Online appendix for "Time-varying Forecast Combination for High-Dimensional Data"

By Bin Chen and Kenwin Maung

1. Proofs

This section contains the proofs for the following results in the main text.

1.1. Proof for Lemma 1

Proof. For $j = 0$,

$$\begin{aligned}
&E\left[(Th)^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} X_s X_s^\top k\left(\frac{s-t}{Th}\right)\right]\\
&= E\left[(Th)^{-1}\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} X_s X_s^\top k\left(\frac{s-t}{Th}\right)\right] - k(0)(Th)^{-1}M(t/T)\\
&= M(t/T)\int k(u)du + o(1)
\end{aligned}$$

The last equality comes from a Taylor expansion on $M(s/T)$ and Riemann approximation of the integral. For $j > 0$ since $s - t = 0$ for equal indices,

$$E(h^{-j}S_j(t/T)) = E\left[(Th)^{-1}\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} X_s X_s^\top \left(\frac{s-t}{Th}\right)^j k\left(\frac{s-t}{Th}\right)\right] = M(t/T)\int u^j k(u)du + o(1).$$

Now to verify that the variance of $h^{-j}S_j(t/T)$ goes to 0. Let $j = 0$, and consider the (m, n)th element of S_0,

$$\begin{aligned}
Var(S_{0,(m,n)}) &= (Th)^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} Var(X_{s,m}X_{s,n})k^2\left(\frac{s-t}{Th}\right)\\
&\quad + 2(Th)^{-2}\sum_{\substack{t-\lfloor Th\rfloor\le s<l\le t+\lfloor Th\rfloor\\ s,l\neq t}} Cov(X_{s,m}X_{s,n}, X_{l,m}X_{l,n})k\left(\frac{s-t}{Th}\right)k\left(\frac{l-t}{Th}\right)\\
&= (Th)^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} Var(X_{s,m}X_{s,n})k^2\left(\frac{s-t}{Th}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ 2(Th)^{-2} \sum_{t-\lfloor Th \rfloor \le s < l \le t+\lfloor Th \rfloor} Cov(X_{s,m}X_{s,n}, X_{l,m}X_{l,n}) k\left(\frac{s-t}{Th}\right) k\left(\frac{l-t}{Th}\right) \\
&- 2(Th)^{-2} \sum_{\substack{t-\lfloor Th \rfloor < s < l < t+\lfloor Th \rfloor \\ s=t \vee l=t}} Cov(X_{s,m}X_{s,n}, X_{l,m}X_{l,n}) k\left(\frac{s-t}{Th}\right) k\left(\frac{l-t}{Th}\right) \\
&\equiv V_{s1} + V_{s2} - V_{s3}
\end{aligned}
$$

Note that

$$
\begin{aligned}
&V_{s1} \le (Th)^{-2} \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Var(X_{s,m}X_{s,n}) k^2\left(\frac{s-t}{Th}\right) = (Th)^{-1} Var(X_{t,m}X_{t,n}) \int k^2(u)du + o(1), \\
&V_{s2} \le 2(Th)^{-2} k^2(0) \sum_j \beta^*(j)^{\delta/(1+\delta)} = o(1),
\end{aligned}
$$

where we have used Lemma (A.2) of Juhl and Xiao (2013) and the mixing condition to bind the covariance terms by the beta coefficients, and the assumption that $k(u) \le k(0), \forall u \in [-1, 1]$. Likewise, we can use the same strategy for V_{s3}, and thus $Var(S_{0,(m,n)}) \to 0$. The proof for $j > 0$, where $Var(h^{-j} S_j(t/T)) \to 0$, is the same (but more convenient) and hence omitted. Therefore, we conclude the proof for (A.3). The proof of (A.4) follows that of Lemma 2 in Cai (2007). □

1.2. Proof for Lemma 2

Proof. Recall the definitions $r(t/T) = (r_0^\top(t/T), r_1^\top(t/T))^\top$ and $\Gamma_j(t/T) = Cov(X_t\varepsilon_{t+1}, X_{t+j}\varepsilon_{t+j+1})$. For our purposes here, define $\Gamma_{s,l} = Cov(X_s\varepsilon_{s+1}, X_l\varepsilon_{l+1})$. We have,

$$
ThVar(r_0(t/T)) = T^{-1}h \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \ne t}}^{t+\lfloor Th \rfloor} V(s/T) k_{st}^2 + 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l \le t+\lfloor Th \rfloor \\ s,l \ne t}} \Gamma_{s,l} k_{st} k_{lt}. \tag{1}
$$

Consider a u_T such that $u_T \to \infty$, $u_T/Th \to 0$, $hu_T \to 0$ and $u_T/\sqrt{T} \to 0$. Since the processes are β-mixing, we use u_T to control the terms that matter asymptotically. In particular, the covariances of data points that are more than u_T indices apart do not. So we decompose the covariance term while taking into consideration the dependence between the reflected pseudo-data and the original. This yields (only the summation is shown to save on notation)

$$
\sum_{\substack{t-\lfloor Th \rfloor \le s < l \le t+\lfloor Th \rfloor \\ s,l \ne t}} = \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} + \sum_{\substack{t < s < l \le t+\lfloor Th \rfloor \\ 1 \le l-s \le d_T}} + \sum_{\substack{t-\lfloor Th \rfloor \le s < t < l \le t+\lfloor Th \rfloor \\ |2t-s-l| \le d_T}} + \sum_{\substack{t-\lfloor Th \rfloor \le s < l \le t+\lfloor Th \rfloor \\ (s,l) \in \Theta_d}},
$$

where $\Theta_u = \{(s,l) : s, l \neq t \wedge \neg(1 \leq l - s \leq u_T \vee |2t - s - l| \leq u_T)\}$. Next, we further decompose

$$\sum_{\substack{t-\lfloor Th\rfloor \leq s<t<l\leq t+\lfloor Th\rfloor \\ |2t-s-l|\leq d_T}} = \sum_{\substack{t-\lfloor Th\rfloor \leq s<t<l\leq t+\lfloor Th\rfloor \\ 2t-s-l=0}} + \sum_{\substack{t-\lfloor Th\rfloor \leq s<t<l\leq t+\lfloor Th\rfloor \\ 1\leq|2t-s-l|\leq d_T}}.$$

Note that because of the data reflection,

$$2 \times T^{-1}h \sum_{\substack{t-\lfloor Th\rfloor \leq s<t<l\leq t+\lfloor Th\rfloor \\ 2t-s-l=0}} \Gamma_{s,l} k_{st} k_{lt} = T^{-1}h \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} V(s/T) k_{st}^2,$$

and

$$\sum_{\substack{t-\lfloor Th\rfloor \leq s<t<l\leq t+\lfloor Th\rfloor \\ 1\leq|2t-s-l|\leq d_T}} = \sum_{\substack{t-\lfloor Th\rfloor \leq s<l<t \\ 1\leq l-s\leq d_T}} + \sum_{\substack{t<s<l\leq t+\lfloor Th\rfloor \\ 1\leq l-s\leq d_T}}.$$

So (1) becomes,

$$\begin{aligned}
ThVar(r_0(t/T)) &= 2T^{-1}h \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} V(s/T) k_{st}^2 \\
&\quad + 2\Bigg(2T^{-1}h \sum_{\substack{t-\lfloor Th\rfloor \leq s<l<t \\ 1\leq l-s\leq d_T}} \Gamma_{s,l} k_{st} k_{lt} + 2T^{-1}h \sum_{\substack{t<s<l\leq t+\lfloor Th\rfloor \\ 1\leq l-s\leq d_T}} \Gamma_{s,l} k_{st} k_{lt}\Bigg) \\
&\quad + 2T^{-1}h \sum_{\substack{t-\lfloor Th\rfloor \leq s<l\leq t+\lfloor Th\rfloor \\ (s,l)\in\Theta_d}} \Gamma_{s,l} k_{st} k_{lt} \\
&\equiv I_{var} + I_{covar} + I_{small}.
\end{aligned}$$

Then, we have,

$$I_{var} = 2T^{-1}h \sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} V(s/T) k_{st}^2 - 2(Th)^{-1} V(t/T) k^2(0) = 2V(t/T) \int k^2(u)du + o(1).$$

For I_{covar}, note that

$$\begin{aligned}
2T^{-1}h \sum_{\substack{t-\lfloor Th\rfloor \leq s<l<t \\ 1\leq l-s\leq d_T}} \Gamma_{s,l} k_{st} k_{lt} &= 2T^{-1}h \sum_{\substack{t-\lfloor Th\rfloor \leq s<l<t \\ 1\leq l-s\leq d_T}} \Gamma_{s,t+l-s} k_{st}^2 \\
&\quad + 2T^{-1}h \sum_{\substack{t-\lfloor Th\rfloor \leq s<l<t \\ 1\leq l-s\leq d_T}} (\Gamma_{s,l} - \Gamma_{s,t+l-s}) k_{st}^2
\end{aligned}$$

$$
+ 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} \Gamma_{s,l} k_{st}(k_{lt} - k_{st})
$$

$$
\equiv J_1 + J_2 + J_3.
$$

Observe that

$$
J_1 = 2T^{-1}h \sum_{t-\lfloor Th \rfloor \le s < t} k_{st}^2 \sum_{1 \le l-s \le u_T} \Gamma_{s,t+l-s} \to 2\int_{-1}^{0} k^2(u)du \sum_{j=1}^{\infty} \Gamma_j(t/T),
$$

where the last relation comes from a Taylor expansion in the first coordinate of Γ around t/T, and recalling that $u_T \to \infty$.

