Online Supplement:

Optimal bandwidth selection for forecasting under parameter instability

NOT FOR PUBLICATION

Yu Bai   Bin Peng   Shuping Shi   Wenying Yao

The Online Supplement is organized as follows. Section S1 provides the proof of Lemma 3. Section S2 reports additional simulation results for the structural break case. Section S3 details the implementation of the forecast combination methods. Section S4 presents an empirical application on real-time inflation forecasting using financial variables.

S1 Proof of Lemma 3

Given the kernel function $\overline{K}$, write $\hat{\theta}_{\overline{K},b,T} = \hat{\theta}_{b,T}$. As in (A.3), the estimator can be decomposed as

$$\begin{aligned}\hat{\theta}_{b,T} - \theta_T &= -H_T(\theta_T)S_T(\theta_T) + o_p(1) \\ &= -H_T(\theta_T)\left(S_T(\theta_t) + B_T\right) + o_p(1),\end{aligned} \tag{S.1}$$

where

$$S_T(\theta_t) = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial \ell_{t,T}\left(\theta_t\right)}{\partial\theta}, H_T(\theta_T) = \left(\frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2 \ell_{t,T}\left(\theta_T\right)}{\partial\theta\partial\theta'}\right)^{-1},$$
$$B_T = \frac{1}{Tb}\sum_{t=1}^{T} k_{tT}\frac{\partial^2 \ell_{t,T}\left(\overline{\theta}_T\right)}{\partial\theta\partial\theta'}\left(\theta_T - \theta_t\right),$$

and $\overline{\theta}_T$ lies between θ_T and θ_t. We will show that

$$\sup_{b\in I_T} \left\|T^{1/2}b^{1/2+\delta}S_T(\theta_t)\right\| = O_p(1), \tag{S.2}$$

$$\sup_{b\in I_T} \left\|H_T(\theta_T)^{-1}\right\| = O_p(1), \tag{S.3}$$

$$\sup_{b\in I_T} \left\|b^{\delta}B_T\right\| = O_p(b), \tag{S.4}$$

for some $0 < \delta < 1/2$. These bounds together with (S.1) prove (A.13).

Proof of (S.2). By Boole's inequality and Chebyshev's inequality, we have, for any $\varepsilon > 0$,

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

where the third inequality follows from the proof of (A.5) since $\left\|\frac{1}{T^{1/2}b^{1/2}}\sum_{t=1}^{T}k_{tT}\frac{\partial\ell_{t,T}(\theta_t)}{\partial\theta}\right\| = O_p(1)$. The final equality follows from Assumption 6.

Proof of (S.3). Recall that

$$\begin{aligned}\tilde{H}_T =&\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\frac{\partial^2\tilde{\ell}_{1,t}\left(\theta_T\right)}{\partial\theta\partial\theta'}\\ =&\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}E\left[\frac{\partial^2\tilde{\ell}_{1,t}\left(\theta_T\right)}{\partial\theta\partial\theta'}\right] + \frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\left(\frac{\partial^2\tilde{\ell}_{1,t}\left(\theta_T\right)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2\tilde{\ell}_{1,T}\left(\theta_T\right)}{\partial\theta\partial\theta'}\right]\right)\\ :=&\tilde{H}_{T,1}\left(I_k+\tilde{\Delta}_T\right),\end{aligned} \tag{S.5}$$

where $\tilde{\Delta}_T = \left(\tilde{H}_{T,1}\right)^{-1}\left(\tilde{H}_T - \tilde{H}_{T,1}\right)$. First, (A.4) holds uniformly over b:

$$\sup_{b\in I_T}\left\|\tilde{H}_{T,1}^{-1}\right\|_{sp} = O_p(1). \tag{S.6}$$

For $\tilde{\Delta}_T$, let $\tilde{\Delta}_t = \frac{\partial^2\tilde{\ell}_{1,t}(\theta_T)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2\tilde{\ell}_{1,T}(\theta_T)}{\partial\theta\partial\theta'}\right]$. Then, for any $\varepsilon > 0$, , by Boole's inequality and

