Model Averaging for Time–Varying Vector Autoregressions*

Yuying Sun[1], Feng Chen[1] and Jiti Gao[2†]

[1]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

[2]*Department of Econometrics and Business Statistics, Monash University*

SUMMARY

This paper proposes a novel time-varying model averaging (TVMA) approach to enhancing forecast accuracy for multivariate time series subject to structural changes. The TVMA method averages predictions from a set of time-varying vector autoregressive models using optimal time-varying combination weights selected by minimizing a penalized local criterion. This allows the relative importance of different models to adaptively evolve over time in response to structural shifts. We establish an asymptotic optimality for the proposed TVMA approach in achieving the lowest possible quadratic forecast errors. The convergence rate of the selected time-varying weights to the optimal weights minimizing expected quadratic errors is derived. Moreover, we show that when one or more correctly specified models exist, our method consistently assigns full weight to them, and an asymptotic normality for the TVMA estimators under some regularity conditions can be established. Furthermore, the proposed approach encompasses special cases including time-varying VAR models with exogenous predictors, as well as time-varying factor augmented VAR (FAVAR) models. Simulations and empirical applications illustrate the proposed TVMA method outperforms some commonly used model averaging and selection methods in the presence of structural changes.

KEY WORDS: Asymptotic optimality; Consistency; Structural change; Time-varying weight

JEL Classification codes: C52, C53.

*All remaining errors are solely ours. Sun thanks support from China NNSF Grant (Nos. 72322016, 72073126, 72091212) and Young Elite Scientists Sponsorship Program by CAST (YESS) [2020QNRC001]. Gao acknowledges the financial support from the Australian Research Council Discovery Programs Scheme under Grant Number: DP250100063.

†Corresponding author: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email: jiti.gao@gmail.com.

# 1 Introduction

There is mounting evidence of structural instabilities in multivariate time series relationships due to factors such as macroeconomic shocks, policy changes, preference shifts, and technological progress (Gao et al., 2024a,b; Paye & Timmermann, 2006; Stock & Watson, 2007, 2016; Su & Wang, 2017). Studies have observed substantial instability in autoregressive models for macroeconomic series (Primiceri, 2005; Stock & Watson, 1996), stock returns and cash flows dynamics (Yu & Yan, 2023), exchange rates (Patton, 2006), and climate data exhibiting time trends (Chen et al., 2022). Popular approaches have emerged to capture the evolutionary behavior of economic and financial time series, including time-varying vector autoregressive (VAR) models with stochastic volatility (Gao et al., 2024a; Primiceri, 2005) and nonparametric time-varying VAR-type models allowing coefficients to change smoothly over time (Fu et al., 2024; Yan et al., 2024). Motivated by the flexibility of these nonparametric-based models, we propose a new model averaging approach for a class of time-varying VAR models.

Model uncertainty poses another significant challenge in multivariate time series forecasting, requiring determination of optimal lagged orders and relevant exogenous predictors amidst numerous candidate models (Liao et al., 2019; Liao & Tsay, 2020). Model selection using information criteria like Akaike information criterion (AIC) and Bayesian information criterion (BIC) overlooks useful alternative models and is sensitive to data perturbations, causing biased and unstable predictions (Yang, 2003; Yuan & Yang, 2005). Instead, model averaging offers a sensible approach to mitigating structural instability and model uncertainty (Ando & Li, 2014; Wan et al., 2010; Zhu et al., 2019). However, most of existing frequentist strategies, like Mallows criteria (Hansen, 2007; Liao et al., 2021; Liao & Tsay, 2020), cross-validation (Cheng & Hansen, 2015; Gao et al., 2016; Liao et al., 2019), and forward-validation (Zhang & Zhang, 2023), are designed to select optimal time-invariant combination weights, failing to address structural changes and model uncertainty in multivariate time series. Intuitively, as the forecasting ability of candidate models fluctuates, it is reasonable to assign time-varying weights, allocating higher weights to well-performing models and lower or zero weights to poorly performing ones at different time periods (Chen & Maung, 2023; Chen et al., 2024).

In this paper, our attempt is to develop an optimal model averaging method for a class of time–varying VAR models, which immediately faces some challenges distinct from the univariate time-varying setting. First, we need to introduce a proper time-varying weight choice criterion that can effectively capture the cross-sectional dependence dynamics among the multivariate time series under study, for which existing local weight choice criteria, such as those discussed in Sun et al. (2021, 2023), and Chen et al. (2024), fail to adequately address such issues. Second, there are no distributional results available for model averaging estimators in multivariate time series settings even with time-invariant weights, not to mention time–varying weights. Third, existing proof techniques for asymptotic optimality and consistency are no longer applicable for time-varying VAR cases, and new developments are required for the establishment of such techniques and tools.

To address these challenges, we propose a penalized time-varying model averaging (TVMA) method within a flexible nonparametric time-varying VAR-based framework. The proposed time-varying criterion is a weighted local quadratic loss function that (i) utilizes the inverse of the time-varying covariance matrix to capture cross-dependence; and (ii) employs penalties to reduce model complexity and select important predictors in high-dimensional settings. This approach allows for the combination of the weights and parameter estimates within the candidate models to adapt smoothly over time, aligning with the dynamic nature of economic structures and the evolving predictive capacities of the models. Furthermore, the proposed approach encompasses scenarios, such as time-varying VAR models with exogenous predictors, as well as time-varying factor-augmented VAR (FAVAR) models. Extensive simulation studies and real data analyses provide strong empirical support for our method.

It is worth discussing some key references and outlining our contributions in relation to the most relevant literature.

(i). The proposed weight choice criterion includes the Mallows-type criterion proposed by Liao et al. (2019) and Liao & Tsay (2020) for time-invariant VAR models. Meanwhile, compared with the cross-validation model averaging in Liao et al. (2019), Sun et al. (2021) and Chen et al. (2024) with time-invariant and time-variant combination weights, respectively, our method is unified, cost-effective and easy to implement. Our approach is also applicable to heteroscedasticity cases.

(ii). We demonstrate that the selected TVMA weights are asymptotically optimal in achieving the lowest possible quadratic loss and consistent even when all candidate models are misspecified. Instead of imposing high–level stochastic mixing conditions, such as those discussed in Appendix A of Gao (2007) for time–invariant models, which are not generally verifiable in such time–varying settings considered in Sun et al. (2021, 2023), and Chen et al. (2024), we will develop a suite of technologies based on a class of time–varying vector moving average (VMA($\infty$)) processes proposed in Gao et al. (2024b) and Yan et al. (2024), which are of independent interest, and easily verifiable and applicable to many scenarios.

(iii). We then establish the consistency and asymptotic normality of the proposed TVMA estimator when correctly specified models are included in the candidates. Specifically, the proposed approach assigns full weight to the correctly specified models. To our knowledge, these findings on asymptotic consistency and normality are the first to be available in the relevant multivariate time series model averaging literature, regardless of whether the selected weights are time-invariant or time-variant.

The rest of this paper is organized as follows. Section 2 introduces the penalized time-varying model avaraging method in a time-varying vector moving averaging (VMA) framework. Section 3 establishes the corresponding asymptotic properties of the proposed TVMA estimator, including asymptotic optimality, consistency, convergence rate and asymptotic normality. Section 4 discusses possible extensions to time-varying VARX models and time-varying FAVAR models. Section 5 presents simulation studies under various structural changes, and Section 6 provides the empirical studies. Section 7 concludes. Throughout the rest of this paper, all convergences occur as the sample size $T \to \infty$, and all mathematical proofs are collected in Appendices A, B and C.

# 2 Time-varying model averaging

## 2.1 Model framework

Suppose the data generating process (DGP) is governed by the following time-varying VAR($\infty$) model

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t \equiv \mathbf{a}(\tau_t) + \sum_{j=1}^{\infty} \mathbf{A}_j(\tau_t)\mathbf{y}_{t-j} + \boldsymbol{\epsilon}_t,\ t = 1, \cdots, T, \tag{2.1}$$

where $\mathbf{y}_t = (y_{t,1}, \cdots, y_{t,K})'$ is a vector of $K$-dimensional observable variables, $\mathbf{a}(\cdot)$, $\mathbf{A}_j(\cdot)$ and $\boldsymbol{\omega}(\cdot)$ are respectively vector and matrices of $K \times 1$, $K \times K$ and $K \times K$ elements, and $\boldsymbol{\epsilon}_t = (\epsilon_{t,1}, \cdots, \epsilon_{t,K})' = \boldsymbol{\omega}(\tau_t)\mathbf{u}_t$. In the relevant literature, such as Gao et al. (2024b), it is often assumed that $\mathbf{a}(\tau) = \mathbf{a}(0)$ and $\mathbf{A}_j(\tau) = \mathbf{A}_j(0)$ for $\tau < 0$, $\boldsymbol{\omega}(\tau)$ is an unknown function that has full row rank uniformly in $\tau \in [0, 1]$, and $\boldsymbol{\Sigma}(\tau) = \boldsymbol{\omega}(\tau)\boldsymbol{\omega}(\tau)'$ is positive definite for all $\tau \in [0, 1]$. We also assume that each component of $\boldsymbol{\omega}(\tau)$ is second–order continuously differentiable on $[0, 1]$, and $\boldsymbol{\omega}(\tau) = \boldsymbol{\omega}(0)$ for $\tau < 0$. The relevant literature also regularly assumes that $\{\mathbf{u}_t\}_{t=-\infty}^{\infty}$ is a martingale differential sequence such that $\mathbb{E}(\mathbf{u}_t|I_{t-1}) = \mathbf{0}$, $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t'|I_{t-1}) = \mathbf{I}_K$ almost surely, and $\sup_{t\geq 1} \mathbb{E}\|\mathbf{u}_t\|^s < \infty$ for some $s > 4$, where $I_{t-1} = \{\mathbf{u}_s\}_{s=-\infty}^{t-1}$.