Next, note that for $t - \lfloor Th \rfloor \le s < l < t$, we have $|\frac{s-t}{T}| < h$ and so by Taylor's theorem $|\Gamma_{s,l} - \Gamma_{s,l+t-s}| = O(h)$. Hence, $|J_2| \le 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} |\Gamma_{s,l} - \Gamma_{s,l+t-s}| k_{st}^2 = O(hu_T)$ since the number of summands is proportional to u_T.

Furthermore, by Lipschitz continuity, we have $|k_{st} - k_{lt}| = \frac{1}{h}|k(\frac{s-t}{Th}) - k(\frac{l-t}{Th})| \le C|\frac{s-l}{Th^2}| \le C\frac{u_T}{Th^2}$. Hence,

$$
\begin{aligned}
|J_3| &\le 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} |\Gamma_{s,l}| k_{st} |k_{lt} - k_{st}| \\
&= 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} |\Gamma_{s,l} - \Gamma_{s,t+l-s}| |k_{st}| k_{lt} - k_{st}| + 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} |\Gamma_{s,t+l-s}| k_{st} |k_{lt} - k_{st}| \\
&= O(\frac{u_T^2}{T}) + O(\frac{u_T}{Th})
\end{aligned}
$$

So we conclude that $2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ 1 \le l-s \le d_T}} \Gamma_{s,l} k_{st} k_{lt} \to 2\int_{-1}^{0} k^2(u)du \sum_{j=1}^{\infty} \Gamma_j(t/T)$, and likewise $2T^{-1}h \sum_{\substack{t < s < l \le t+\lfloor Th \rfloor \\ 1 \le l-s \le d_T}} \Gamma_{s,l} k_{st} k_{lt} \to 2\int_{0}^{1} k^2(u)du \sum_{j=1}^{\infty} \Gamma_j(t/T)$. Thus, $I_{covar} \to 2(2\nu_0 \sum_{j=1}^{\infty} \Gamma_j(t/T))$.

Lastly, for I_{small},

$$
\begin{aligned}
I_{small} = 2T^{-1}h &\sum_{\substack{t-\lfloor Th \rfloor \le s < l < t \\ l-s > u_T}} \Gamma_{s,l} k_{st} k_{lt} + 2T^{-1}h \sum_{\substack{t < s < l \le t+\lfloor Th \rfloor \\ l-s > u_T}} \Gamma_{s,l} k_{st} k_{lt} \\
&+ 2T^{-1}h \sum_{\substack{t-\lfloor Th \rfloor \le s < t < l \le t+\lfloor Th \rfloor \\ |2t-l-s| > u_T}} \Gamma_{s,l} k_{st} k_{lt} \\
&\equiv I_{s1} + I_{s2} + I_{s3}
\end{aligned}
$$

For the (m,n)th component of I_{s1}, we have $|I_{s1(m,n)}| \leq Ck^2(0)(Th)^{-1}\sum_{j=u_T}^{Th} j^2\beta^*(j)^{\delta/(1+\delta)}$. The situation is the same for I_{s2} and I_{s3}.

Taken together, we have

$$ThVar(r_0(t/T)) \to 2\Bigg(V(t/T) + 2\sum_{j=1}^{\infty}\Gamma_j(t/T)\Bigg)\nu_0 = 2\Omega(t/T)\nu_0.$$

Likewise,

$$ThVar(h^{-1}r_1(t/T)) \to 2\Omega(t/T)\nu_2,$$

and

$$ThCovar(r_0(t/T), h^{-1}r_1(t/T)) \to 2\Omega(t/T)\nu_1 = 0.$$

□

1.3. Proof for Lemma 3

Proof. First, to get an expression for $\hat{\beta}_s$, we do the block matrix inversion for $S(s/T)$. However, since we are only interested in the first $(d+1)$ components of $\hat{\gamma}_s$, we focus on the blocks in the top row of $S^{-1}(s/T)$. For notational convenience, define $A_j(s/T) \equiv A_{js}$, where A is an arbitrary matrix or vector. Then we have,

$$\begin{aligned}
\hat{\beta}_s &= (e_1^\top \otimes I_{(d+1)})\hat{\gamma}_s \\
&= (e_1^\top \otimes I_{(d+1)})S^{-1}(s/T)R(s/T) \\
&= \underbrace{[S_{0s}^{-1} + S_{0s}^{-1}S_{1s}^\top(S_{2s} - S_{1s}S_{0s}^{-1}S_{1s}^\top)^{-1}S_{1s}S_{0s}^{-1}, \quad -S_{0s}^{-1}S_{1s}^\top(S_{2s} - S_{1s}S_{0s}^{-1}S_{1s}^\top)^{-1}]}_{(d+1)\times 2(d+1)} \underbrace{\begin{bmatrix} R_{0s} \\ R_{1s} \end{bmatrix}}_{2(d+1)\times 1}. \qquad (2)
\end{aligned}$$

Denote $\Delta_1 \equiv S_{0s}^{-1}S_{1s}^\top(S_{2s} - S_{1s}S_{0s}^{-1}S_{1s}^\top)^{-1}S_{1s}S_{0s}^{-1}$, and $\Delta_2 \equiv S_{0s}^{-1}S_{1s}^\top(S_{2s} - S_{1s}S_{0s}^{-1}S_{1s}^\top)^{-1}$. From Lemma 1, both Δ_1 and Δ_2 are $O_p((Th)^{-1/2}) = O_p(h^2)$. Recall that we have derived the following expression, $R_{js} = S_{js}\beta_s + S_{(j+1)s}\beta_s' + B_{js} + D_{js} + r_{js}$. Let $E_{js} \equiv B_{js} + D_{js} = T^{-1}\sum_{\substack{l=s-\lfloor Th\rfloor \\ l\neq s}}^{s+\lfloor Th\rfloor} X_l X_l^\top (\frac{l-s}{T})^j$ $k_{ls}Q_{l,s}^*$, where $Q_{l,s}^* = \beta(l/T) - \beta(s/T) - (\frac{l-s}{T})\beta'(s/T)$ is a non-random function. Thus, we have

$$\begin{aligned}
\hat{\beta}_s &= \beta_s + S_{0s}^{-1}(E_{0s} + r_{0s}) + \bar{R}_s \\
&= \beta_s + S_{0s}^{-1}T^{-1}\sum_{\substack{l=s-\lfloor Th\rfloor \\ l\neq s}}^{s+\lfloor Th\rfloor} k_{ls}X_l \underbrace{[X_l^\top Q_{l,s}^* + \varepsilon_{l+1}]}_{\equiv\, \varepsilon_{l,s}^*} + \bar{R}_s
\end{aligned}$$

$$\equiv \beta_s + D_s + \bar{R}_s$$

where $\bar{R}_s = S_{0s}^{-1} S_{1s} \beta_s' + \Delta_1 R_{0s} - \Delta_2 R_{1s} = \Delta_1(E_{0s} + r_{0s}) + \Delta_2(E_{1s} + r_{1s})$. Hence, $\bar{R}_s$ is smaller than $S_{0s}^{-1}(E_{0s} + r_{0s})$ by a factor of h^2 since $E_{0s} + r_{0s}$ and $E_{1s} + r_{1s}$ are of the same order. So instead, we consider $\tilde{\beta}_s = \beta_s + D_s$. The cross-validation statistic with $\tilde{\beta}_s$ is given by

$$\begin{aligned}
\tilde{CV}(h) &= T^{-1} \sum_{s=1}^{T} (y_{s+1} - X_s^\top \tilde{\beta}_s)^2 \\
&= T^{-1} \sum_{s=1}^{T} (\varepsilon_{s+1} + X_s^\top \underbrace{(\beta_s - \tilde{\beta}_s)}_{=-D_s})^2 \\
&= T^{-1} \sum_{s=1}^{T} \varepsilon_{s+1}^2 + T^{-1} \sum_{s=1}^{T} D_s^\top X_s X_s^\top D_s - 2T^{-1} \sum_{s=1}^{T} \varepsilon_{s+1} X_s^\top D_s.
\end{aligned}$$