Chebyshev's inequality, we have

$$\mathbb{P}\left(\sup_{b\in I_T}\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\|>\varepsilon\right)\leqslant\sum_{b\in I_T}\mathbb{P}\left(\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\|>\varepsilon\right)$$
$$\leqslant\underbrace{|I_T|}_{O(b^{\delta})}\times\underbrace{\sup_{b\in I_T}\mathbb{P}\left(\left\|\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\tilde{\Delta}_t\right\|>\varepsilon\right)}_{o(1)}=o(1). \tag{S.7}$$

To sum up, we continue from (A.7):

$$\sup_{b\in I_T}\left\|\tilde{H}_T^{-1}\right\|_{sp}\leqslant\underbrace{\sup_{b\in I_T}\left\|\tilde{H}_{T,1}^{-1}\right\|_{sp}}_{O_p(1)\text{ by (S.6)}}\left(1-\underbrace{\sup_{b\in I_T}\left\|\tilde{\Delta}_T\right\|_{sp}}_{o_p(1)\text{ by (S.7)}}\right)^{-1}=O_p(1).$$

This also implies (S.3).

Proof of (S.4). Recall that the stationary approximation of B_T is $\tilde{B}_T$, where $\tilde{B}_T=\tilde{B}_{T,1}+\tilde{B}_{T,2}$:

$$\tilde{B}_{T,1}=\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}\left(\frac{\partial^2\tilde{\ell}_{1,T}(\theta_T)}{\partial\theta\partial\theta'}-E\left[\frac{\partial^2\tilde{\ell}_{1,T}(\theta_T)}{\partial\theta\partial\theta'}\right]\right)(\theta_T-\theta_t),$$
$$\tilde{B}_{T,2}=\frac{1}{Tb}\sum_{t=1}^{T}k_{tT}E\left[\frac{\partial^2\tilde{\ell}_{1,T}(\theta_T)}{\partial\theta\partial\theta'}\right](\theta_T-\theta_t).$$

For $\tilde{B}_{T,1}$, again, similarly as in (S.2), we have

$$\mathbb{P}\left(\sup_{b\in I_T}\left\|\tilde{B}_{T,1}\right\|>\varepsilon\right)\leqslant\sum_{b\in I_T}\mathbb{P}\left(\left\|\tilde{B}_{T,1}\right\|>\varepsilon\right)\leqslant|I_T|\times\sup_{b\in I_T}\mathbb{P}\left(\left\|\tilde{B}_{T,1}\right\|>\varepsilon\right)=o(1).$$

Moving to $\tilde{B}_{T,2}$, since for some $0<\delta<1/2$, we have

$$\mathbb{P}\left(\sup_{b\in I_T}\left\|b^{\delta}\tilde{B}_{T,2}\right\|>\varepsilon\right)\leqslant\sum_{b\in I_T}\mathbb{P}\left(\left\|\tilde{B}_{T,2}\right\|>b^{-\delta}\varepsilon\right)\leqslant|I_T|\times\sup_{b\in I_T}\mathbb{P}\left(\left\|\tilde{B}_{T,2}\right\|>b^{-\delta}\varepsilon\right)=O(b).$$

Thus, we have

$$\sup_{b\in I_T}\left\|\tilde{B}_T\right\|\leqslant\sup_{b\in I_T}\left\|\tilde{B}_{T,1}\right\|+\sup_{b\in I_T}\left\|\tilde{B}_{T,2}\right\|=O_p(b^{1-\delta}),$$

which implies (S.4).

S2 Additional simulation results

The parameters a_t and b_t for the DGPs (13) are summarized in Table 1. DGP C1 uses constant parameter values $a_t = 0.9$ and $b_t = 1$ for all t. DGPs C2–C4 allow for a one-time structural break in the parameters a_t and b_t at different time points: $\tau = T/4$, $T/2$, and $3T/4$. DGP C5–C7 have a one-time break at the same time points but with smaller-sized break in the parameters.

Table S1: Specification of DGPs: C1–C7.