Consider $S$ candidate models with time-varying parameters to approximate the DGP in (2.1), where $S$ is allowed to diverge with the sample size $T$. The $s$–th ($1 \leq s \leq S$) candidate model is

$$\mathbf{y}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^{s} \mathbf{A}_j(\tau_t)\mathbf{y}_{t-j} + \boldsymbol{\epsilon}_t^{(s)} = \mathbf{A}^{(s)\prime}(\tau_t)\mathbf{z}_{t-1}^{(s)} + \boldsymbol{\epsilon}_t^{(s)}, \tag{2.2}$$

where $\mathbf{z}_{t-1}^{(s)} = \left[1, \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-s}'\right]'$, $\mathbf{A}^{(s)}(\tau_t) = [\mathbf{a}(\tau_t), \mathbf{A}_1(\tau_t), \ldots, \mathbf{A}_s(\tau_t)]'$, and $S + 1 \leq t \leq T$. Let $\mathbf{Y} = (\mathbf{y}_{S+1}, \mathbf{y}_{S+2}, \cdots, \mathbf{y}_T)'$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_{S+1}, \boldsymbol{\mu}_{S+2}, \cdots, \boldsymbol{\mu}_T)'$, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_{S+1}, \boldsymbol{\epsilon}_{S+2}, \cdots, \boldsymbol{\epsilon}_T)'$. We note that our framework allows for both nested and non-nested candidate models. Without loss of generality, we present the case of nested candidate models as a simple case.

Back to model (2.2), for the $s$–th candidate model, we make the best use of the local linear kernel method to estimate $\mathbf{A}^{(s)}(\tau_t)$ in the same way as in the relevant literature (see, for example, Yan et al. (2024) and Gao et al. (2024b)). Intuitively, when $\tau_t$ is in a small neighbourhood of $\tau$, we can write (2.1) as follows:

$$\mathbf{y}_t \approx \left[\mathbf{A}^{(s)'}(\tau), l\dot{\mathbf{A}}^{(s)'}(\tau)\right] \widetilde{\mathbf{z}}_{t-1}^{(s)} + \boldsymbol{\epsilon}_t^{(s)}, \tag{2.3}$$

where $\dot{\mathbf{A}}^{(s)}(\cdot)$ is the first derivative of $\mathbf{A}^{(s)}(\cdot)$ and $\widetilde{\mathbf{z}}_{t-1}^{(s)} = \left[\mathbf{z}_{t-1}^{(s)'}, \frac{\tau_t-\tau}{l}\mathbf{z}_{t-1}^{(s)'}\right]'$. The local linear estimator of $\mathbf{A}^{(s)}(\tau)$ is then given by

$$\widehat{\mathbf{A}}^{(s)}(\tau) = \left( \ \mathbf{I}_{(Ks+1)}, \mathbf{0}_{(Ks+1)} \ \right) \left(\widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_t\widetilde{\mathbf{Z}}^{(s)}\right)^{-1} \widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_t\mathbf{Y}, \tag{2.4}$$

in which $\widetilde{\mathbf{Z}}^{(s)} = \left(\widetilde{\mathbf{z}}_S^{(s)}, \widetilde{\mathbf{z}}_{S+1}^{(s)}, \cdots, \widetilde{\mathbf{z}}_{T-1}^{(s)}\right)'$ with its dimensionality being $(T-S) \times 2(Ks+1)$, $\mathbf{K}_t = \mathrm{diag}\left\{k_{(S+1)t}, k_{(S+2)t}, \ldots, k_{Tt}\right\}$ is the weighting matrix with its dimensionality being $(T-S) \times (T-S)$ and $(r,t)$–element being $k_{rt} = k((r-t)/(Tl))$, $k(\cdot)$ is a symmetric and positive kernel function, and $l$ is a bandwidth.

The estimator of $\boldsymbol{\mu}_t$ in the $s$–th candidate model is

$$\widehat{\boldsymbol{\mu}}_t^{(s)} = \widehat{\mathbf{A}}^{(s)'}(\tau_t)\mathbf{z}_{t-1}^{(s)}.$$

We then have $\widehat{\boldsymbol{\mu}}^{(s)} \equiv \left(\widehat{\boldsymbol{\mu}}_{S+1}^{(s)}, \widehat{\boldsymbol{\mu}}_{S+2}^{(s)}, \cdots, \widehat{\boldsymbol{\mu}}_T^{(s)}\right)' = \mathbf{P}^{(s)}\mathbf{Y}$, where

$$\mathbf{P}^{(s)} = \begin{pmatrix} \mathbf{z}_S^{(s)'} \left( \ \mathbf{I}_{(Ks+1)}, \mathbf{0}_{(Ks+1)} \ \right) \left(\widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_{S+1}\widetilde{\mathbf{Z}}^{(s)}\right)^{-1} \widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_{S+1} \\ \mathbf{z}_{S+1}^{(s)'} \left( \ \mathbf{I}_{(Ks+1)}, \mathbf{0}_{(Ks+1)} \ \right) \left(\widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_{S+2}\widetilde{\mathbf{Z}}^{(s)}\right)^{-1} \widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_{S+2} \\ \vdots \\ \mathbf{z}_{T-1}^{(s)} \left( \ \mathbf{I}_{(Ks+1)}, \mathbf{0}_{(Ks+1)} \ \right) \left(\widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_T\widetilde{\mathbf{Z}}^{(s)}\right)^{-1} \widetilde{\mathbf{Z}}^{(s)'}\mathbf{K}_T \end{pmatrix}.$$

For each $t$, let the weight vector $\mathbf{w} = \left(w^{(1)}, ..., w^{(S)}\right)'$, belonging to the set $\mathcal{W} = \{\mathbf{w} \in [0,1]^S : \sum_{s=1}^S w^{(s)} = 1\}$. Then, the model average estimators of $\boldsymbol{\mu}_t$ and $\mathbf{A}(\tau_t)$ at any given time $t$ can be respectively expressed by

$$\widehat{\boldsymbol{\mu}}_t(\mathbf{w}) = \sum\nolimits_{s=1}^S w^{(s)}\widehat{\boldsymbol{\mu}}_t^{(s)}, \tag{2.5}$$

and

$$\widehat{\mathbf{A}}\left(\tau_t, \mathbf{w}\right) = \sum_{s=1}^S w^{(s)}\widehat{\mathbf{A}}^{(s)}(\tau_t). \tag{2.6}$$

## 2.2 Penalized time-varying weight choice criterion

To account for the interrelationships among the univariate time series components, we propose the following penalized time-varying weight choice criterion:

$$\begin{aligned} \mathrm{TVMA}_{t,T}(\mathbf{w}) &= \mathrm{trace}\left\{\sqrt{\mathbf{K}_t}\left[\mathbf{Y}-\widehat{\boldsymbol{\mu}}(\mathbf{w})\right]\boldsymbol{\Sigma}_t^{-1}\left[\mathbf{Y}-\widehat{\boldsymbol{\mu}}(\mathbf{w})\right]'\sqrt{\mathbf{K}_t}\right\}+\lambda_T K^2\sum_{s=1}^{S} w^{(s)}s \\ &= \sum_{r=S+1}^{T} k_{rt}\left(\mathbf{y}_r-\widehat{\boldsymbol{\mu}}_r(\mathbf{w})\right)'\boldsymbol{\Sigma}_t^{-1}\left(\mathbf{y}_r-\widehat{\boldsymbol{\mu}}_r(\mathbf{w})\right)+\lambda_T K^2\sum_{s=1}^{S} w^{(s)}s, \end{aligned} \tag{2.7}$$

where $\widehat{\boldsymbol{\mu}}(\mathbf{w})=\sum_{s=1}^{S} w^{(s)}\widehat{\boldsymbol{\mu}}^{(s)}$, $\boldsymbol{\Sigma}_t=\boldsymbol{\Sigma}(\tau_t)$, $s$ is the lag order and $\lambda_T$ is the tuning parameter. The penalty term serves as a penalization on the complexity of the candidate models.

It is worth noting that the proposed TVMA criterion nests several existing information criteria as special cases in time-invariant settings, i.e., time-invariant parameters with multivariate least squares estimation, time-invariant $\mathbf{w}$ and $\boldsymbol{\Sigma}_t$ with an identity matrix $\mathbf{K}_t$. Specifically, when a single element of the weight vector $\mathbf{w}$ equals one with all others being zero, $\mathrm{TVMA}_{t,T}(\mathbf{w})$ reduces to the AIC-type criterion if $\lambda_T=2$ and BIC-type if $\lambda_T=\ln(T)$. More generally, for $\mathbf{w}\in\mathcal{W}$, $\mathrm{TVMA}_{t,T}(\mathbf{w})$ with $\lambda_T=2$ becomes the MMA criterion in Hansen (2007) for the random sampling case, and Liao & Tsay (2020) for the time–invariant multivariate time series setting. Furthermore, if $K=1$, $\mathrm{TVMA}_{t,T}(\mathbf{w})$ reduces to the weight choice criterion in Zhang et al. (2020) for univariate analysis. Minimizing $\mathrm{TVMA}_{t,T}(\mathbf{w})$, we obtain

$$\widehat{\mathbf{w}}_t=\mathrm{argmin}_{w\in\mathcal{W}}\mathrm{TVMA}_{t,T}(\mathbf{w}). \tag{2.8}$$

Then, the time-varying model averaging estimators of $\boldsymbol{\mu}_t$ and $\mathbf{A}(\tau_t)$ for any given time point $t$ are respectively $\widehat{\boldsymbol{\mu}}_t(\widehat{\mathbf{w}}_t)$ and $\widehat{\mathbf{A}}(\tau_t,\widehat{\mathbf{w}}_t)$. The time-varying $\boldsymbol{\Sigma}_t$ allows for heteroscedasticity, and our approach offers a cost-effective and easily implementable alternative to the cross-validation method (Chen et al., 2024; Liao et al., 2019) in such cases.