Here, let $M^{-1}(s/T) = M_s^{-1}$ and note that

$$\begin{aligned}
D_s &= M_s^{-1}\, T^{-1} \sum_{\substack{l=s-\lfloor Th \rfloor \\ l \neq s}}^{s+\lfloor Th \rfloor} k_{ls} X_l \varepsilon_{l,s}^* + (S_{0s}^{-1} - M_s^{-1})\, T^{-1} \sum_{\substack{l=s-\lfloor Th \rfloor \\ l \neq s}}^{s+\lfloor Th \rfloor} k_{ls} X_l \varepsilon_{l,s}^* \\
&= D_{s,1} + D_{s,2},
\end{aligned}$$

where $S_{0s} - M_s = O_p((Th)^{-1/2}) = O_p(h^2)$ by Lemma 1 and Chebyshev (or by Lemma A below). So $D_{s,2}$ is smaller than $D_{s,1}$ by a factor of h^2. Therefore, we use $\hat{D}_s = M_s^{-1}\, T^{-1} \sum_{\substack{l=s-\lfloor Th \rfloor \\ l \neq s}}^{s+\lfloor Th \rfloor} k_{ls} X_l \varepsilon_{l,s}^*$ instead. The corresponding CV statistic is thus

$$\begin{aligned}
\hat{\tilde{CV}}(h) &= T^{-1} \sum_{s=1}^{T} \varepsilon_{s+1}^2 + T^{-1} \sum_{s=1}^{T} \hat{D}_s^\top X_s X_s^\top \hat{D}_s - 2T^{-1} \sum_{s=1}^{T} \varepsilon_{s+1} X_s^\top \hat{D}_s \\
&= T^{-1} \sum_{s=1}^{T} \varepsilon_{s+1}^2 + CV_1 - CV_2.
\end{aligned}$$

□

1.4. Proof for Lemma 4

Proof. For the first term in $\hat{\tilde{CV}}(h)$, $E[T^{-1} \sum_{s=1}^{T} \varepsilon_{s+1}^2] = \int_0^1 \sigma^2(\tau) d\tau + O(T^{-1})$, where the approximation order is smaller than $T^{-4/5}$. Next, for CV_2,

$$E[CV_2] = 2T^{-2} \sum_{s=1}^{T} \sum_{\substack{l=s-\lfloor Th \rfloor \\ l \neq s}}^{s+\lfloor Th \rfloor} E[\varepsilon_{s+1} X_s^\top M_s^{-1} X_l \varepsilon_{l,s}^*] k_{ls}$$

$$= 2T^{-2}\sum_{s=1}^{T}\sum_{\substack{l=s-\lfloor Th\rfloor\\ l\neq s}}^{s+\lfloor Th\rfloor} E[\varepsilon_{s+1}X_s^\top M_s^{-1}X_lX_l^\top]Q_{l,s}^*k_{ls} + 2T^{-2}\sum_{s=1}^{T}\sum_{\substack{l=s-\lfloor Th\rfloor\\ l\neq s}}^{s+\lfloor Th\rfloor} E[\varepsilon_{s+1}X_s^\top M_s^{-1}X_l\varepsilon_{l+1}]k_{ls}$$

$$= CV_{21} + CV_{22}$$

We decompose CV_{21} as such,

$$CV_{21} = 2T^{-2}\sum_{s=1}^{T}\left\{\sum_{l=s+1}^{s+\lfloor Th\rfloor} E[\varepsilon_{s+1}X_s^\top M_s^{-1}X_lX_l^\top]Q_{l,s}^*k_{ls} + \sum_{l=s-\lfloor Th\rfloor}^{s-1} E[\varepsilon_{s+1}X_s^\top M_s^{-1}X_lX_l^\top]Q_{l,s}^*k_{ls}\right\}.$$

Written this way, we can easily see that $CV_{21} = 0$ because we can apply the law of iterated expectations conditioned on $\mathcal{I}_s$ to the first term, and on $\mathcal{I}_l$ to the second term. In the former case where $l > s$, X_l will be measurable with respect to $\mathcal{I}_s$ because we are using data reflection. In addition, we get the same result for $CV_{22} = 0$. Hence, $E[CV_2] = 0$.

Now, for CV_1 we have,

$$\begin{aligned}
CV_1 = T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor} Q_{m,s}^{*\top}X_mX_m^\top M_s^{-1}X_sX_s^\top M_s^{-1}X_nX_n^\top Q_{n,s}^*k_{ms}k_{ns}\\
+ T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor} \varepsilon_{m+1}X_m^\top M_s^{-1}X_sX_s^\top M_s^{-1}X_nX_n^\top Q_{n,s}^*k_{ms}k_{ns}\\
+ T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor} Q_{m,s}^{*\top}X_mX_m^\top M_s^{-1}X_sX_s^\top M_s^{-1}X_n\varepsilon_{n+1}k_{ms}k_{ns}\\
+ T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor} \varepsilon_{m+1}X_m^\top M_s^{-1}X_sX_s^\top M_s^{-1}X_n\varepsilon_{n+1}k_{ms}k_{ns}\\
= CV_{bias} + CV_{11} + CV_{12} + CV_{variance}.
\end{aligned}$$

By taking a Taylor expansion of $\beta(\frac{\cdot}{T})$ in $Q_{\cdot,s}^*$ around $\frac{s}{T}$, we can rewrite $Q_{\cdot,s}^* = \frac{1}{2}\beta_s''(\frac{\cdot-s}{T})^2 + O(T^{-3})$, and we ignore the last term which is a smaller order term. For convenience, let $M_s^{-1}X_sX_s^\top M_s^{-1} \equiv U_s$, $X_mX_m^\top \equiv V_m$, and $k_{ms}k_{ns}(\frac{m-s}{T})^2(\frac{n-s}{T})^2 \equiv K_{s,m,n}$. Consider CV_{bias} first. We have 2 cases to consider: (1) $m < s < n$, and (2) $s < n < m$. We consider only the first case:

$E(CV_{bias}^1)$

$$
\begin{aligned}
&= \frac{1}{4T^3}\sum_{s=1}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1}\sum_{n=s+1}^{s+\lfloor Th\rfloor} E(\beta_s^{''\top} X_m X_m^\top M_s^{-1} X_s X_s^\top M_s^{-1} X_n X_n^\top \beta_s^{''}) k_{ms} k_{ns} \left(\frac{m-s}{T}\right)^2 \left(\frac{n-s}{T}\right)^2 \\
&= \frac{1}{4T^3}\sum_{m<s<n} \mathrm{Tr}\{\beta_s^{''}\beta_s^{''\top} E(U_s)E(V_m)E(V_n)\} K_{s,m,n} \\
&+ \frac{1}{4T^3}\sum_{m<s<n} \mathrm{Tr}\{\beta_s^{''}\beta_s^{''\top}[E(U_sV_m)E(V_n^\top) - E(U_s)E(V_m)E(V_n)]\} K_{s,m,n} \\
&+ \frac{1}{4T^3}\sum_{m<s<n} \mathrm{Tr}\{\beta_s^{''}\beta_s^{''\top}[E(U_sV_mV_n^\top) - E(U_sV_m)E(V_n^\top)]\} K_{s,m,n} \\
&= C_1^1 + C_2^1 + C_3^1
\end{aligned}
$$

We start with C_3^1. Let $\bar{D}$ represent the maximum adjacent distance between two indices. Here, we have 2 subcases: (1a) $s-m \geq n-s$, and (1b) $n-s \geq s-m$. Considering case (1a), we let $\bar{D} = s-m$, and invoke Lemma A.2 of Juhl and Xiao (2013) to get,

$$
\begin{aligned}
|C_3^{1a}| &\leq \frac{1}{4T^3}\sum_{\substack{s=1\\ s-m=\bar{D}}}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1}\sum_{n=s+1}^{s+\lfloor Th\rfloor} |\mathrm{Tr}(\beta_s^{''}\beta_s^{''\top})| M_1^{1/1+\delta}\beta^*(s-m)^{\delta/1+\delta} k_{ms}k_{ns}\left(\frac{m-s}{T}\right)^2\left(\frac{n-s}{T}\right)^2 \\
&= \frac{Ch^3}{4T^3h}\sum_{\substack{s=1\\ s-m=\bar{D}}}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1}\beta^*(s-m)^{\delta/1+\delta}\underbrace{k\left(\frac{m-s}{Th}\right)}_{\leq k(0)}\underbrace{\left(\frac{m-s}{Th}\right)^2}_{\leq 1}\sum_{n=s+1}^{s+\lfloor Th\rfloor} k\left(\frac{n-s}{Th}\right)\left(\frac{n-s}{Th}\right)^2 \\
&= O\left(\frac{h^3}{T}\right),
\end{aligned}
$$

where $M_1 = \max(M_1^a, M_1^b)$, $M_1^a = \sup_{s,m,n}\int U_sV_mV_n^\top dF(X_s,X_m,X_n)$, and $M_1^b = \sup_{s,m,n}\int\int U_s V_mV_n^\top dF(X_s,X_m)dF(X_n)$, and by assumption A.5, we know that $M_1 < \infty$. The case for (1b) would be the same since it affects only the β-coefficient terms. Repeating the same steps for C_2^1 leads to the same conclusion.