DGP	a_t	b_t
C1	0.9	1
C2	$0.9 - T^{-0.2}\,\mathbf{1}(t \geqslant T/4 + 1)$	$1 + T^{-0.2}\,\mathbf{1}(t \geqslant T/4 + 1)$
C3	$0.9 - T^{-0.2}\,\mathbf{1}(t \geqslant T/2 + 1)$	$1 + T^{-0.2}\,\mathbf{1}(t \geqslant T/2 + 1)$
C4	$0.9 - T^{-0.2}\,\mathbf{1}(t \geqslant 3T/4 + 1)$	$1 + T^{-0.2}\,\mathbf{1}(t \geqslant 3T/4 + 1)$
C5	$0.9 - T^{-0.5}\,\mathbf{1}(t \geqslant T/4 + 1)$	$1 + T^{-0.5}\,\mathbf{1}(t \geqslant T/4 + 1)$
C6	$0.9 - T^{-0.5}\,\mathbf{1}(t \geqslant T/2 + 1)$	$1 + T^{-0.5}\,\mathbf{1}(t \geqslant T/2 + 1)$
C7	$0.9 - T^{-0.5}\,\mathbf{1}(t \geqslant 3T/4 + 1)$	$1 + T^{-0.5}\,\mathbf{1}(t \geqslant 3T/4 + 1)$

Forecast comparisons for DGPs C1-C7 are presented in Table S2. In general, the Gaussian kernel delivers the best forecasting results. DGP C1 does not have any parameter instability, and hence the benchmark full-sample estimator is expected to perform well. This is the case when $h = 1$, nonetheless when the forecast horizon is long with $h = 5$, all four local estimators lead to forecasts with smaller MSEs than the benchmark no matter the sample size. When there is a one-time structural break in the parameters, the local estimators tend to produce better forecasts when the size of the break is larger (DGPs C2–C4), and when the break point is closer to the end of the sample. Similar to the results shown in Table 2, the improvement of using local estimators compared to the benchmark forecasts is larger as the forecasting horizon h increases.

S3 Forecast combination methods

Let $\omega_{i,t}$ be the combination weight for model i at time t. For equal-weighted (EW) combinations, we set $\omega_{i,t} = 1/N$, where N is the number of candidate models.

Table S2: Forecasting performance of the local estimators for DGPs C1–C7.

DGP	$h=1$				$h=5$			
	Opt_R	Opt_G	Opt_E	Opt_T	Opt_R	Opt_G	Opt_E	Opt_T
$T=200$								
C1	1.092	1.040	1.105	1.112	0.891	0.878	0.904	0.908
C2	0.883	0.856	0.889	0.893	0.719	0.708	0.721	0.726
C3	0.727	0.707	0.732	0.735	0.533	0.528	0.534	0.535
C4	0.653	0.701	0.656	0.659	0.478	0.501	0.480	0.482
C5	1.063	1.024	1.077	1.087	0.837	0.826	0.848	0.851
C6	1.039	1.001	1.051	1.061	0.798	0.790	0.807	0.810
C7	1.053	1.009	1.063	1.069	0.768	0.771	0.773	0.775
$T=400$								
C1	1.070	1.040	1.075	1.080	0.880	0.874	0.885	0.887
C2	0.827	0.815	0.828	0.831	0.670	0.665	0.667	0.671
C3	0.730	0.718	0.728	0.730	0.498	0.499	0.500	0.500
C4	0.705	0.697	0.708	0.709	0.491	0.497	0.493	0.494
C5	1.050	1.024	1.056	1.059	0.830	0.827	0.830	0.832
C6	1.036	1.014	1.043	1.048	0.799	0.794	0.802	0.804
C7	1.024	1.001	1.028	1.031	0.783	0.782	0.782	0.782
$T=800$								
C1	1.042	1.025	1.047	1.050	0.876	0.874	0.875	0.878
C2	0.834	0.832	0.837	0.839	0.637	0.643	0.633	0.637
C3	0.760	0.753	0.761	0.761	0.508	0.509	0.507	0.508
C4	0.732	0.726	0.731	0.732	0.481	0.483	0.480	0.480
C5	1.035	1.017	1.039	1.041	0.828	0.830	0.826	0.829
C6	1.007	0.997	1.011	1.012	0.805	0.807	0.801	0.805
C7	1.022	1.007	1.027	1.029	0.791	0.794	0.788	0.789

Note: Ratios of MSEs against the benchmark forecasts using full-sample least square estimators. Opt_R: rolling window selection method proposed by Inoue et al. (2017); Opt_G: optimal bandwidth selection with Gaussian kernel; Opt_E: optimal bandwidth selection with Epanechnikov kernel; Opt_T: optimal bandwidth selection with triangular kernel.