However, in practice, $\boldsymbol{\Sigma}_t$ is unknown. We explore the proposal of Liao et al. (2019) for the time–invariant setting to estimate the unknown $\boldsymbol{\Sigma}_t$ in the largest model (i.e., the $S$–th candidate model if all candidate models are nested) by using the local information. Then, for any given $t$, the estimator of $\boldsymbol{\Sigma}_t$ is

$$\widehat{\boldsymbol{\Sigma}}_t=\left(\mathrm{trace}(\mathbf{K}_t)\right)^{-1}\left(\mathbf{Y}-\widehat{\boldsymbol{\mu}}^{(S)}\right)'\mathbf{K}_t\left(\mathbf{Y}-\widehat{\boldsymbol{\mu}}^{(S)}\right). \tag{2.9}$$

The feasible penalized time-varying weight choice criterion is given by

$$\begin{aligned}\mathrm{TVMA}_{t,T}^{F}(\mathbf{w}) &= \mathrm{trace}\left\{\sqrt{\mathbf{K}_t}\left[\mathbf{Y}-\widehat{\boldsymbol{\mu}}(\mathbf{w})\right]\widehat{\boldsymbol{\Sigma}}_t^{-1}\left[\mathbf{Y}-\widehat{\boldsymbol{\mu}}(\mathbf{w})\right]'\sqrt{\mathbf{K}_t}\right\}+\lambda_T K^2\sum_{s=1}^{S}w^{(s)}s \\ &= \sum_{r=S+1}^{T}k_{rt}\left(\mathbf{y}_r-\widehat{\boldsymbol{\mu}}_r(\mathbf{w})\right)'\widehat{\boldsymbol{\Sigma}}_t^{-1}\left(\mathbf{y}_r-\widehat{\boldsymbol{\mu}}_r(\mathbf{w})\right)+\lambda_T K^2\sum_{s=1}^{S}w^{(s)}s. \qquad (2.10)\end{aligned}$$

Accordingly, the corresponding feasible time-varying weight vector is

$$\widehat{\mathbf{w}}_t^F=\mathrm{argmin}_{w\in\mathcal{W}}\mathrm{TVMA}_{t,T}^{F}(\mathbf{w}). \qquad (2.11)$$

Then, the feasible time-varying model averaging estimators of $\boldsymbol{\mu}_t$ and $\mathbf{A}(\tau_t)$ for any given time point $t$ are respectively $\widehat{\boldsymbol{\mu}}_t(\widehat{\mathbf{w}}_t^F)$ and $\widehat{\mathbf{A}}(\tau_t,\widehat{\mathbf{w}}_t^F)$.

# 3 Asymptotic properties

## 3.1 Asymptotic optimality

To establish an asymptotic optimality for the proposed TVMA estimators, we consider the following infeasible locally weighted quadratic error loss:

$$L_{t,T}(\mathbf{w})=\mathrm{trace}\left\{\sqrt{\mathbf{K}_t}\left[\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}(\mathbf{w})\right]\boldsymbol{\Sigma}_t^{-1}\left[\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}(\mathbf{w})\right]'\sqrt{\mathbf{K}_t}\right\}. \qquad (3.1)$$

Denote $\boldsymbol{\mu}^*(\mathbf{w})=\left(\boldsymbol{\mu}_{S+1}^*(\mathbf{w}),\boldsymbol{\mu}_{S+2}^*(\mathbf{w}),\ldots,\boldsymbol{\mu}_T^*(\mathbf{w})\right)'$, where $\boldsymbol{\mu}_t^{(s)*}=\widehat{\boldsymbol{\mu}}_t^{(s)}\big|_{\widehat{\mathbf{A}}^{(s)}(\tau_t)=\mathbf{A}^{(s)*}(\tau_t)}$, $\boldsymbol{\mu}_t^*(\mathbf{w})=\sum_{s=1}^{S}w^{(s)}\boldsymbol{\mu}_t^{(s)*}$, and $\mathbf{A}^{(s)*}(\tau_t)$ is defined in Condition (C.1) below. We then define $L_{t,T}^*(\mathbf{w})=\mathrm{trace}\left\{\sqrt{\mathbf{K}_t}\left[\boldsymbol{\mu}-\boldsymbol{\mu}^*(\mathbf{w})\right]\boldsymbol{\Sigma}_t^{-1}\left[\boldsymbol{\mu}-\boldsymbol{\mu}^*(\mathbf{w})\right]'\sqrt{\mathbf{K}_t}\right\}$, and $\xi_{t,T}=\inf_{\mathbf{w}\in\mathcal{W}}\mathbb{E}L_{t,T}^*(\mathbf{w})$. Let $\zeta_{\min}(\mathbf{A})$ and $\zeta_{\max}(\mathbf{A})$ represent the minimum and maximum singular values of matrix $\mathbf{A}$, respectively. The symbols $\otimes$ and vec signify the Kronecker product and the operation of stacking all the columns of a given matrix into a vector, respectively. All asymptotic behaviours below are considered in the context of $T\to\infty$.

**Condition (C.1).** *For the $s$–th candidate model, given any time $t$, there exists a limit $\mathbf{A}^{(s)*}(\tau_t)$ such that $\left\|\widehat{\mathbf{A}}^{(s)}(\tau_t)-\mathbf{A}^{(s)*}(\tau_t)\right\|=O_p\left(S^{1/2}T^{-1/2}l^{-1/2}\right)$.*

**Condition (C.2).** *Let $k(\cdot)$ be a symmetric and positive kernel function with bounded support on $[-1,1]$ and $\int_{-1}^{1}k(u)\,du=1$. Moreover, $k(\cdot)$ is Lipschitz continuous on $[-1,1]$.*

**Condition (C.3).** *The bandwidth $l = cT^{-\lambda}$ for some positive $\lambda$ and $0 < c < \infty$.*

**Condition (C.4).** *$S \to \infty$, $S(Tl)^{-1/2} \to 0$, $\lambda_T \to \infty$, $S\lambda_T \xi_{t,T}^{-1} \to 0$ and $S^4 Tl \xi_{t,T}^{-2} \to 0$.*

**Remark 1.** *Condition (C.1) is concerned with the convergence (including the rate) of local linear kernel estimation method proposed in Gao et al. (2024b). Following the proofs of Proposition 1 and Theorem 1.2 of Gao et al. (2024b), it can be shown that Condition (C.1) is satisfied for the case where $S \to \infty$ and $ST^{-1}l^{-1} \to 0$. More discussions and verifications of Condition (C.1) are available in Appendix A. In addition, with the under-smoothing bandwidth, the squared bias term $\widehat{\mathbf{A}}^{(s)}(\tau_t) - \mathbf{A}^{(s)*}(\tau_t)$ can be dominated by the variance term $O_p\left(S^{1/2}T^{-1/2}l^{-1/2}\right)$. Thus, the bias term is ignored in Condition (C.1).*

**Remark 2.** *Condition (C.2) includes commonly used kernels with compact support $[-1, 1]$, such as the Epanechnikov, Uniform and Triangular kernels. Furthermore, Condition (C.2) implies that $\int_{-1}^{1} uk(u)\,du = 0$, $\int_{-1}^{1} u^2 k(u)\,du < \infty$, and $\int_{-1}^{1} uk^2(u)\,du < \infty$. Condition (C.3) includes the commonly used bandwidth $l \propto T^{-1/5}$.*

**Remark 3.** *Condition (C.4) imposes some restrictions on the speed of growth of the number of candidate models $S$, the tuning parameter $\lambda_T$ and the limit risk $\xi_{t,T}$. Like Ando & Li (2014), we consider a case with $\xi_{t,T} = T^{1-\widetilde{\delta}}$, $0 < \widetilde{\delta} < 1/2$, and then we obtain $S = O(T^{1/20})$. Note that Condition (C.4) implies that $\xi_{t,T} \to \infty$, which indicates that all candidate models are misspecified. If there were any correctly specified candidate model, the limiting parameter value $\mathbf{A}^{(s)*}(\tau_t)$ would coincide with the true parameter value at any time $t$, and then $\xi_{t,T}$ would not increase with $T$. Specifically, suppose that the $(S_0 + 1)$–th model is correctly specified, then it follows that*

$$\begin{aligned}
\xi_{t,T} &= \inf_{\mathbf{w} \in \mathcal{W}} \mathbb{E} trace \left\{ \sqrt{\mathbf{K}_t} \left[\boldsymbol{\mu} - \boldsymbol{\mu}^*(\mathbf{w})\right] \boldsymbol{\Sigma}_t^{-1} \left[\boldsymbol{\mu} - \boldsymbol{\mu}^*(\mathbf{w})\right]' \sqrt{\mathbf{K}_t} \right\} \\
&\leq \mathbb{E} trace \left\{ \sqrt{\mathbf{K}_t} \left[\boldsymbol{\mu} - \boldsymbol{\mu}^{(S_0+1)*}\right] \boldsymbol{\Sigma}_t^{-1} \left[\boldsymbol{\mu} - \boldsymbol{\mu}^{(S_0+1)*}(\mathbf{w})\right]' \sqrt{\mathbf{K}_t} \right\} = 0. \qquad (3.2)
\end{aligned}$$

*As a result, Condition (C.4) is violated when one of the candidate models is correctly specified. Note that Condition (C.4) is concerned with the degree of misspecification which controls the local distance between the quasi-true value $\boldsymbol{\mu}_t^{(s)*}$ and the true value $\boldsymbol{\mu}_t$. This does not* *contradict Condition (C.1). Furthermore, Condition (C.4) is consisent with Condition 3 of Gao et al. (2023) if $S$ is fixed, with both being weaker than Condition (7) of Ando & Li (2014) and Condition (A3) of Ando & Li (2017).*

**Condition (C.5).** $\zeta_{\max}\left(\widehat{\boldsymbol{\Sigma}}_t^{-1} - \boldsymbol{\Sigma}_t^{-1}\right) STl\xi_{t,T}^{-1} = o_p(1)$.