Similar derivation shows that $C_i^j = O(h^3/T)$ for $i = 2,3$, and cases $j = 1,2$. Note that the results can be generalized to cases where the positions of m and n are swapped or when s is the largest index. Also, the cases where $n = m$ would yield an order smaller by T^{-1}, and thus are negligible. Hence, we are left with the first term for all the cases,

$$
C_1 \equiv \frac{1}{4T^3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor} \mathrm{Tr}\{\beta_s^{''}\beta_s^{''\top}E(U_s)E(V_{m,s})E(V_{n,s})\}K_{s,m,n}.
$$

Using a Taylor approximation around s/T on M_m and M_n we get,

$$
\begin{aligned}
C_1 &= \frac{1}{4T^3}\sum_s\sum_m\sum_n \mathrm{Tr}\{M_s\beta_s^{''}\beta_s^{''\top}\}k_{ms}k_{ns}\left(\frac{m-s}{T}\right)^2\left(\frac{n-s}{T}\right)^2 + O\left(\frac{h^4}{T}\right)\\
&= \frac{h^4}{4T}\sum_s \mathrm{Tr}\{M_s\beta_s^{''}\beta_s^{''\top}\}(\int u^2k(u)du)^2 + o(1)\\
&= \frac{h^4\mu_2^2}{4}\int \mathrm{Tr}\{M(\tau)\beta^{''}(\tau)\beta^{''\top}(\tau)\}d\tau + o(1).
\end{aligned}
$$

Therefore, we conclude that $E(CV_{bias}) = \frac{h^4\mu_2^2}{4}\int \mathrm{Tr}\{M(\tau)\beta^{''}(\tau)\beta^{''}(\tau)\}d\tau + O(\frac{h^3}{T})$.

Now we consider $CV_{variance}$. To begin, define two set of indices as $\Theta_1 = \{(m,n) : m = n\}$ and $\Theta_2 = \{(m,n) : m+n = 2s\}$. Note that the summands in $E(CV_{variance})$ with indices that do not belong in $\Theta_1 \cup \Theta_2$ are 0 by the law of iterated expectations. So we decompose $CV_{variance} = V_{\Theta_1} + V_{\Theta_2}$ whose definition will be clear presently.

Consider the first case where $s > m$ instead of $m > s$. Then, we have

$$
\begin{aligned}
E(V^1_{\Theta_1}) &= \frac{1}{T^3}\sum_{s=1}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1}\mathrm{Tr}\{M_s^{-1}E(X_sX_s^\top)M_s^{-1}E(\varepsilon_{m+1}^2X_m^\top X_m)\}k_{ms}^2\\
&+\frac{1}{T^3}\sum_{s=1}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1}\mathrm{Tr}\{E(M_s^{-1}X_sX_s^\top M_s^{-1}\varepsilon_{m+1}^2X_m^\top X_m) - M_s^{-1}E(X_sX_s^\top)M_s^{-1}E(\varepsilon_{m+1}^2X_m^\top X_m)\}k_{ms}^2\\
&= C^{1(1)}_{\Theta_1} + C^{1(2)}_{\Theta_1}.
\end{aligned}
$$

Next,

$$
\begin{aligned}
|C^{1(2)}_{\Theta_1}| &\le \frac{1}{T^3}\sum_{s=1}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1} V_{(1)}^{1/1+\delta}\beta^*(s-m)^{\delta/1+\delta}k_{m,s}^2\\
&\le \frac{C}{T^3h^2}\sum_{s=1}^{T}\sum_{m=s-\lfloor Th\rfloor}^{s-1}\beta^*(s-m)^{\delta/1+\delta} = O((Th)^{-2}) = O(h^3/T),
\end{aligned}
$$

where $V_{(1)} = \max(V_1^a, V_1^b) < \infty$, $V_1^a = \sup_{s,m} E|M_s^{-1}X_sX_s^\top M_s^{-1}\varepsilon_{m+1}^2X_mX_m^\top|$, and $V_1^b = \sup_{s,m}$ $\int\int M_s^{-1}X_sX_s^\top M_s^{-1}\varepsilon_{m+1}^2X_mX_m^\top dF(X_s)dF(\varepsilon_{m+1},X_m)$. Note that the same result holds for the second case $(m > s)$ too. Hence we are left with

$$
C^{1(1)}_{\Theta_1} + C^{2(1)}_{\Theta_1} = \frac{1}{T^3h^2}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\mathrm{Tr}\{M_s^{-1}V(m/T)\}k^2\left(\frac{m-s}{Th}\right) + O\left(\frac{1}{T^2h}\right)
$$

$$= \frac{\nu_0}{Th} \int \mathrm{Tr}\{M_\tau^{-1} V(\tau)\} d\tau + o(1).$$

We consider V_{Θ_2} next with $s > m$ as the first case,

$$E(V_{\Theta_2}^1) = \frac{1}{T^3} \sum_{s=1}^{T} \sum_{m=s-\lfloor Th \rfloor}^{s-1} \mathrm{Tr}\{M_s^{-1} E(X_s X_s^\top) M_s^{-1} E(\varepsilon_{m+1} X_m^\top X_{2s-m} \varepsilon_{2s-m+1})\} k_{m,s} k_{2s-m,s}$$
$$+ \frac{1}{T^3} \sum_{s=1}^{T} \sum_{m=s-\lfloor Th \rfloor}^{s-1} \mathrm{Tr}\{E(M_s^{-1} X_s X_s^\top M_s^{-1} \varepsilon_{m+1} X_m^\top X_{2s-m} \varepsilon_{2s-m+1})$$
$$- M_s^{-1} E(X_s X_s^\top) M_s^{-1} E(\varepsilon_{m+1} X_m^\top X_{2s-m} \varepsilon_{2s-m+1})\} k_{m,s} k_{2s-m,s}$$
$$= C_{\Theta_2}^{1(1)} + C_{\Theta_2}^{1(2)}.$$

Since we are using data reflection, we have that $X_m = X_{2s-m}$ and $\varepsilon_{m+1} = \varepsilon_{2s-m+1}$. Hence, we will have $C_{\Theta_2}^{1(2)} = O(h^3/T)$. The result will be the same for case 2 where $m > s$.

So we have,

$$C_{\Theta_2}^{1(1)} + C_{\Theta_2}^{2(1)} = C_{\Theta_1}^{1(1)} + C_{\Theta_1}^{2(1)} = \frac{\nu_0}{Th} \int \mathrm{Tr}\{M_\tau^{-1} V(\tau)\} d\tau + o(1),$$

and we conclude $E(CV_{variance}) = \frac{2\nu_0}{Th} \int \mathrm{Tr}\{M_\tau^{-1} V(\tau)\} d\tau + O(\frac{h^3}{T})$.

For CV_{11}, note that when $m = \max\{s, n, m\}$, $E(CV_{11})$ is 0 by the law of iterated expectations. So we only have to consider the other 2 cases: (1) $m < s < n$ and (2) $n < m < s$. The derivation for case (1) is similar to the derivations above with CV_{bias}, and by noting that $E(\varepsilon_{m+1} X_m^\top) = 0$ we can show that it will yield an order of $O(h/T)$. Thus, for completeness, we show the second case. Consider the subcase of (2) where $s - m \geq m - n$ (i.e. $s - m = \bar{D}$). Here, we apply Lemma A.2 of Juhl and Xiao (2013) twice to obtain

$$E(CV_{11}^{2a}) = \frac{1}{2T^3} \sum_{\substack{s=1 \\ s-m=\bar{D}}}^{T} \sum_{m=s-\lfloor Th \rfloor+1}^{s-1} \sum_{n=s-\lfloor Th \rfloor}^{m-1} \mathrm{Tr}\{E(M_s^{-1} X_s X_s^\top M_s^{-1} X_n X_n^\top \beta_s'' \varepsilon_{m+1} X_m^\top)\} \left(\frac{n-s}{T}\right)^2$$
$$\times k_{m,s} k_{n,s}$$
$$\leq \frac{1}{2T^3} \sum_{\substack{n<m<s \\ s-m=\bar{D}}} \mathrm{Tr}\{M_s^{-1} E(X_s X_s^\top) M_s^{-1} E(X_n X_n^\top) \beta_s'' \underbrace{E(\varepsilon_{m+1} X_m^\top)}_{=0}\} \left(\frac{n-s}{T}\right)^2 k_{m,s} k_{n,s}$$
$$+ \frac{1}{2T^3} \sum_{\substack{n<m<s \\ s-m=\bar{D}}} R_1^{1/1+\delta} \beta^* (s-m)^{\delta/1+\delta} \left(\frac{n-s}{T}\right)^2 k_{m,s} k_{n,s} + \frac{1}{2T^3} \sum_{\substack{n<m<s \\ s-m=\bar{D}}} R_2^{1/1+\delta} \beta^* (s-m)^{\delta/1+\delta} \left(\frac{n-s}{T}\right)^2$$

$$\times k_{m,s}k_{n,s} = O\left(\frac{h}{T}\right)$$

where $R_1 = \max(R_1^a, R_1^b)$, $R_1^a = \sup_{s,m,n} \int M_s^{-1}X_sX_s^\top M_s^{-1}X_nX_n^\top \beta_s'' \varepsilon_{m+1}X_m^\top dF(X_m, X_s, X_n, \varepsilon_{m+1})$, $\sup_{s,m,n} \int\int M_s^{-1}X_sX_s^\top M_s^{-1}X_nX_n^\top \beta_s'' \varepsilon_{m+1}X_m^\top dF(X_m, \varepsilon_{m+1}, X_n)dF(X_s)$, $R_2 = \max(R_2^a, R_2^b)$, $R_2^a = \sup_{m,n} \int X_nX_n^\top \varepsilon_{m+1}X_m^\top dF(X_n, \varepsilon_{m+1}, X_m)$, and $R_2^b = \sup_{m,n} \int\int X_nX_n^\top \varepsilon_{m+1}X_m^\top dF(X_n)dF(\varepsilon_{m+1}, X_m)$. Furthermore, by repeating the same steps for the other cases (e.g. when s is the smallest index) and for CV_{12}, we will have that both $E(C_{11})$ and $E(C_{12})$ are $O(h/T)$, which are smaller orders. So we conclude that pointwise in h, $E(CV(h)) = IMSCFE(h)_L + o_p(T^{-4/5})$. □

1.5. Proof for Lemma 5

Before we begin the proof proper, we state and prove an intermediate result to deal with the random denominator in $CV(h)$.