For the discounted MSE (DMSE) combining method (Stock and Watson, 2004; Rapach et al., 2010), the weight $\omega_{i,t}$ is computed according to

$$\omega_{i,t} = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^{N} \phi_{j,t}^{-1}}, \qquad \text{with} \quad \phi_{i,t} = \sum_{s=T_0}^{t-1} \rho^{t-1-s} (y_{s+h} - \hat{y}_{i,s+h|s})^2,$$

where ρ is a discounting factor, h is the forecast horizon, y_{s+h} is the true value, and $\hat{y}_{i,s+h|s}$ is the forecast from model i. This method assigns higher weight to an individual model whose forecasts have lower MSEs over the holdout out-of-sample period. When $\rho = 1$, there is no discounting and these weights are exactly the same as Bates and Granger (1969) for the case where the forecasts from one given model are uncorrelated. When $\rho < 1$, higher weights are attached to the more recent forecast accuracy measures for each model. In both applications, we set $\rho = 0.9$.

S4 Forecasting inflation

Real-time price index data are obtained from the Federal Reserve Bank of Philadelphia's Real-Time Dataset for Macroeconomists (RTDSM), described in more detail by Croushore and Stark (2001). We use quarterly data from 1985:Q1 to 2023:Q4. Inflation at time t is measured as $400 \times \ln(P_t/P_{t-1})$, where P_t is the GDP price index.[1] Following Romer and Romer (2000) among many others, we use the second available estimate in the RTDSM to compute the actual inflation and measure the forecast accuracy.[2]

The forecasts are computed using the auto-regressive distributed lag (ARDL) model with time-varying coefficients:

$$y_{t+h} = \theta_{0,t} + \theta_{1,t} y_{t-1} + \theta_{2,t} x_t + \varepsilon_{t+h}, \tag{S.8}$$

where x_t is a scalar predictor and h is the forecast horizon.

We consider the four kernel functions as in (17), with the same optimal bandwidth selection procedure as in the simulation studies in Section 5. Parameter estimation and bandwidth selection are implemented recursively using an expanding window. The bandwidth parame-

[1] For simplicity, "GDP price index" refers to the price index series for GNP/GDP. For some of the sample the measure is based on GNP and a fixed weight deflator.

[2] For example, the first available estimate for 2019:Q4 price index is in the 2020:Q1 vintage, and the second available estimate for 2019:Q4 price index is in the 2020:Q2 vintage. This is what we use to calculate 2019:Q4 inflation.

ter is set to be $b = cT^{-1/3}$, with c varying from 1 to 5 in increments of 0.1. The (left-sided) Epanechnikov kernel with fixed bandwidth parameter $\tilde{b} = 1.06T^{-1/5}$ is used for the local linear estimator $\tilde{\theta}_T$. The parameter estimation and the bandwidth selection are carried out recursively with an expanding window and are updated in each period.

The benchmark forecasts are obtained from a simple AR(1) model by setting $\theta_{2,t} = 0$ in (S.8), estimated using full-sample non-local least square. We also consider forecast combinations from models in which each scalar predictor x_t is used one at a time. In addition, we report forecasts computed with AR(1) model estimated using local estimators for comparison.

We consider a set of predictors inspired by Stock and Watson (2003), which includes interest rates, default spread, stock market variables, commodity prices, exchange rates and monetary variables. Unlike GDP price index, asset prices are not revised, hence we rely on the currently available time series. A detailed description of the list of predictors can be found in Table S3. The initial estimation sample is from 1959:Q3 to 1984:Q4, and the first available individual forecast is computed for 1985:Q1. We use 40 observations as the hold-out out-of-sample to obtain the weights for forecast combination based on the DMSE. Therefore, the forecast evaluation period runs from 1995:Q1 to 2023:Q4. We report results for forecasts at one quarter ($h = 1$) and one year ($h = 4$) ahead.

Table S3: The list of predictors and variable transformation in forecasting the U.S. inflation.

