSV model. The row "AR(4)" presents the ratio of the RMSFEs of the forecasts from an AR(4) model relative to the benchmark. In other columns, the numbers are also the ratios of the RMSFEs relative to the benchmark.

Table D.2: Forecasting performance for inflation in Canada: 2020Q1-2023Q1

UCSV	51.517				
AR(4)	0.826				
	Non-local	$R = 40$	Opt-R	Opt-G	Opt-E
Asset prices					
BANK_RATE_L	0.827	0.829	0.920	0.799	0.851
TBILL_3M	0.828	0.875	0.993	0.807	0.902
GOV_AVG_10pY	0.832	0.853	0.894	0.825	0.890
G_AVG_5.10.Bank_rate	0.829	0.964	0.931	0.833	0.984
G_AVG_10p.TBILL_3M	0.828	0.923	0.941	0.834	0.936
TSX_CLO	0.816	0.954	0.960	0.822	1.070
Real economic activity					
REAL_GDP	1.005	1.583	1.145	1.021	1.598
hhold_dispo_income	0.794	0.899	0.944	0.854	0.949
REAL_I	0.850	1.054	1.119	0.954	1.143
CANPROINDQISMEI	0.839	0.916	0.817	0.780	0.907
LFEMTTTTCAQ647S	1.465	3.180	3.191	2.256	3.727
UNEMP_CAN	0.842	0.679	0.768	0.704	0.808
hstart_CAN	0.823	0.940	0.943	0.829	0.960
Price indices					
CPI_ALL_CAN	0.837	0.902	0.947	0.775	0.912
IPPI_CAN	0.832	0.942	0.954	0.872	0.968
C_PRICE	0.785	1.008	1.022	0.908	1.043
Monetary measures					
MBASE1	2.703	21.638	59.313	11.931	67.582
CRED_BUS_cb	0.827	0.955	0.939	0.843	0.994
CRED_HOUS_cb	0.860	0.833	0.775	0.787	0.870

Notes: The description of predictors is detailed in Table 14. $R = 40$: rolling window with fixed window size; Opt-R: optimal selection with $K_1(u)$; Opt-G: optimal selection with $K_2(u)$; Opt-E: optimal selection with $K_3(u)$. We set $b = cT^{-1/3}$ with c ranging from 1 to 5 (width 0.05). The tuning parameter used to compute $\tilde{\theta}_T$ is set to $\tilde{b} = 1.06T^{-1/5}$. The row "UCSV" presents the exact RMSFEs of the forecasts from the UCSV model. The row "AR(4)" presents the ratio of the RMSFEs of the forecasts from an AR(4) model relative to the benchmark. In other columns, the numbers are also the ratios of the RMSFEs relative to the benchmark.