Lemma A. *Define $\hat{M}_s(\tau) = X_sX_s^\top \frac{1}{h}k(\frac{s/T-\tau}{h})$ and recall that $M(t/T) = E[X_tX_t^\top]$, then*

$$\sup_{\tau\in[0,1]} \left\| T^{-1} \sum_{\substack{s=\tau T-\lfloor Th\rfloor \\ s\neq \tau T}}^{\tau T+\lfloor Th\rfloor} \hat{M}_s(\tau) - M(\tau) \right\| = O_p(h^2).$$

Proof. For convenience, treat the matrices $\hat{M}(\tau)$ and $M(\tau)$ as the $(m,n)^{th}$ element, so that they are scalar-valued. Cover $\tilde{\tau} \equiv [0,1]$ by L(T) intervals $I_k \equiv I_{k,T}$ of length $\ell_T = O(h^4)$. Note that $L(T) = O(\ell_T^{-1})$. Let τ_k be the midpoint of I_k. We have

$$\sup_{\tau\in[0,1]} \left| T^{-1} \sum_{\substack{s=\tau T-\lfloor Th\rfloor \\ s\neq \tau T}}^{\tau T+\lfloor Th\rfloor} \hat{M}_s(\tau) - M(\tau) \right| \leq \max_{1\leq j\leq L(T)} \sup_{\tau\tilde{\tau}\cap I_k} \left| T^{-1} \sum_{\substack{s=\tau T-\lfloor Th\rfloor \\ s\neq \tau T}}^{\tau T+\lfloor Th\rfloor} \hat{M}_s(\tau_k) - M(\tau_k) \right|$$
$$+ \max_{1\leq j\leq L(T)} \sup_{\tau\tilde{\tau}\cap I_k} T^{-1} \sum_{\substack{s=\tau T-\lfloor Th\rfloor \\ s\neq \tau T}}^{\tau T+\lfloor Th\rfloor} \left\{ |\hat{M}_s(\tau) - \hat{M}_s(\tau_k)| + |M(\tau) - M(\tau_k)| \right\}$$
$$\equiv O_1 + O_2.$$

It is straightforward to check that by the Lipschitz condition on the kernel and $M(\tau)$ (by assumption A.2(iii)) that $O_2 = O_p(h^2)$.

Define $\mathcal{B}_T = \eta_1(2Th)^{1/\delta}$, for $\eta_1 > 0$ and $\delta > 2$. Define the shorthand notation $\hat{\mathcal{M}}_s(\tau) = \hat{M}_s(\tau) - M(\tau)$. Then, using the truncation trick

$$\hat{\mathcal{M}}_s(\tau_k) = \hat{\mathcal{M}}_s(\tau_k)\mathbf{1}_{\{|X_sX_s^\top{}_{(m,n)}|\leq\mathcal{B}_T\}} + \hat{\mathcal{M}}_s(\tau_k)1_{\{|X_sX_s^\top{}_{(m,n)}|>\mathcal{B}_T\}} \equiv O_{3,s} + O_{4,s}.$$

Let τ^* be the maximizer in $\tilde{T} \cap I_k$, then for any $\epsilon > 0$ we have

$$P\bigg(T^{-1}\sum_{\substack{s=\tau^*T-\lfloor Th\rfloor\\ s\neq\tau^*T}}^{\tau^*T+\lfloor Th\rfloor} O_{4,s} > h^2\epsilon\bigg) \leq \sum_{\substack{s=\tau^*T-\lfloor Th\rfloor\\ s\neq\tau^*T}}^{\tau^*T+\lfloor Th\rfloor} P(|X_sX_s^\top{}_{(m,n)}| > \mathcal{B}_T) \leq \frac{E|X_sX_s^\top{}_{(m,n)}|^\delta}{(2\eta_1^\delta)} < \frac{\eta}{2},$$

by picking $\eta_1 > (E|X_sX_s^\top{}_{(m,n)}|^\delta)^{1/\delta}\eta^{-1/\delta}$, and an arbitrarily small η. For $O_{3,s}$, note that $|Q_{3,s}| \leq C_1\mathcal{B}_T/h$ and $Var(Q_{3,s}) \leq C_1/h$, for a $C_1 > 0$. Then, we use a Bernstein inequality (Merlevède et al., 2009) for mixing sequences that need not be stationary,

$$\begin{aligned}
P\bigg(T^{-1}\sum_{\substack{s=\tau^*T-\lfloor Th\rfloor\\ s\neq\tau^*T}}^{\tau^*T+\lfloor Th\rfloor} Q_{3,s} > h^2\epsilon\bigg) &\leq \exp\bigg(-\frac{C(Th^2\epsilon)^2}{v^2(2Th) + (C\mathcal{B}_T/h)^2 + (Th^2\epsilon)(C\mathcal{B}_T/h)(\log\lfloor Th\rfloor)^2}\bigg)\\
&= \exp\bigg(-\frac{C\epsilon^2}{\frac{2hv^2}{Th^4\epsilon^2} + \frac{C^2(2Th)^{2/\delta}}{T^2h^6\epsilon^2} + \frac{(C(2Th)^{1/\delta}/h)(\log\lfloor Th\rfloor)^2}{Th^2\epsilon}}\bigg)\\
&= o(1),
\end{aligned}$$

where $C > 0$ and $v^2 = \sup_s(Var(Q_{3,s}) + 2\sum_{s^*>s}|Cov(Q_{3,s}, Q_{3,s^*})|) = O(h^{-1})$. The last equality is attained by first observing that $v^2 = O(h^{-1})$ so that $\frac{2hv^2}{Th^4\epsilon} = o(1)$. Next, note that $\frac{C^2(2Th)^{2/\delta}}{T^2h^6\epsilon^2} = \frac{C^22^{2/\delta}}{T^{8/5\delta-4/5}\epsilon^2} = o(1)$ since $\delta > 2$. Likewise, for any $\delta > 2$, we can show that the last term in the denominator goes to 0. Hence, concluding the proof. □

Proof of (i)

Recall that $\beta_s = \hat{\beta}_s + D_s + \bar{R}_s$ and $D_s - \hat{D}_s = (S_{0s}^{-1} - M_s^{-1})T^{-1}\sum_{\substack{l=s-\lfloor Th\rfloor\\ l\neq s}}^{s+\lfloor Th\rfloor} k_{ls}X_l\varepsilon^*_{l,s}$ from the proof of Lemma 3. Note that $S_{0s}^{-1} - M_s^{-1} = -M_s^{-1}(S_{0s} - M_s)M_s^{-1} + O(\|S_{0s} - M_s\|^2)$, which implies that, uniformly in s, the leading order of $S_{0s}^{-1} - M_s^{-1}$ is h^2 by Lemma A.

$$\begin{aligned}
CV(h) - \hat{\tilde{CV}}(h) &= T^{-1}\sum_{s=1}^{T}(\beta_s - \hat{\beta}_s + \hat{D}_s)^\top X_sX_s^\top(\beta_s - \hat{\beta}_s + \hat{D}_s) + 2T^{-1}\sum_{s=1}^{T}\varepsilon_{s+1}X_s^\top(\beta_s - \hat{\beta}_s + \hat{D}_s)\\
&= T^{-1}\sum_{s=1}^{T}(-D_s - \bar{R}_s + \hat{D}_s)^\top X_sX_s^\top(-D_s - \bar{R}_s - \hat{D}_s) + 2T^{-1}\sum_{s=1}^{T}\varepsilon_{s+1}X_s^\top(-D_s - \bar{R}_s + \hat{D}_s)\\
&= T^{-1}\sum_{s=1}^{T}(D_s - \hat{D}_s)^\top X_sX_s^\top(D_s + \hat{D}_s) + T^{-1}\sum_{s=1}^{T}(D_s - \hat{D}_s)^\top X_sX_s^\top\bar{R}\\
&+ T^{-1}\sum_{s=1}^{T}\bar{R}^\top X_sX_s^\top(D_s + \hat{D}_s) + T^{-1}\sum_{s=1}^{T}\bar{R}_s^\top X_sX_s^\top\bar{R}_s
\end{aligned}$$

$$
- 2T^{-1}\sum_{s=1}^{T}\varepsilon_{s+1}X_s^\top(D_s-\hat{D}_s) - 2T^{-1}\sum_{s=1}^{T}\varepsilon_{s+1}X_s^\top\bar{R}_s
$$
$$
\equiv Y_1+Y_2+Y_3+Y_4+Y_5+Y_6.
$$