**Condition (C.5)′.** $\zeta_{\max}\left(\widehat{\boldsymbol{\Sigma}}_t^{-1} - \boldsymbol{\Sigma}_t^{-1}\right) = o_p(1)$.

**Remark 4.** *Condition (C.5) is consistent with Condition 5 of Gao et al. (2023) in time-invariant settings, which restricts the relationship between the largest singular value of $\widehat{\boldsymbol{\Sigma}}_t^{-1} - \boldsymbol{\Sigma}_t^{-1}$, $STl$ and $\xi_{t,T}^{-1}$. Condition (C.5) is weaker than Condition (C.8) of Liao et al. (2019) and the assumption in the Appendix of Liao & Tsay (2020), which assume the covariance matrix is consistently estimated. When $\widehat{\boldsymbol{\Sigma}}_t \xrightarrow{p} \boldsymbol{\Sigma}_t$, Conditions (C.5) and (C.5)′ hold since the map $\mathbf{A} \mapsto \mathbf{A}^{-1}$ is continuous on the set of invertible matrices. When $\widehat{\boldsymbol{\Sigma}}_t$ is not a consistent estimator for $\boldsymbol{\Sigma}_t$, we could restrict the degree of misspecification $\xi_{t,T}$ to be large to satisfy Condition (C.5). More discussions are given in Appendix A.*

**Theorem 1.** *Let Conditions (C.1)-(C.5) hold. Then for any given time $t$, the proposed feasible TVMA estimator satisfies the asymptotic optimality (OPT) property, i.e.,*

$$\frac{L_{t,T}\left(\widehat{\mathbf{w}}_t^F\right)}{\inf_{\mathbf{w}\in\mathcal{W}} L_{t,T}(\mathbf{w})} \xrightarrow{p} 1, \text{ as } T \to \infty. \tag{3.3}$$

Theorem 1 demonstrates that the feasible TVMA estimator achieves asymptotic optimality. This is elucidated through the comparison of its local quadratic error loss, which means its local quadratic error loss is asymptotically equivalent to the infeasible optimal averaging estimator. This implies that the feasible TVMA estimator is asymptotically optimal within the class of time-varying model averaging estimators constructed from a set of time-varying VAR candidate models. Furthermore, in Proposition 1 of Appendix B below, we have also shown that the infeasible TVMA estimator with $\widehat{\mathbf{w}}_t$ also satisfies the OPT property.

## 3.2 Convergence of weights

In this section, we establish the consistency of the TVMA estimators. Define the optimal weight $\mathbf{w}_t^0 = \operatorname{argmin}_{\mathbf{w}\in\mathcal{W}} \mathbb{E}L_{t,T}(\mathbf{w})$, and $\widetilde{\xi}_{t,T} = \min_{\mathbf{w}\in\mathcal{W}} \mathbb{E}L_{t,T}(\mathbf{w})$. The following conditions are presented for the establishment of Theorem 2.

**Condition (C.6).** *$S \to \infty$, $\lambda_T \to \infty$, $S^3\widetilde{\xi}_{t,T}^{-1}(Tl)^{-2\delta} = o(1)$ and $S^{3/2}\lambda_T\widetilde{\xi}_{t,T}^{-1/2}(Tl)^{-1/2-\delta} = o(1)$, where $\delta$ is a positive constant.*

**Condition (C.7).** *For some positive constants $\kappa_1$ and $\kappa_2$, we have*

$$0 < \kappa_1 < \zeta_{\min}\left(T^{-1}l^{-1}\mathbf{\Lambda}'\left(\mathbf{I}_K \otimes \mathbf{K}_t\right)\mathbf{\Lambda}\right) \leq \zeta_{\max}\left(T^{-1}l^{-1}\mathbf{\Lambda}'\left(\mathbf{I}_K \otimes \mathbf{K}_t\right)\mathbf{\Lambda}\right) < \kappa_2 < \infty,$$

*with probability approaching 1, where $\mathbf{\Lambda} = \left(\operatorname{vec}\left(\widehat{\boldsymbol{\mu}}^{(1)}\right), \ldots, \operatorname{vec}\left(\widehat{\boldsymbol{\mu}}^{(S)}\right)\right)$.*

**Condition (C.8).** $0 < \kappa_3 < \zeta_{\min}\left(\widehat{\mathbf{\Sigma}}_t^{-1}\right) \leq \zeta_{\max}\left(\widehat{\mathbf{\Sigma}}_t^{-1}\right) < \kappa_4 < \infty$ *with probability approaching 1.*

**<u>Remark</u> 5.** *Condition (C.6) restricts the relationship among $\widetilde{\xi}_{t,T}$, $Tl$, $\lambda_T$ and $S$. Similar conditions can be found in Li et al. (2022) and Sun et al. (2023). Condition (C.7) requires that the minimum and maximum singular values of $T^{-1}l^{-1}\mathbf{\Lambda}'\left(\mathbf{I}_K \otimes \mathbf{K}_t\right)\mathbf{\Lambda}$ are asymptotically bounded, which is common in the existing literature (Liao et al., 2019). For instance, if $\mathbf{K}_t$ reduces to an identity matrix, Condition (C.7) simplifies to: $0 < \kappa_1 < \zeta_{\min}\left(T^{-1}\mathbf{\Lambda}'\mathbf{\Lambda}\right) \leq \zeta_{\max}\left(T^{-1}\mathbf{\Lambda}'\mathbf{\Lambda}\right) < \kappa_2 < \infty$, which is identical to Condition (C.9) in Liao et al. (2019). Condition (C.8) imposes regularity assumptions on the minimum and maximum singular values of $\widehat{\mathbf{\Sigma}}_t^{-1}$, which follows from $0 < \zeta_{\min}\left(\mathbf{\Sigma}_t\right) \leq \zeta_{\max}\left(\mathbf{\Sigma}_t\right) < \infty$ under Condition (C.5).*

**<u>Theorem</u> 2.** *Let Conditions (C.1)-(C.3), and Conditions (C.6)-(C.8) hold. Then there exists a local minimizer $\widehat{\mathbf{w}}_t^F$ of $\mathrm{TVMA}_{t,T}^F(\mathbf{w})$ such that*

$$\left\|\widehat{\mathbf{w}}_t^F - \mathbf{w}_t^0\right\| = O_p\left(\widetilde{\xi}_{t,T}^{1/2}(Tl)^{-1/2+\delta}\right), \tag{3.4}$$

*where $\delta$ is a positive constant given in Condition (C.6).*

Theorem 2 shows that the feasibly selected $\widehat{\mathbf{w}}_t^F$ converges to the optimal weight $\mathbf{w}_t^0$ at the rate $\widetilde{\xi}_{t,T}^{1/2}(Tl)^{-1/2+\delta}$ for any given time $t$. The slower the rate of $\widetilde{\xi}_{t,T} \to \infty$, the faster the rate of $\widehat{\mathbf{w}}_t$ converging to $\mathbf{w}_t^0$ as $T \to \infty$. Additionally, it is worth noting that $\widetilde{\xi}_{t,T} \to \infty$

is not a necessary condition for Theorem 2, although it is permitted here. Furthermore, in Proposition 2 of Appendix B, we have also established that the convergence of the infeasibly selected weights $\widehat{\mathbf{w}}_t$ remains true for any fixed time $t$.

We then would like to comment that the proposed averaging prediction asymptotically allocates all weights to the correctly specified models if they are included in the candidate model set (i.e., the true models and over-parameterized models). Let $\mathcal{D}$ be the subset of $\{1, \ldots, S\}$ that consists of the indices of the correctly specified candidate models and $\mathcal{D}^c$ be the complement of $\mathcal{D}$. Define

$$\mathcal{W}^* = \left\{ \mathbf{w} \in [0,1]^S : \sum_{s \in \mathcal{D}^c} w^{(s)} = 1 \text{ and} \sum_{s \in \mathcal{D}} w^{(s)} = 0 \right\}$$

and $\xi^*_{t,T} = \inf_{\mathbf{w} \in \mathcal{W}^*} \mathbb{E} L^*_{t,T}(\mathbf{w})$. We further need the following condition.

**Condition (C.9).** *$S \to \infty$, $\lambda_T \to \infty$, $S\lambda_T \left(\xi^*_{t,T}\right)^{-1} \to 0$, and $S^2 T^{1/2} l^{1/2} \left(\xi^*_{t,T}\right)^{-1} \to 0$.*

**Remark 6.** *Condition (C.9) restricts the growth rate of the minimum risk when we consider only misspecified models as candidate models. Condition (C.9) is equivalent to Condition (C.4) when $\mathcal{D}$ is empty, that is, all candidate models are misspecified and then $\xi^*_{t,T} = \xi_{t,T}$. In other words, Condition (C.4) is a special case of Condition (C.9). The similar conditions can be found in Sun et al. (2023) for the univariate time–varying setting.*

**Theorem 3.** *Suppose Conditions (C.1)-(C.3), (C.5)′ and (C.9) hold and $\mathcal{D}$ is non-empty. Then for any given time $t$, we have $\sum_{s \in \mathcal{D}} \widehat{w}_t^{F(s)} \xrightarrow{p} 1$ as $T \to \infty$, where $\widehat{w}_t^{F(s)}$ is the $s$–th element of $\widehat{\mathbf{w}}_t^F$.*

Theorem 3 demonstrates that when there are correctly specified models, the proposed feasible TVMA criteria asymptotically assign all weights to the correctly specified models. If there is only one correctly specified model, the proposed feasible TVMA would asymptotically select this correctly specified model, which is analogous to the consistency property of model selection. Furthermore, in Proposition 3 of Appendix B, we have also shown that the asymptotic convergence of the infeasibly selected weights $\widehat{\mathbf{w}}_t$ remains true for any given time $t$.