Variable	Description	Source	Transform
FFR	Effective federal funds rate	FRED-QD	level
TmSpd	10-year minus 3-month Treasury bill rates	FRED-QD	level
DfSpd	BAA- minus AAA-rated corporate bond yields	FRED-QD	level
S&P500	S&P500 composite index	FRED-QD	$100\Delta \ln$
PE	Price-earnings ratio for S&P500 composite stocks	FRED-QD	$100\Delta \ln$
CAD	Canada/U.S. exchange rate	FRED-QD	$100\Delta \ln$
GBP	U.K./U.S. exchange rate	FRED-QD	$100\Delta \ln$
COM	Moody's commodity price index	GFD	$100\Delta \ln$
M1REAL	Real M1 money stock, deflated by CPI	FRED-QD	$100\Delta \ln$
M2REAL	Real M2 money stock, deflated by CPI	FRED-QD	$100\Delta \ln$

Note: The FRED-QD data set is developed by McCracken et al. (2021) and maintained by the Federal Reserve Bank of St. Louis. GFD refers to the Global Financial Database.

Table S4 reports the ratio of MSEs of each model to that of the benchmark forecasts.

Apart from the full-sample least square estimator (OLS), we consider the fixed-rolling window estimator with window size 40 ($R = 40$), optimal rolling window selection method proposed by Inoue et al. (2017) (Opt_R), and the local estimator with optimally selected bandwidth using the the Gaussian kernel (Opt_G), the Epanechnikov kernel (Opt_E), and the Triangular kernel (Opt_T). The first row represents the AR(1) model, rows two through eleven correspond to the model in Equation (S.8) with different predictors, and the final two rows represent the forecast combinations of the forecasts from different predictors given above.

Table S4: Forecasting performance for U.S. inflation: 1985:Q1–2023:Q4.

	$h = 1$				$h = 4$			
	Opt_R	Opt_G	Opt_E	Opt_T	Opt_R	Opt_G	Opt_E	Opt_T
AR	0.954	0.935	0.948	0.959	0.748*	0.738*	0.738*	
FFR	0.870*	0.876*	0.898	0.897	0.780*	0.811*	0.767*	0.751*
TmSpd	0.998	0.961	0.997	0.996	0.786*	0.808*	0.754*	0.746*
DfSpd	1.006	0.924	1.108	1.117	0.769*	0.784*	0.790*	0.791*
S&P500	1.002	0.924	1.040	1.061	0.735*	0.759*	0.726*	0.713*
PE	1.009	0.947	1.008	1.007	0.721*	0.774*	0.700*	0.691*
CAD	0.983	0.948	1.000	1.014	0.706*	0.745*	0.687*	0.681*
GBP	0.985	0.919	1.007	1.028	0.751*	0.766*	0.739*	0.729*
COM	0.822	0.800*	0.836	0.848	0.680*	0.710*	0.657*	0.651*
M1REAL	3.081	2.509	2.853	3.583	0.759*	0.795*	0.780*	0.756*
M2REAL	1.116	0.968	1.216	1.475	0.788*	0.806*	0.794*	0.797*
Comb-EW	0.968	0.936	0.981	1.017	0.719*	0.761*	0.704*	0.693*
Comb-DMSE	0.974	0.939	0.986	1.025	0.717*	0.761*	0.702*	0.691*

Note: Ratio of MSEs against the benchmark forecasts of AR(1) model estimated using non-local least square. Opt_R: rolling window selection method proposed by Inoue et al. (2017); Opt_G: optimal bandwidth selection with Gaussian kernel; Opt_E: optimal bandwidth selection with Epanechnikov kernel; Opt_T: optimal bandwidth selection with triangular kernel. Differences in forecasting accuracy that are significant at the 5% level using the DM test are marked by an asterisk. The grey-shaded cells denote the best forecasting performance for each group.

There are several issues worth mentioning. First, using local estimators improves forecast accuracy for the AR(1) model. Gains are always significant, and are larger for one year ahead forecast ($h = 4$). Second, adding additional predictor is not always useful. Choice of predictor really matters. The commodity price index is the most reliable predictor, which delivers the best forecasting performance. The gains also become more evident for $h = 4$. Using Gaussian kernel is the best for $h = 1$, while triangular kernel is preferred for $h = 4$.