Based on the argument of Lemma 3, we know that Y_1 contains the leading term and Y_2 to Y_6 are smaller orders. So we check Y_1,

$$
\begin{aligned}
Y_1 &= T^{-1}\sum_{s=1}^{T}\Bigg[T^{-1}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\varepsilon^*_{m,s}k_{ms}X_m^\top(S_{0s}^{-1}-M_s^{-1})^\top\Bigg]X_sX_s^\top\Bigg[2T^{-1}M_s^{-1}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor}\varepsilon^*_{n,s}k_{ns}X_n \\
&\quad + (S_{0s}^{-1}-M_s^{-1})T^{-1}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor}\varepsilon^*_{n,s}k_{ns}X_n\Bigg] \\
&= 2T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor}\varepsilon^*_{m,s}k_{ms}X_m^\top(S_{0s}^{-1}-M_s^{-1})^\top X_sX_s^\top M_s^{-1}X_nk_{ns}\varepsilon^*_{n,s} \\
&\quad + T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor}\varepsilon^*_{m,s}k_{ms}X_m^\top(S_{0s}^{-1}-M_s^{-1})^\top X_sX_s^\top(S_{0s}^{-1}-M_s^{-1})X_nk_{ns}\varepsilon^*_{ns} \\
&= Y_{11}+Y_{12}.
\end{aligned}
$$

Since $\varepsilon_{l,s}=X_l^\top Q^*_{l,s}+\varepsilon_{l+1}$, consider the first cross product,

$$
\begin{aligned}
Y_{11}^{(1)} &\equiv 2T^{-3}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor}\beta_s''^\top X_mX_m^\top(S_{0s}^{-1}-M_s^{-1})^\top X_sX_s^\top M_s^{-1}X_nX_n^\top\beta_s'' \\
&\quad\times\left(\frac{m-s}{T}\right)^2\left(\frac{n-s}{T}\right)^2k_{ms}k_{ns}.
\end{aligned}
$$

So

$$
\begin{aligned}
E(Y_{11}^{(1)^2}) &= 4T^{-6}\sum_{s,\dots,c}E\Bigg[\beta_s''^\top V_m(S_{0s}^{-1}-M_s^{-1})^\top V_sM_s^{-1}V_n\beta_s''\beta_a''^\top V_b(S_{0a}^{-1}-M_a^{-1})^\top V_aM_a^{-1}V_c\beta_a'' \\
&\times\left(\frac{m-s}{T}\right)^2\left(\frac{n-s}{T}\right)^2\left(\frac{b-a}{T}\right)^2\left(\frac{c-a}{T}\right)^2k_{ms}k_{ns}k_{ba}k_{ca}\Bigg].
\end{aligned}
$$

By our usual argument using mixing inequalities, we can show that $E(Y_{11}^{(1)^2})=O(T^{-12/5})$, likewise for the other cross product terms from Y_{11}. So we conclude that $E[(CV(h)-\hat{\tilde{CV}}(h))^2]=O(T^{-12/5})$.

Proof of (ii)-(vi)

We state some intermediate results that would be convenient for computation:

$$
\begin{aligned}
|k_{ls}(h^*) - k_{ls}(\tilde{h}_j)| &= \left|\frac{1}{h^*}k\left(\frac{l-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{l-s}{T\tilde{h}_j}\right)\right| \\
&\leq \left|\frac{1}{h^*}k\left(\frac{l-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{l-s}{Th^*}\right)\right| + \left|\frac{1}{\tilde{h}_j}k\left(\frac{l-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{l-s}{T\tilde{h}_j}\right)\right|
\end{aligned}
$$

where

$$
\left|\frac{1}{h^*}k\left(\frac{l-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{l-s}{Th^*}\right)\right| \leq \left|\frac{\tilde{h}_j - h^*}{h^*\tilde{h}_j}\right| k(0)
$$

and

$$
\begin{aligned}
\left|\frac{1}{\tilde{h}_j}k\left(\frac{l-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{l-s}{T\tilde{h}_j}\right)\right| &= \frac{1}{\tilde{h}_j}\left|k\left(\frac{l-s}{Th^*}\right) - k\left(\frac{l-s}{T\tilde{h}_j}\right)\right| \\
&\leq \frac{C}{\tilde{h}_j}\left|\frac{l-s}{Th^*} - \frac{l-s}{T\tilde{h}_j}\right| \\
&= \frac{C}{\tilde{h}_j}\frac{|\tilde{h}_j - h^*||l-s|}{Th^*\tilde{h}_j}.
\end{aligned}
$$

So

$$
|k_{ls}(h^*) - k_{ls}(\tilde{h}_j)| \leq \frac{C_1}{\tilde{h}_j}\underbrace{\left[\frac{|\tilde{h}_j - h^*|}{h^*} + \frac{|\tilde{h}_j - h^*||l-s|}{Th^*\tilde{h}_j}\right]}_{=O(T^{-3/5})}, \tag{3}
$$

since $|\tilde{h}_j - h^*| \leq l_T = O(T^{-4/5})$.

For the squared version, add and subtract twice:

$$
\begin{aligned}
&|k_{ms}(h^*)k_{ns}(h^*) - k_{ms}(\tilde{h}_j)k_{ns}(\tilde{h}_j)| \\
&\leq \frac{1}{\tilde{h}_j}\frac{\tilde{h}_j}{h^*}k\left(\frac{m-s}{Th^*}\right)\left|\frac{1}{h^*}k\left(\frac{n-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{n-s}{T\tilde{h}_j}\right)\right| + \frac{1}{\tilde{h}_j}k\left(\frac{n-s}{T\tilde{h}_j}\right)\left|\frac{1}{h^*}k\left(\frac{m-s}{Th^*}\right) - \frac{1}{\tilde{h}_j}k\left(\frac{m-s}{T\tilde{h}_j}\right)\right| \\
&\leq \frac{1}{\tilde{h}_j^2}\left[\frac{\tilde{h}_j}{h^*} + 1\right]k(0)O(T^{-3/5}) = \tilde{h}_j^{-2}O(T^{-3/5}). \qquad (4)
\end{aligned}
$$

We start with the proof of (ii) $E[(CV_{bias}(\tilde{h}_j) - CV_{bias}(h^*))^2]$. Again, we use the notation $M_s^{-1}X_sX_s^\top M_s^{-1} = U_s$, and $X_mX_m^\top = V_m$. Then,

$$
(CV_{bias}(\tilde{h}_j) - CV_{bias}(h^*))^2
$$

$$= T^{-6}\sum_{s=1}^{T}\sum_{\substack{m=s-\lfloor Th\rfloor\\ m\neq s}}^{s+\lfloor Th\rfloor}\sum_{\substack{n=s-\lfloor Th\rfloor\\ n\neq s}}^{s+\lfloor Th\rfloor}\sum_{a=1}^{T}\sum_{\substack{b=a-\lfloor Th\rfloor\\ b\neq a}}^{a+\lfloor Th\rfloor}\sum_{\substack{c=a-\lfloor Th\rfloor\\ c\neq a}}^{a+\lfloor Th\rfloor}\beta_s^{''\top}V_mU_sV_n\beta_s^{''}\beta_a^{''\top}V_bU_aV_c\beta_s^{''}$$
$$\times\left(\frac{m-s}{T}\right)^2\left(\frac{n-s}{T}\right)^2\left(\frac{b-a}{T}\right)^2\left(\frac{c-a}{T}\right)^2[k_{ms}(\tilde{h}_j)k_{ns}(\tilde{h}_j)-k_{ms}(h^*)k_{ns}(h^*)][k_{ba}(\tilde{h}_j)k_{ca}(\tilde{h}_j)$$
$$-k_{ba}(h^*)k_{ca}(h^*)]$$

Define $G_{msn}=\beta_s^{''\top}V_mU_sV_n\beta_s^{''}$ and $\tilde{K}_{msn}=k_{ms}(\tilde{h}_j)k_{ns}(\tilde{h}_j)-k_{ms}(h^*)k_{ns}(h^*)$. Consider the following case where $m<s<n<b<a<c$, and $n-s=\bar{D}$ is the maximum adjacent distance. Note that this can be generalized to other cases. By adding and subtracting terms, we get

$$E[(CV_{bias}(\tilde{h}_j)-CV_{bias}(h^*))^2]$$
$$=T^{-6}\sum_{m,\dots,c}\left\{E[G_{msn}G_{bac}]-E[G_{msn}]E[G_{bac}]\right\}\left(\frac{m-s}{T}\right)^2\dots\left(\frac{c-a}{T}\right)^2\tilde{K}_{msn}\tilde{K}_{bac}$$
$$+\dots+T^{-6}\sum_{m,\dots,c}\beta_s^{''\top}M_s^{-1}\beta_s^{''}\beta_a^{''\top}M_a^{-1}\beta_a^{''}\left(\frac{m-s}{T}\right)^2\dots\left(\frac{c-a}{T}\right)^2\tilde{K}_{msn}\tilde{K}_{bac}.$$

The last term is the leading term because we can use the mixing inequality on all the other terms as we have done in other proofs. Label the last term EC_1, and see that for a constant C,

$$EC_1\le\tilde{h}_j^4T^{-6}\sum_{m,\dots,c}C\left(\frac{m-s}{T\tilde{h}_j}\right)^2\dots\left(\frac{c-a}{T\tilde{h}_j}\right)^2O(T^{-6/5})=O(h^8\times T^{-6/5})=O(T^{-14/5}),$$

where we have used (4).