**Remark 7.** *Note that we allow DGP to accommodate structural breaks, i.e.,* $\mathbf{A}_j(\tau_t) = \sum_{i=1}^{K_0+1} \mathbf{A}_{j,i}(\tau_t) \mathbb{I}(r_{i-1} < \tau_t \leq r_i)$, *where* $\mathbb{I}(\cdot)$ *denotes the indicator function, each element of* $\mathbf{A}_{j,i}(\tau)$ *is second order continuously differentiable for* $\tau \in [0, 1]$, $K_0$ *is a finite and positive integer that represents the number of (unknown) structural breaks, and* $0 < r_1 < \cdots < r_{K_0} < 1$ *are the scaled time points of the structural breaks with* $r_0 = 0$ *and* $r_{K_0+1} = 1$. *By allowing for discontinuities across some regimes, we can deal with significant events such as COVID-19 pandemic. Theorems 1 and 2 remain valid to DGP with structural breaks, when all candidate models are misspecified and parameters are estimated using local linear estimation.*

*When the candidate model incorporates structural breaks, Theorem 3 continues to hold. Consider the scenario where the correctly specified model has a single break. If the break date is known in advance, one can estimate the parameters of the candidate model by dividing the sample and conducting separate estimations before and after the break date. In cases where the break date is unknown, a consistent estimate of the breakpoint can first be obtained, followed by parameter estimation based on the estimated break date.*

## 3.3 Asymptotic distribution

In this subsection, we analyze the scenario where the set of nested candidate models encompasses under-fitted specifications, over-fitted specifications, as well as the true models. For simplicity, we assume that the $(S_0+1)$–th model is the true model, and the $s$–th candidate model is under-fitted if $s < S_0 + 1$, while it is over-fitted if $s > S_0 + 1$. In this subsection, we assume that the number of candidate models $S$ is large but fixed. In other subsections, however, $S$ is allowed to diverge with the sample size.

**Condition (C.10).** *For any* $s$, $\mathbf{A}^{(s)*}(\tau)$ *is second-order continuously differentiable on* $[0, 1]$.

**Condition (C.11).** *For any given time* $t$, $\sqrt{Tl}\left[\mathrm{vec}\Big(\widehat{\mathbf{A}}^{(s)}(\tau_t) - \mathbf{A}^{(s)*}(\tau_t) - \frac{1}{2}l^2\widetilde{c}_2\ddot{\mathbf{A}}^{(s)*}(\tau_t)\Big)\right] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_t^{(s)})$ *in the* $s$*–th candidate model, where* $\boldsymbol{\Omega}_t^{(s)}$ *is a finite, symmetric and positive definite matrix,* $\widetilde{c}_2 = \int_{-1}^{1} u^2 k(u) du$ *and* $\ddot{\mathbf{A}}^{(s)*}$ *is the second-order derivative of* $\mathbf{A}^{(s)*}$.

**Condition (C.12).** $\lambda_T \to \infty$ *and* $\lambda_T T^{-\frac{1}{2}} l^{-\frac{1}{2}} \to 0$.

**Remark** 8. *Condition (C.10) is the same as Assumption 1 in Gao et al. (2024b), which allows parameter values to change slowly in different small segments of each continuity time point. Similar conditions are found in Sun et al. (2021, 2023) with $K = 1$ and local constant estimators. Condition (C.11) pertains to the asymptotic distribution of the local linear kernel estimation method established in Theorem 1.2 of Gao et al. (2024b), where $S$ is large but fixed. Condition (C.12) implies the relationship between $\lambda_T$ and $Tl$, which is equivalent to Condition (C.6) in Sun et al. (2023). For example, $\lambda_T = \ln(Tl)$, Condition (C.12) holds, which is consistent with Zhang et al. (2020) in time-invariant setting.*

**Lemma** 1. *Let Conditions (C.1)-(C.3), (C.5)′, (C.10) and (C.12) hold. Then for any given time $t$, we have $\widehat{w}_t^{F(s)} = o_p(T^{-\frac{1}{2}}l^{-\frac{1}{2}})$ for $s < S_0 + 1$.*

Lemma 1 implies that, under certain regularity conditions, for any given time $t$, the time-varying weights for all under-fitted models, which exclude regressors with nonzero parameters, converge to zero at a faster rate than stated in Theorem 2. This is consistent with Theorem 1 in Zhang & Liu (2019) for univariate regressions with time-invariant parameters and Lemma 3 in Racine et al. (2023) for nonparametric spline regression models.

**Lemma** 2. *Let Conditions (C.1)-(C.3), (C.5)′, (C.10) and (C.12) hold. Then for any given time $t$, we have $\widehat{w}_t^{F(s)} = O_p(\lambda_T^{-1})$ for $s > S_0 + 1$.*

Lemma 2 demonstrates that the proposed TVMA weights for over-fitted models, which include irrelevant variables but do not omit any relevant ones, are $O_p(\lambda_T^{-1})$. This implies that the TVMA estimator asymptotically assigns zero weight to all over-fitted models as $\lambda_T \to \infty$.

**Theorem** 4. *Let Conditions (C.1)-(C.3), (C.5)′ and (C.10)-(C.12) hold. Then for any given time $t$, we have*

$$\sqrt{Tl}\left[\operatorname{vec}\left(\widehat{\mathbf{A}}\left(\tau_t, \widehat{\mathbf{w}}_t^F\right) - \mathbf{A}\left(\tau_t\right) - \frac{1}{2}l^2\widetilde{c}_2\ddot{\mathbf{A}}\left(\tau_t\right)\right)\right] \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega}_t^{(S_0+1)}), \text{ as } T \to \infty.$$

Theorem 4 establishes that the proposed TVMA estimator is asymptotically normal with the same covariance matrix as that for the local linear estimator under the true model (i.e., $(S_0 + 1)$–th model). Its proof is given in Appendix C of the supplementary document.

# 4 TVMA for time-varying VARX models

In this section, we discuss how to extend our approach to a wider class of time–varying VAR models. Consider the following DGP:

$$\mathbf{y}_t = \mathbf{a}\left(\tau_t\right) + \sum_{j=1}^{\infty} \mathbf{A}_j(\tau_t)\mathbf{y}_{t-j} + \mathbf{A}_{\mathbf{x}}(\tau_t)\mathbf{x}_t + \boldsymbol{\epsilon}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \tag{4.1}$$

where $\mathbf{x}_t$ is a $m$-dimensional vector of stationary exogenous variables and $m$ is fixed, $\{\mathbf{A}_{\mathbf{x}}\left(\tau\right)\}$ are the $K \times m$ coefficient matrices and $\{\mathbf{A}_{\mathbf{x}}\left(\tau\right)\} = \{\mathbf{A}_{\mathbf{x}}\left(0\right)\}$ for $\tau < 0$. We also assume that the roots of $\mathbf{I}_d - \sum_{j=1}^{\infty} \mathbf{A}_{\mathbf{y},j}\left(\tau\right) L^j = \mathbf{0}$ all lie outside the unit circle uniformly in $\tau \in [0, 1]$. The innovation setting follows Section 2.1.

Consider the $s$–th $(1 \leq s \leq S)$ candidate model

$$\mathbf{y}_t = \mathbf{a}\left(\tau_t\right) + \sum_{j=1}^{p_s} \mathbf{A}_j(\tau_t)\mathbf{y}_{t-j} + \mathbf{A}_{\mathbf{x}}(\tau_t)\mathbf{x}_t^{(s)} + \boldsymbol{\epsilon}_t^{(s)} = \mathbf{A}^{(s)\prime}(\tau_t)\mathbf{z}_t^{(s)} + \boldsymbol{\epsilon}_t^{(s)}, \tag{4.2}$$

where $\mathbf{z}_t^{(s)} = \left[1, \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p_s}', \mathbf{x}_t^{(s)\prime}\right]'$, $\mathbf{A}^{(s)}(\tau_t) = [\mathbf{a}(\tau_t), \mathbf{A}_1(\tau_t), \cdots, \mathbf{A}_{p_s}(\tau_t), \mathbf{A}_{\mathbf{x}}(\tau_t)]'$, $\mathbf{x}_t^{(s)}$ is a subset of $\mathbf{x}_t$ and $S+1 \leq t \leq T$. Note that our framework covers VAR as well as non-nested VARX models with time-varying parameters. For each candidate model, the time-varying parameters are estimated by a local linear method, which is the same as (2.4) in Section 2. Based on (2.7)-(2.8) and (2.10)-(2.11), we obtain the corresponding infeasible time-varying weight vector $\widehat{\mathbf{w}}_t$ and feasible time-varying weight vector $\widehat{\mathbf{w}}_t^F$.

When the VAR-based framework is generalized to the VARX-based framework by incorporating exogenous predictors, the corresponding asymptotic properties are analogous to those derived for the VAR-based framework. This is formally stated in the following Corollary.

**Corollary 1.** *Let the dimension of $\mathbf{x}_t$ increase to infinity at a rate slower than or equal to the largest lag $S$. Then the conclusions of Theorems 1-3 remain valid for any given time $t$ when employing the time-varying vector of weights $\widehat{\mathbf{w}}_t^F$ in the context of time-varying VARX models.*

If the dimension $m$ of $\mathbf{x}_t$ increases to infinity at a rate slower than or equal to the largest lag $S$, Conditions (C.1)- (C.12) still hold, preserving the validity of Theorems 1-3; see more discussions on Appendix C. Moreover, if $m$ grows faster than $S$, we expect similar asymptotic properties with some modified conditions. For instance, the convergence rate in Condition (C.1) should change to $(S+m)^{1/2}T^{-1/2}l^{-1/2}$.

**Remark 9.** *In data-rich environments with numerous predictors, factor models provide an appealing way to summarize information. These models assume a small set of latent common factors can explain a large portion of the comovement among observed variables, effectively reducing dimensionality while capturing core driving forces behind the data interdependencies. The proposed approach encompasses time-varying FAVAR models; see Appendix A for details.*

# 5 Simulation studies

This section compares the finite sample performance of the proposed TVMA method against several competing model selection and averaging methods.