Finally, forecast combinations improve the forecast accuracy in nearly all cases, except Opt_T for $h = 1$.

Table S5 presents forecasting evaluation results for the period up to 2019:Q4, excluding COVID-19 observations to avoid pandemic-related distortions. The overall conclusions are similar, with a few noticeable differences. First, using local estimators improves forecast accuracy in all cases. Second, DMSE combining method delivers the best results. Inoue et al. (2017)'s method is overall the best for $h = 1$, while using triangular kernel is the best for $h = 4$. However, when we test the equal forecast accuracy between the best performing case and the second best case (AR Opt_R for $h = 1$ and AR Opt_T for $h = 4$), the results are only significant for $h = 1$. This implies that exogenous predictors are not so useful once we control for parameter instability, especially for longer-horizon forecasts.

Table S5: Forecasting performance for U.S. inflation: 1985:Q1–2019:Q4.

	$h=1$				$h=4$			
	Opt_R	Opt_G	Opt_E	Opt_T	Opt_R	Opt_G	Opt_E	Opt_T
AR	0.768*	0.789*	0.777	0.784	0.578*	0.628*	0.548*	0.537*
FFR	0.775*	0.779*	0.811	0.826	0.624*	0.689*	0.610*	0.598*
TmSpd	0.808*	0.809*	0.808	0.811	0.645*	0.688*	0.594*	0.578*
DfSpd	0.876	0.799*	1.029	1.031	0.554*	0.594*	0.624*	0.637*
S&P500	0.840	0.805*	0.903	0.937	0.541*	0.598*	0.522*	0.516*
PE	0.744*	0.774*	0.766*	0.777	0.625*	0.668*	0.621*	0.623*
CAD	0.814*	0.809*	0.838	0.863	0.571*	0.608*	0.528*	0.518*
GBP	0.818	0.793*	0.857	0.886	0.606*	0.625*	0.588*	0.583*
COM	0.816	0.804*	0.841	0.868	0.535*	0.556*	0.520*	0.520*
M1REAL	0.759*	0.785*	0.792	0.822	0.559*	0.641*	0.536*	0.519*
M2REAL	0.776	0.779*	0.800	0.813	0.633*	0.655*	0.605*	0.598*
Comb-EW	0.739*	0.764*	0.763	0.780	0.533*	0.604*	0.505*	0.495*
Comb-DMSE	0.737*	0.764*	0.763*	0.781	.529*	0.601*	0.501*	0.492*

Note: Ratio of MSEs against the benchmark forecasts of AR(1) model estimated using non-local least square. Opt_R: rolling window selection method proposed by Inoue et al. (2017); Opt_G: optimal bandwidth selection with Gaussian kernel; Opt_E: optimal bandwidth selection with Epanechnikov kernel; Opt_T: optimal bandwidth selection with triangular kernel. Differences in forecasting accuracy that are significant at the 5% level using the DM test are marked by an asterisk. The grey-shaded cells denote the best forecasting performance for each group.

References

Bates, J. M. and Granger, C. W. (1969), 'The combination of forecasts', *Journal of the operational research society* **20**(4), 451–468.

Croushore, D. and Stark, T. (2001), 'A real-time data set for macroeconomists', *Journal of econometrics* **105**(1), 111–130.

Inoue, A., Jin, L. and Rossi, B. (2017), 'Rolling window selection for out-of-sample forecasting with time-varying parameters', *Journal of econometrics* **196**(1), 55–67.

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), 1–44.

Rapach, D. E., Strauss, J. K. and Zhou, G. (2010), 'Out-of-sample equity premium prediction: combination forecasts and links to the real economy', *The Review of Financial Studies* **23**(2), 821–862.

Romer, C. D. and Romer, D. H. (2000), 'Federal reserve information and the behavior of interest rates', *American economic review* **90**(3), 429–457.

Stock, J. H. and Watson, M. W. (2003), 'Forecasting output and inflation: the role of asset prices', *Journal of economic literature* **41**(3), 788–829.

Stock, J. H. and Watson, M. W. (2004), 'Combination forecasts of output growth in a seven-country data set', *Journal of forecasting* **23**(6), 405–430.