For (iii), define $\varepsilon_{m+1}X_m^\top=W_m^\top$, and we have

$$E[(CV_{variance}(\tilde{h}_j)-CV_{variance}(h^*))^2]=T^{-6}\sum_{m,\dots,c}\left\{E[W_m^\top U_sW_nW_b^\top U_aW_c]\right\}\tilde{K}_{msn}\tilde{K}_{bac}.$$

Note that in the following cases where: (1) $m\neq n$ or $b\neq c$, (2) $m\neq b$ or $n\neq c$, and (3) $m\neq c$ or $n\neq b$, we can see that the expression will be equivalent to 0 by the law of iterated expectations. So we take $m=n$ and $b=c$ for granted, and consider $m<s<b<a$ with $m-s=\bar{D}$ as the maximal adjacent distance. Then

$$E[(CV_{variance}(\tilde{h}_j)-CV_{variance}(h^*))^2]$$
$$=T^{-6}\sum_{m,s,b,a}\left\{E[\varepsilon_{m+1}^2X_m^\top U_sX_m\varepsilon_{b+1}^2X_b^\top U_aX_b]-E[\varepsilon_{m+1}^2X_m^\top U_sX_m]E[\varepsilon_{b+1}^2X_b^\top U_aX_b]\right\}\tilde{K}_{msm}\tilde{K}_{bab}$$

$$+\ldots+T^{-6}\sum_{m,s,b,a}\mathrm{Tr}\{M_s^{-1}V(m/T)M_a^{-1}V(b/T)\}\tilde{K}_{msm}\tilde{K}_{bab},$$

where we recall that $V(t/T)=E[\varepsilon_{t+1}X_tX_t^\top]$. Again, we focus on the final term (label it EC_2), and observe that

$$EC_2\le CT^{-6}\tilde{h}_j^{-4}\sum_{m,s,b,a}O(T^{-6/5})=O(T^{-2}\times h^{-2}\times T^{-6/5})=O(T^{-14/5}).$$

Likewise, we can apply similar steps to (iv) and (v) to yield the orders $O(T^{-14/5})$. A similar calculation for (vi) yields $O(T^{-3})$. □

1.6. Proof for Lemma 6

Proof. This proof is approached in a similar manner to the uniform consistency proof in Lemma A. For notational convenience, we define $A_T\equiv\sqrt{\frac{\log h^{-1}}{Th}}$. Next, cover the compact space $\tilde{T}\equiv[0,1]$ by $L(T)$ intervals $I_k\equiv I_{k,T}$ of length $l_T=A_Th^2$. Thus we have $L(T)=O(l_T^{-1})$. Let τ_k be the midpoint of I_k. Then,

$$\begin{aligned}
&\max_{1\le j\le p_T+1}\sup_{\tau\in[0,1]}\left|T^{-1}\sum_{\substack{s=\tau T-\lfloor Th\rfloor\\ s\ne\tau T}}^{\tau T+\lfloor Th\rfloor}Z_{sj}(\tau,l)\right|\le\max_{1\le j\le p_T+1}\max_{1\le k\le L(T)}\sup_{\tau\in\tilde{T}\cap I_k}\left|T^{-1}\sum_{\substack{s=\tau T-\lfloor Th\rfloor\\ s\ne\tau T}}^{\tau T+\lfloor Th\rfloor}Z_{sj}(\tau_k,l)\right|\\
&+\max_{1\le j\le p_T+1}\max_{1\le k\le L(T)}\sup_{\tau\in\tilde{T}\cap I_k}T^{-1}\sum_{\substack{s=\tau T-\lfloor Th\rfloor\\ s\ne\tau T}}^{\tau T+\lfloor Th\rfloor}\left|Z_{sj}(\tau,l)-Z_{sj}(\tau_k,l)\right|\\
&\equiv Z_1+Z_2.
\end{aligned}$$

By the Lipschitz condition on the kernel, we can show that $Z_2=O_p(l_T/h^2)=O_p(A_T)$.

For Z_1, we use truncation. Define $B_T=\xi_1(2Th(p_T+1))^{1/\psi}$ where $\xi_1>0$ and $\psi>2+\frac{\delta_1}{1-\delta_2}+\delta$, and

$$\begin{aligned}
\tilde{Z}_{sj}(\tau_k,l)&=Z_{sj}(\tau_k,l)-Z_{sj}(\tau_k,l)\mathbf{1}_{|H_sX_{sj}|\le B_T}\\
&\equiv Z_{sj}(\tau_k,l)-Z_{sj}^{\mathbf{1}}(\tau_k,l)
\end{aligned}$$

where $H_s\equiv y_{s+1}-\alpha_0(s/T)^\top X_s$. Let τ^* be the maximizer in $\tilde{T}\cap I_k$ and we get

$$Z_1\le\max_{1\le j\le p_T+1}\max_{1\le k\le L(T)}\left|T^{-1}\sum_{\substack{s=\tau^*T-\lfloor Th\rfloor\\ s\ne\tau^*T}}^{\tau^*T+\lfloor Th\rfloor}[Z_{sj}^{\mathbf{1}}(\tau_k,l)-E(Z_{sj}^{\mathbf{1}}(\tau_k,l))]\right|$$

$$+\max_{1\leq j\leq p_T+1}\max_{1\leq k\leq L(T)}\left|T^{-1}\sum_{\substack{s=\tau^*T-\lfloor Th\rfloor\\ s\neq\tau^*T}}^{\tau^*T+\lfloor Th\rfloor}[\tilde{Z}_{sj}(\tau_k,l)-E(\tilde{Z}_{sj}(\tau_k,l))]\right|$$

$$\equiv Z_3+Z_4$$

The proof for Z_4 is the same as the proof in Li et al. (2015), which implies that $Z_4=O_p(A_T)$.

Subsequently, define $Q_s\equiv Z^{\mathbf{1}}_{sj}(\tau_k,l)-E(Z^{\mathbf{1}}_{sj}(\tau_k,l))$. Note that $|Q_s|\leq C_1B_T/h$ since if the indicator function equals one then $|H_s(\tau_k)X_{sj}|\leq B_T$. The constant C_1 is used to bind the kernel and the non-random time indices. Furthermore, note that $Var(Z^{\mathbf{1}}_{sj}(\tau_k,l))\leq C_1/h$.

Next, we use Theorem 2 of Merlevède et al. (2009), which provides a Bernstein inequality for mixing sequences that are not necessarily stationary, to bind the tail probability of Z_3:

$$\begin{aligned}P(Z_3>\xi A_T)&\leq\exp\left(-\frac{C(T\xi A_T)^2}{v^2(2Th)+(C_1B_T/h)^2+(T\xi A_T)(C_1B_T/h)(\log[2Th])^2}\right)\\&=\exp\left(-\frac{C\xi^2\log h^{-1}}{v^2(2h^2)+C_1^2B_T^2/(Th)+\xi C_1A_TB_T(\log[2Th])^2}\right)\\&=o(1),\end{aligned}$$

where $v^2=\sup_s(Var(Q_s)+2\sum_{s^*>s}|Cov(Q_s,Q_{s^*})|)=O(h^{-1})$. We arrive at the last line by noting that $v^2(2h^2)\to0$, and by construction of ψ we get $B_T^2/(Th)\to0$.

Also, we have that $A_TB_T(\log[2Th])^2=\sqrt{\frac{\log h^{-1}(2Th(p_T+1)^{2/\psi})(\log(2Th))^4}{Th}}\to0$, which can be obtained if we choose a sufficiently large ψ. Hence, we conclude that $Z_1=O_p(A_T)$.

□

2. Additional simulations

We consider 3 modifications to the DGP introduced in section 7.1. In particular, we look at the following cases,

1. No bias and constant weights: $y_{t+1}=0.6f_{1,t}+0.3f_{2,t}+\varepsilon_{t+1}$.
2. Constant bias and weights: $y_{t+1}=0.3+0.6f_{1,t}+0.3f_{2,t}+\varepsilon_{t+1}$
3. Same as original, but without the time-varying bias or intercept.