## 5.1 Competing methods

We consider the following alternative model selection and averaging methods, including information criterition (IC) of Gao et al. (2024b), AIC, BIC, Hannan–Quinn (HQ) of Tsay (2013), smoothed AIC (sAIC), smoothed BIC (sBIC), smoothed HQ (sHQ), and simple averaging (SA). Specifically, for the $s$–th candidate model, the IC used in Gao et al. (2024b) is $\mathrm{IC}^{(s)} = \ln(\mathrm{RSS}(s)) + s\chi_T$, where $\mathrm{RSS}(s) = \frac{1}{T-S}\sum_{t=S+1}^{T}(\mathbf{y}_t - \widehat{\boldsymbol{\mu}}_t^{(s)})'(\mathbf{y}_t - \widehat{\boldsymbol{\mu}}_t^{(s)})$, $\chi_T$ is the penalty term, and $\chi_T = \max\left\{l^4, \frac{\ln(T-S)}{(T-S)l}\right\} \cdot \ln(\ln((T-S)l))$ based on Gao et al. (2024b).

In addition, we define $\widehat{\mathbf{\Sigma}}^{(s)} = (T-S)^{-1}\left(\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{(s)}\right)'\left(\mathbf{Y} - \widehat{\boldsymbol{\mu}}^{(s)}\right)$ as the residual covariance matrix from the $s$–th candidate model. Then for the $s$–th candidate model, the three popular order selection criteria (i.e., AIC, BIC, and HQ) are expressed as $\mathrm{AIC}^{(s)} = \ln\left|\widehat{\mathbf{\Sigma}}^{(s)}\right| + 2sK^2/(T-S)$, $\mathrm{BIC}^{(s)} = \ln\left|\widehat{\mathbf{\Sigma}}^{(s)}\right| + (\ln(T-S))sK^2/(T-S)$, and $\mathrm{HQ}^{(s)} = \ln\left|\widehat{\mathbf{\Sigma}}^{(s)}\right| + 2(\ln\ln(T-S))sK^2/(T-S)$. Furthermore, sAIC assigns the weight $w_{\mathrm{AIC}}^{(s)} = \frac{\exp(-\mathrm{AIC}^{(s)}/2)}{\sum_{s=1}^{S}\exp(-\mathrm{AIC}^{(s)}/2)}$ to the $s$–th model, and the weights of sBIC and sHQ are defined similarly. We compare the TVMA approach with these aforementioned competing methods.

## 5.2 Multistep prediction and forecast evaluation

For time series models, rolling forecasting is commonly employed for out-of-sample prediction. More specifically, for the $s$–th ($1 \leq s \leq S$) candidate model, the $h$–step-ahead out-of-sample rolling prediction of $\mathbf{y}_{T+h}$ is defined as

$$\widehat{\mathbf{y}}_{T+h}^{(s)} = \widehat{\mathbf{a}}^{1,s}(\tau_T) + \sum_{j=1}^{s} \widehat{\mathbf{A}}_j^{1,s}(\tau_T)\, \widehat{\mathbf{y}}_{T+h-j}^{*}, \tag{4.3}$$

where $\widehat{\mathbf{y}}_{T+h-j}^{*} = \mathbf{y}_{T+h-j}$ for $h \leq j$, and $\widehat{\mathbf{A}}_j^{1,s}$ together with $\widehat{\mathbf{a}}^{1,s}$ are estimated based on

$$\mathbf{y}_t = \mathbf{a}(\tau_t) + \sum_{j=1}^{s} \mathbf{A}_j(\tau_t)\, \mathbf{y}_{t-j} + \boldsymbol{\epsilon}_t^{(s)}, S+1 \leq t \leq T. \tag{4.4}$$

To compare the forecast accuracies of different model selection and averaging methods, we compute the ratio of the root mean squared prediction errors (RMSPE),

$$\text{RMSPE} = \frac{\sqrt{\sum_{d=1}^{D} \left\| \widehat{y}_{T+h,k}^{d} - y_{T+h,k}^{d} \right\|^2 / D}}{\sqrt{\sum_{d=1}^{D} \left\| \widehat{y}_{T+h,k,\text{AIC}}^{d} - y_{T+h,k}^{d} \right\|^2 / D}}, \tag{4.5}$$

where $\widehat{y}_{T+h,k}^{d}$ and $y_{T+h,k}^{d}$ respectively denote the forecasts from our interested method and the true value in the $d$th replication, $\widehat{y}_{T+h,k,\text{AIC}}^{d}$ is the forecast from the AIC method which serves as a popular benchmark approach, and $D$ is the number of replications. A method with an RMSPE less than 1 outperforms the AIC method. The lower the RMSPE value, the better the performance of the method under consideration.

## 5.3 Simulation design and results

We consider two DPGs as follows:

DGP1 (Time-varying VAR in Gao et al. (2024b)):

$$\mathbf{y}_t = \mathbf{a}(\tau) + \mathbf{A}_1(\tau)\mathbf{y}_{t-1} + \mathbf{A}_2(\tau)\mathbf{y}_{t-2} + \boldsymbol{\eta}_t$$

with $\boldsymbol{\eta}_t = \boldsymbol{\omega}(\tau)\boldsymbol{\epsilon}_t$ for $t = 1, \ldots, T$, where $\boldsymbol{\epsilon}_t$'s are iid draws from $N(\mathbf{0}_{2\times 1}, \mathbf{I}_2)$,
$\boldsymbol{\omega}(\tau) = \begin{bmatrix} 1.5 + 0.2\exp(0.5-\tau) & 0 \\ 0.1\exp(0.5-\tau) & 1.5 + 0.5(\tau-0.5)^2 \end{bmatrix}$, $\mathbf{a}(\tau) = (0.5\sin(2\pi_\tau \tau), 0.5\cos(2\pi_\tau \tau))'$,

$\mathbf{A}_1(\tau) = \begin{bmatrix} 0.8\exp(-0.5+\tau) & 0.8(\tau-0.5)^3 \\ 0.8(\tau-0.5)^3 & 0.8+0.3\sin(\pi\tau) \end{bmatrix}$, and $\mathbf{A}_2(\tau) = \begin{bmatrix} -0.2\exp(-0.5+\tau) & 0.8(\tau-0.5)^2 \\ 0.8(\tau-0.5)^2 & -0.4+0.3\cos(\pi\tau) \end{bmatrix}$. We generate $\pi_t$ by $\pi_t = \epsilon_{t,1} + e_t$, where $\{e_t\}$ is a sequence of iid standard normal variables.

DGP2 (Bivariate time-varying ARMA (2,1)):

$$\mathbf{y}_t = \mathbf{a}(\tau) + \mathbf{A}_1(\tau)\mathbf{y}_{t-1} + \mathbf{A}_2(\tau)\mathbf{y}_{t-2} + \boldsymbol{\eta}_t - \boldsymbol{\theta}\boldsymbol{\eta}_{t-1},$$

with $\boldsymbol{\theta} = \begin{bmatrix} -0.6 & 0.3 \\ 0.3 & 0.6 \end{bmatrix}$ which is a time–varying version of Liao et al. (2019), and the other time-varying parameters are the same as in DGP1.

In this context, DGP1 is a time-varying VAR model with finite lag order while DGP2 can be transformed into a time-varying VAR model with an infinite lag order. Let $K = 2$, $T =$ 100, 300 and 500, and $D$=1000. For $T = 100$, we set $S = 5$, and for $T = 300$ and 500, we set $S = 10$. The choice of $S = 5$ when $T = 100$ is to ensure that we can efficiently estimate the associated coefficients using local linear method, which is consistent with Liao et al. (2019). The forecast horizon $h$ =1, 2, 3 and 4. For simplicity, we follow the spirit of Zhang et al. (2020) to set $\lambda_T = 2\ln(Tl)$, due to the local linear estimation of time-varying coefficients with $2s$ regressors for the $s$–th candidate model.

Table 1 reports the forecast results under DGP1 with different forecast horizons $h$ and vector dimensions $k$ (where $k = 1, 2$). First, TVMA estimator is the top-performing strategy in the vast majority of cases. For instance, given $T = 100$, $S = 5$, $k = 1$ and $h = 1$, the RMSPE of TVMA is 0.3632, while the RMSPE of the second-best approach is 0.4101, indicating an improvement of approximately 11%. When $T = 100$, $S = 5$, $k = 2$ and $h = 4$, the RMSPE of TVMA is 0.1075, whereas the RMSPE of the second-best approach is 0.1676, suggesting an improvement of about 36%. Besides, the IC method in Gao et al. (2024b) generally has the second-lowest RMSPE, while the AIC, BIC, and HQ methods consistently perform poorly. Furthermore, the SA outperforms the sAIC, sBIC, and sHQ model averaging methods with time-invariant weights, but it performs worse than the proposed TVMA. This underscores the importance of time-varying combination weights and reinforces the asymptotic optimality stated in Theorem 1.

Table 2 reports the RMSPE results under DGP2, which allows for time-dependence among innovations and potential endogeneity. Similarly, the average improvements in forecasts using our TVMA method exist across almost all forecast horizons. However, under DGP2, the IC method generally does not outperform the sAIC, sBIC, and sHQ methods. Additionally, when the sample size $T$ is large, the sAIC, sBIC, and sHQ methods often perform better than the traditional model selection methods, including AIC, BIC, and HQ. One possible explanation is that the DGP is a time-varying VAR model with an infinite lag order, which complicates model selection. Consequently, model averaging methods tend to achieve better forecasting performance than model selection methods.