The results are presented in Table 1 below. The results suggest that NPRf performs better when the DGP does indeed possess time-varying weights, while least squares-regression based methods perform better otherwise. Although, NPRf tends to be a close contender even in those cases. The inclusion or exclusion of a bias does not appear to affect the results.

Table 1: Simulation results for alternative cases with low-dimensional data (2 forecasts).

	Case 1			Case 2			Case 3		
Sample size (T):	200	300	500	200	300	500	200	300	500
Adaptive									
NPRf	1.02	1.01	0.99	1.02	1.01	0.99	1.03	1.03	1.01
	(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.21)
BG	1.11	1.11	1.10	1.06	1.07	1.05	1.05	1.06	1.05
	(0.22)	(0.21)	(0.23)	(0.22)	(0.21)	(0.22)	(0.22)	(0.20)	(0.22)
TVGRregconst	1.01	1.01	0.99	1.01	1.01	0.99	1.04	1.04	1.02
	(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.20)	(0.22)	(0.20)	(0.21)
TVGRreg	1.01	1.01	0.99	1.04	1.04	1.02	1.03	1.03	1.02
	(0.20)	(0.19)	(0.20)	(0.21)	(0.20)	(0.21)	(0.21)	(0.20)	(0.21)
TVGRregconstr	1.11	1.11	1.09	1.06	1.06	1.04	1.05	1.05	1.04
	(0.22)	(0.21)	(0.23)	(0.22)	(0.21)	(0.22)	(0.22)	(0.20)	(0.22)
Static									
GRregconst	1.02	1.01	0.99	1.02	1.01	0.99	1.04	1.04	1.02
	(0.21)	(0.20)	(0.20)	(0.21)	(0.20)	(0.20)	(0.22)	(0.20)	(0.21)
GRreg	1.01	1.01	0.99	1.04	1.04	1.02	1.04	1.04	1.02
	(0.20)	(0.19)	(0.20)	(0.21)	(0.20)	(0.21)	(0.22)	(0.20)	(0.21)
GRregconstr	1.11	1.11	1.09	1.06	1.06	1.04	1.06	1.06	1.04
	(0.22)	(0.21)	(0.23)	(0.22)	(0.21)	(0.22)	(0.22)	(0.20)	(0.22)
EQ	1.14	1.14	1.14	1.12	1.12	1.11	1.07	1.08	1.08
	(0.23)	(0.22)	(0.24)	(0.23)	(0.22)	(0.23)	(0.22)	(0.21)	(0.22)
Best	TVGR-reg	TVGR-reg	TVGR-reg	TVGR-regconst	TVGR-regconst	TVGR-regconst	NPRf	NPRf	NPRf

Notes: (1) Mean of ASCFE and standard deviation in parentheses from 500 iterations for cases 1 to 3. (2) NPRf: nonparametric estimator with data reflection, BG: Bates and Granger (1969) adaptive etimator, GRregconst: OLS regression with intercept, GRreg: OLS regression without intercept, GRregconstr: OLS regression with sum of coefficients constrained to unity, TV: time-varying weights, EQ: equal weights.

3. Data description for empirical applications

3.1. Forecasting inflation

Following Ang et al. (2007), we generate forecasts from an ARMA(1,1) model; a forecast of annual inflation based on a regression on lagged inflation and GDP growth (PC1) from the Phillips curve framework; forecasts derived from lagged inflation, GDP growth and the short rate (TS1) as a term structure-implied forecast; and the SPF for a survey-based forecast. All forecasts are generated with an expanding window. We consider three CPI-based measures of inflation: CPI for all urban consumers (PUNEW), CPI less housing and shelter (PUXHS), and CPI less food and energy (PUXX). A summary of the models and the inflation indices are presented in table 2.

Annual inflation is defined as $\pi_{t+4,t} = 0.25 \sum_{k=1}^{4} \pi_{t+k}$, where π_t is the log difference of quarterly indices. The ARMA-implied forecasts of annual inflation is generated from the sum of one to four-steps ahead predictions. For PC1 and TS1, the lag polynomial $\beta(L)$ in table 2 is determined by conventional BIC at each time period, and the same lag order is assumed for all regressors. The forecasts of annual inflation are constructed in an adaptive manner as in the simulations. Since participants of the SPF are asked to forecast average quarterly changes in the CPI for all urban consumers, a forecast of annual inflation, as detailed in table 2, is obtained in a similar fashion to that of realized annual inflation.

The SPF began forecasting inflation in 1981Q3, so we generate the first forecasts from the other three methods with data from 1970Q1 to 1981Q2, and allow this window to gradually expand. This results in a series of forecasts spanning from 1981Q3 to 2018Q2. Note that at the final time point, the SPF forecast, by construction, uses quarterly SPF data till 2019Q1 and is a forecast of annual inflation in 2019Q2. We consider twelve quarters (or three years) for the pseudo out-of-sample period from 2015Q3 to 2018Q2. We use 1998Q3 to 2015Q2 as the estimation sample, and because the nonparametric estimator require a holdout period for CV, we use 1981Q1 to 1998Q2 for that purpose.

Table 2: Summary of the inflation indices and models.

Infl. index		Description	Data source (Mnemonic)
PUNEW		CPI all urban	St. Louis FRED (CPIAUCSL)
PUXHS		CPI less housing and shelter	St. Louis FRED (CUSR0000SA0L2)
PUXX		CPI less food and energy	St. Louis FRED (CPILFESL)
Model	**Category**	**Description**	**Data source**
ARMA(1,1)	Time series	On quarterly inflation $\pi_{t+1} = \alpha + \gamma\pi_t + \phi\varepsilon_t + \varepsilon_{t+1}$	
PC1	Phillips curve	$\pi_{t+4,t} = \mu + \beta(L)^\top X_t + u_{t+4,t}$ $X_t^\top = [\pi_t, GDPG_t]^\top$	St. Louis FRED (GDPC1)
TS1	Term structure	same as PC1 $X_t^\top = [\pi_t, GDPG_t, RATE_t]^\top$	St. Louis FRED (GDPC1) CRSP risk-free rates 3 Month
SPF	Survey	$f_t^S = 0.25 \sum_{k=0}^{3} f_{t+k}^q$ f_t^q: quarterly median forecast	Federal Reserve Bank of Philadelphia

Notes: (1) $\pi_{t+4,t}$ is annual inflation while π_t refers to quarterly inflation. (2) Naming convention follows Ang et al. (2007).

3.2. Predictability of stock returns

We use 13 predictors following Rapach et al. (2010): (i) Dividend-price ratio (D/P): the log difference between the 12-month moving sum of dividends and the index; (ii) Dividend-yield ratio (D/Y): the log difference between dividends and the lagged index; (iii) Earnings-price ratio (E/P): the log difference between the 12-month moving sum of earnings on the S&P 500 index and stock prices; (iv) Stock variance (SVAR): the sum of squared daily returns on the index; (v) Book-to-market ratio (B/M): the ratio of the book value to the market value for the Dow Jones Industrial Average; (vi) Net equity expansion (NTIS): the ratio of 12-month moving sum of net issues to total end-of-year market capitalization for NYSE-listed stocks; (vii) Treasury bill rate (TBL): the 3-month Treasury bill rate (secondary market); (viii) Long-term return (LTR): the return on long-term government bonds; (ix) Term spread (TMS): the difference between long-term yield and the Treasury bill rate; (x) Default yield spread (DFY): the difference in yields between BAA- and AAA-rated corporate bond; (xi) Default return spread (DFR): the difference between long-term corporate bond yields and long-term government bond returns; (xii) Inflation (INFL): the CPI inflation for all urban consumers; (xiii) Investment-to-capital ratio (I/K): the ratio of fixed investment to capital for the whole economy[1].

We investigate quarterly data starting in 1947Q2. Using the initial period of 1947Q2 to 1965Q1 (18 years), we generate out-of-sample forecasts of stock return by running univariate regressions on the 13 predictors. We use an expanding window and produce individual forecasts from 1965Q2 to 2018Q3. We require a holdout period from 1965Q2 to 1986Q3 for the nonparametric estimator and use the period from 1986Q4 to 2013Q3 as our estimation sample. The out-of-sample evaluation is a 5 year period that spans from 2013Q4 to 2018Q3.

[1]Following Elliott et al. (2013), we exclude the dividend-payout ratio and the long term yield to avoid multicollinearity.

References

Ang, A., Bekaert, G., Wei, M., 2007. Do macro variables, asset markets, or surveys forecast inflation better? Journal of monetary Economics 54, 1163–1212.

Bates, J.M., Granger, C.W., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

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

Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. Journal of Econometrics 177, 357–373.

Juhl, T., Xiao, Z., 2013. Nonparametric tests of moment condition stability. Econometric Theory 29, 90–114.

Li, D., Ke, Y., Zhang, W., 2015. Model selection and structure specification in ultra-high dimensional generalised semi-varying coefficient models. The Annals of Statistics 43, 2676–2705.

Merlevède, F., Peligrad, M., Rio, E., et al., 2009. Bernstein inequality and moderate deviations under strong mixing conditions. Institute of Mathematical Statistics.