For roubstness check, we consider different bandwidths and present the related RMSPE results in Tables D.1-D.6 of Appendix D. The bandwidth selection is based on the rule $l = cT^{-1/5}$, where $c$ takes the values 0.75, 1, and 1.25 respectively. It is found that the forecast performance of TVMA is robust to the choice of bandwidths. Overall, the TVMA method consistently delivers the best forecasting performance, while the AIC, BIC and HQ methods frequently produce the poorest results.

# 6 Empirical application

## 6.1 Prediction on U.S. macroeconomic dynamic systems

In this section, we employ our TVMA method to the quarterly US macroeconomic data set from 1959Q1 to 2015Q4, previously analyzed by Hansen (2016) and Liao et al. (2019). The dependent variables include Gross Domestic Product (GDP), the GDP deflator (GDPD), and Federal Funds Rates (FF). Consistent with Liao et al. (2019), we implement the same data transformations detailed therein, set $T = 100$ and use the same out-of-sample periods spanning from 1984:Q2+h-1 to 2005:Q4 for $h \in \{1, 2, 3, 4\}$. We compare our method against IC, AIC, BIC, HQ, sAIC, sBIC, sHQ and SA, with all methods evaluated on the same set of candidate nested time-varying VAR models with $S = 5$. Futhermore, the out-of-sample period is extendedto 2015Q4. The forecast evaluation criterion used here is the root mean squared prediction error (rMSPE).

Tables 3 and 4 report the out-of-sample forecast performance for GDP, GDPD, and FF forecasts using various methods. For ease of comparison, the best performing strategy in each case is highlighted in bold. First, it is observed that the TVMA method is consistently ranked first against the other 8 competing models during different forecast periods. The Diebold-Mariano test $p$-values reported indicate that the TVMA's relative reductions in rMSFE are statistically significant in a majority of cases. For instance, our proposed approach yields a near two-fold improvement in out-of-sample forecast precision compared to the IC method of Gao et al. (2024b) at the one-step-ahead forecast horizon. Second, we document that the TVMA forecasts frequently dominate the SA forecasts, especially at shorter horizons (e.g., $h = 1, 2$). This finding contradicts the forecast combination puzzle of Stock & Watson (2004), wherein simple averaging outperforms sophisticated adaptive weighting schemes. For GDP forecasts at $h = 1$, the TVMA's gains range from 19%–54% relative to the SA benchmark. A potential explanation is that all candidate models are likely misspecified to some degree, with their relative forecast rankings varying over time. By allowing the weights to adapt optimally, the TVMA can upweight models during periods when they perform well while downweighting them when they underperform. Furthermore, the model selection criteria, including AIC, BIC and HQ, consistently yield the poorest predictive performance. This highlights the merits of reducing model uncertainty and increasing the robustness of model averaging.

## 6.2 Prediction on government spending and GDP

In this section, we apply our method to the U.S. government spending and GDP data analyzed by Gao et al. (2024b), Blanchard & Perotti (2002), and Ramey & Zubairy (2018). The prediction of these two variables is crucial, as we follow Ramey & Zubairy (2018) and Gao et al. (2024b) to measure government spending multipliers, i.e., the change in output resulting from a one-dollar change in government spending. It is widely acknowledged that government spending multipliers are essential for fiscal policy analysis since they gauge the extent to which government purchases can stimulate private activity (Ramey & Zubairy, 2018). The dataset comprises observations from 1954:Q1 to 2015:Q4. We consider two variables time-varying VAR inlcuding government spending and real per capita GDP, both adjusted by dividing by trend GDP.

The optimal lag based on IC is set to 2, i.e., $\widehat{s} = 2$, consistent with Gao et al. (2024b). The existing literature often assumes the lag to vary from 2 to 4, so we set the maximum lag order S to 5 in our candidate models. We follow Sun et al. (2021, 2023) by using rolling estimation to forecast the two variables with different forecast horizons, i.e., $h$ = 1, 2, 3 and 4. We use rMSPE to evaluate the performance of TVMA methods and other alternative methods. The out-of-sample begins from 1985Q1, 1990Q1, 1995Q1, 2000Q1 and 2005Q1, and all end at 2015Q4. Furthermore, we also consider two forecast periods: the Great Moderation from 1983Q1 to 2006Q4 and post-Great Moderation from 2007Q1 to 2015Q4.

The forecast results are reported in Tables 5-8. First, it is evident that the rMSFE obtained from the TVMA approach produces the smallest values across all seven forecast periods at different forecast horizons. The $p$-value of the Diebold-Mariano test is almost smaller than 10%, which provides strong evidence for the superior performance of the TVMA method to other competing methods. For instance, in Table 5, when $h = 4$, TVMA's predictive performance is approximately 18% to 33% better than that of IC, the second-best method. Similarly, in Table 6, the improvement ranges from 21% to 26% at the forecast horizon $h = 4$. Second, our approach tends to exhibit larger gains over model averaging schemes with time-varying parameters but constant weights during periods of heightened economic fluctuations. For instance, relative to methods allowing time-varying coefficients but constant weights (sAIC, sBIC, sHQ), the TVMA's forecast accuracy gains for GDP at $h = 1$ are substantial at 30% during the volatile "Great Moderation" period, but more modest at 10% in the "Post Great Moderation". This highlights the importance of time-varying weights. A potential explanation is that individual model's forecast performance can shift over time, especially in the presence of structural instabilities. Furthermore, the IC approach frequently delivers the second-best forecast performance in most cases. The sAIC, sBIC, and sHQ methods outperform the related model selections AIC, BIC, and HQ across all forecast horizons and periods.

## 7 Conclusion

This paper proposes a novel penalized time-varying model averaging method for a class of VAR models associated with time–varying features, allowing for the forecast combination weights to evolve over time. Under a set of general and easily verifiable conditions, we establish asymptotic properties, including asymptotic optimality, consistency, and asymptotic normality when both the number of candidate models and the dimension of regressors diverge. To our knowledge, we are the first to show that when the set of candidate models includes correctly specified models, our method consistently assigns all weight to these optimal choices and establish the asymptotic distributions of the combined parameter estimators. Furthermore, we demonstrate the suitability of the proposed TVMA for time-varying VARX and time-varying FAVAR frameworks. Simulation studies and empirical applications illustrate the superior performance of TVMA compared to other competing model averaging and model selection methods.

The forecast combination method developed in this paper opens up several avenues for future research. One interesting direction is to develop a time-varying cross-validation approach to model averaging, which may provide additional benefits in the presence of highly persistent dependence, albeit at the cost of greater computational complexity. Additionally, this paper has only considered a global bandwidth for the TVMA estimator, which may be severely affected by the existence of structural changes. It would be interesting to use a localized bandwidth at each time point.

## References


Ando, T. & Li, K. (2017). A weight-relaxed model averaging approach for high-dimensional generalized linear models. *Annals of Statistics* **45**, 2654–2679.

Ando, T. & Li, K.-C. (2014). A model-averaging approach for high-dimensional regression. *Journal of the American Statistical Association* **109**, 254–265.

Blanchard, O. & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics* **117**, 1329–1368.

Chen, B. & Maung, K. (2023). Time-varying forecast combination for high-dimensional data. *Journal of Econometrics* **237**, 105418.

Chen, J., Li, D., Linton, O. & Lu, Z. (2018). Semiparametric ultra-high dimensional model averaging of nonlinear dynamic time series. *Journal of the American Statistical Association* **113**, 919–932.

Chen, L., Gao, J. & Vahid, F. (2022). Global temperatures and greenhouse gases: a common features approach. *Journal of Econometrics* **230**, 240–254.

Chen, Q., Hong, Y. & Li, H. (2024). Time-varying forecast combination for factor-augmented regressions with smooth structural changes. *Journal of Econometrics* **240**, 105693.

Cheng, X. & Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics* **186**, 280–293.

Fan, J. & Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics* **32**, 928–961.

Fu, Z., Su, L. & Wang, X. (2024). Estimation and inference on time-varying favar models. *Journal of Business & Economic Statistics* **42**, 533–547.

Gao, J. (2007). *Nonlinear Time Series: Semi– and Non–Parametric Methods.* Chapman & Hall/CRC.

Gao, J., Peng, B., Wu, W. B. & Yan, Y. (2024a). Time-varying multivariate causal processes. *Journal of Econometrics* **240**, 105671.

Gao, J., Peng, B. & Yan, Y. (2024b). Estimation, inference, and empirical analysis for time-varying VAR models. *Journal of Business & Economic Statistics* **42**, 310–321.

Gao, Y., Zhang, X., Wang, S. & Zou, G. (2016). Model averaging based on leave-subject-out cross-validation. *Journal of Econometrics* **192**, 139–151.

Gao, Z., Zou, J., Zhang, X. & Ma, Y. (2023). Frequentist model averaging for envelope models. *Scandinavian Journal of Statistics* **50**, 1325–1364.

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

Hansen, B. E. (2016). Stein combination shrinkage for vector autoregressions. *Manuscript, University of Wisconsin-Madison* .

Li, J., Lv, J., Wan, A. T. & Liao, J. (2022). Adaboost semiparametric model averaging prediction for multiple categories. *Journal of the American Statistical Association* **117**, 495–509.

Liao, J., Zong, X., Zhang, X. & Zou, G. (2019). Model averaging based on leave-subject-out cross-validation for vector autoregressions. *Journal of Econometrics* **209**, 35–60.

Liao, J., Zou, G., Gao, Y. & Zhang, X. (2021). Model averaging prediction for time series models with a diverging number of parameters. *Journal of Econometrics* **223**, 190–221.

Liao, J.-C. & Tsay, W.-J. (2020). Optimal multistep var forecast averaging. *Econometric Theory* **36**, 1099–1126.

Patton, A. J. (2006). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics* **21**, 147–173.

Paye, B. S. & Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance* **13**, 274–315.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* **72**, 821–852.

Racine, J. S., Li, Q., Yu, D. & Zheng, L. (2023). Optimal model averaging of mixed-data kernel-weighted spline regressions. *Journal of Business & Economic Statistics* **41**, 1251–1261.

Ramey, V. A. & Zubairy, S. (2018). Government spending multipliers in good times and in bad: evidence from us historical data. *Journal of Political Economy* **126**, 850–901.

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. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**, 405–430.

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

Stock, J. H. & Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of Macroeconomics*, vol. 2. Elsevier, pp. 415–525.

Su, L. & Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics* **198**, 84–101.

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

Sun, Y., Hong, Y., Wang, S. & Zhang, X. (2023). Penalized time-varying model averaging. *Journal of Econometrics* **235**, 1355–1377.

TSAY, R. S. (2013). *Multivariate Time Series Analysis: with R and Financial Applications.* John Wiley & Sons.

WAN, A. T. K., ZHANG, X. & ZOU, G. (2010). Least squares model averaging by Mallows criterion. *Journal of Econometrics* **156**, 277–283.

WHITE, H. (2014). *Asymptotic Theory for Econometricians.* Academic Press.

YAN, Y., GAO, J. & PENG, B. (2024). Asymptotics for time–varying vector MA($\infty$) processes. *Econometric Theory* **43**, 1–33.

YANG, Y. (2003). Regression with multiple candidate models: selecting or mixing? *Statistica Sinica* **33**, 783–809.

YU, D., LIAN, H., SUN, Y., ZHANG, X. & HONG, Y. (2024). Post-averaging inference for optimal model averaging estimator in generalized linear models. *Econometric Reviews* **43**, 98–122.

YU, D. & YAN, Y. (2023). Joint dynamics of stock returns and cash flows: A time-varying present-value framework. *Financial Management* **52**, 513–541.

YUAN, Z. & YANG, Y. (2005). Combining linear regression models: When and how? *Journal of the American Statistical Association* **100**, 1202–1214.

ZHANG, X. & LIU, C. A. (2019). Inference after model averaging in linear regression models. *Econometric Theory* **35**, 816–841.

ZHANG, X., WAN, A. T. K. & ZOU, G. (2013). Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics* **174**, 82–94.

ZHANG, X. & ZHANG, X. (2023). Optimal model averaging based on forward-validation. *Journal of Econometrics* **237**, 105295.

ZHANG, X., ZOU, G., LIANG, H. & CARROLL, R. J. (2020). Parsimonious model averaging with a diverging number of parameters. *Journal of the American Statistical Association* **115**, 972–984.

ZHU, R., WAN, A. T., ZHANG, X. & ZOU, G. (2019). A Mallows–type model averaging estimator for the varying-coefficient partially linear model. *Journal of the American Statistical Association* **114**, 882–892.

Table 1: Forecast evaluation RMSPE under DGP1

| $h$ | $k$ | TVMA | IC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|
| | | | | T=100, S=5 | | | | | |
| 1 | 1 | **0.3632** | 0.4101 | 0.9978 | 1.0000 | 0.4783 | 0.4596 | 0.4707 | 0.4329 |
| 1 | 2 | **0.3776** | 0.4518 | 0.9983 | 0.9997 | 0.4901 | 0.4713 | 0.4824 | 0.4424 |
| 2 | 1 | **0.2679** | 0.3066 | 0.9984 | 1.0000 | 0.3866 | 0.3671 | 0.3787 | 0.3416 |
| 2 | 2 | **0.2253** | 0.2755 | 0.9960 | 1.0000 | 0.3485 | 0.3281 | 0.3401 | 0.3008 |
| 3 | 1 | **0.1839** | 0.2189 | 0.9398 | 1.0000 | 0.3030 | 0.2814 | 0.2942 | 0.2509 |
| 3 | 2 | **0.1844** | 0.1962 | 0.9962 | 1.0000 | 0.2968 | 0.2739 | 0.2874 | 0.2365 |
| 4 | 1 | 0.0276 | **0.0230** | 0.9999 | 1.0000 | 0.2631 | 0.2393 | 0.2534 | 0.2010 |
| 4 | 2 | **0.1075** | 0.1676 | 0.9999 | 1.0000 | 0.2645 | 0.2408 | 0.2548 | 0.2069 |
| | | | | T=300, S=10 | | | | | |
| 1 | 1 | **0.4073** | 0.4481 | 0.9908 | 1.0000 | 0.5262 | 0.5132 | 0.5209 | 0.5036 |
| 1 | 2 | **0.6277** | 0.6509 | 0.9973 | 1.0000 | 0.7173 | 0.7088 | 0.7139 | 0.6995 |
| 2 | 1 | **0.4297** | 0.4601 | 0.9883 | 1.0000 | 0.5527 | 0.5404 | 0.5477 | 0.5311 |
| 2 | 2 | **0.5930** | 0.6263 | 0.9916 | 1.0000 | 0.6660 | 0.6568 | 0.6622 | 0.6450 |
| 3 | 1 | **0.4562** | 0.4692 | 0.9903 | 0.9999 | 0.6012 | 0.5881 | 0.5959 | 0.5774 |
| 3 | 2 | **0.7060** | 0.8265 | 0.9831 | 0.9999 | 0.7162 | 0.7124 | 0.7146 | 0.7093 |
| 4 | 1 | 0.5763 | **0.5748** | 0.9979 | 0.9999 | 0.6773 | 0.6678 | 0.6735 | 0.6597 |
| 4 | 2 | **0.6124** | 0.6314 | 0.9971 | 0.9998 | 0.7058 | 0.6978 | 0.7026 | 0.6907 |
| | | | | T=500, S=10 | | | | | |
| 1 | 1 | **0.4931** | 0.6147 | 0.9899 | 1.0004 | 0.6169 | 0.6078 | 0.6133 | 0.5922 |
| 1 | 2 | **0.7722** | 0.8424 | 0.9984 | 0.9999 | 0.8435 | 0.8382 | 0.8414 | 0.8186 |
| 2 | 1 | **0.6139** | 0.7625 | 0.9941 | 0.9995 | 0.7090 | 0.7016 | 0.7061 | 0.6836 |
| 2 | 2 | **0.8679** | 0.9463 | 1.0002 | 1.0000 | 0.9086 | 0.9047 | 0.9071 | 0.8856 |
| 3 | 1 | **0.6956** | 0.7586 | 0.9945 | 0.9999 | 0.7656 | 0.7599 | 0.7634 | 0.7469 |
| 3 | 2 | **0.8148** | 0.9485 | 0.9976 | 1.0000 | 0.8667 | 0.8615 | 0.8647 | 0.8408 |
| 4 | 1 | **0.6787** | 0.7657 | 0.9990 | 1.0000 | 0.7682 | 0.7615 | 0.7656 | 0.7370 |
| 4 | 2 | **0.5768** | 0.6046 | 0.9990 | 1.0000 | 0.6284 | 0.6238 | 0.6266 | 0.6137 |

Notes:

(1) Eight methods include TVMA, IC in Gao et al. (2024b), BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) RMSPE is computed as RMSPE $= \frac{\sqrt{\sum_{d=1}^{D}\left\|\widehat{y}_{T+h,k}^{d}-y_{T+h,k}^{d}\right\|^2/D}}{\sqrt{\sum_{d=1}^{D}\left\|\widehat{y}_{T+h,k,\mathrm{AIC}}^{d}-y_{T+h,k}^{d}\right\|^2/D}}$, where $\widehat{y}_{T+h,k}^{d}$ and $y_{T+h,k}^{d}$ respectively denote the forecasts from competing method and the true value in the $d$th replication, $\widehat{y}_{T+h,k,\mathrm{AIC}}^{d}$ is the forecast from the AIC method, and $D = 1000$ is the number of replications.

Table 2: Forecast evaluation RMSPE under DGP2

| $h$ | $k$ | TVMA | IC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|
| | | | | T=100, S=5 | | | | | |
| 1 | 1 | **0.3995** | 0.5598 | 0.9969 | 1.0000 | 0.4923 | 0.4746 | 0.4851 | 0.4383 |
| 1 | 2 | **0.4071** | 0.5498 | 0.9987 | 1.0000 | 0.5028 | 0.4857 | 0.4958 | 0.4501 |
| 2 | 1 | **0.2400** | 0.3229 | 0.9981 | 1.0000 | 0.3774 | 0.3566 | 0.3689 | 0.3198 |
| 2 | 2 | **0.2287** | 0.3655 | 0.9958 | 1.0000 | 0.3732 | 0.3517 | 0.3644 | 0.3122 |
| 3 | 1 | **0.1176** | 0.2116 | 0.9873 | 1.0000 | 0.2893 | 0.2664 | 0.2800 | 0.2259 |
| 3 | 2 | **0.0455** | 0.0821 | 0.9986 | 0.9999 | 0.2657 | 0.2424 | 0.2562 | 0.2036 |
| 4 | 1 | **0.1491** | 0.1811 | 0.9539 | 1.0000 | 0.2887 | 0.2654 | 0.2792 | 0.2168 |
| 4 | 2 | **0.1336** | 0.1743 | 0.9717 | 1.0000 | 0.2806 | 0.2569 | 0.2710 | 0.2102 |
| | | | | T=300, S=10 | | | | | |
| 1 | 1 | **0.4490** | 0.4624 | 0.9885 | 1.0000 | 0.5650 | 0.5529 | 0.5601 | 0.5365 |
| 1 | 2 | **0.6730** | 0.7102 | 0.9953 | 1.0000 | 0.7254 | 0.7169 | 0.7220 | 0.7004 |
| 2 | 1 | **0.4243** | 0.4275 | 0.9896 | 0.9996 | 0.5647 | 0.5513 | 0.5593 | 0.5323 |
| 2 | 2 | **0.6159** | 0.6517 | 0.9955 | 0.9998 | 0.7207 | 0.7105 | 0.7166 | 0.6878 |
| 3 | 1 | 0.4238 | **0.4204** | 0.9942 | 0.9999 | 0.5600 | 0.5471 | 0.5548 | 0.5293 |
| 3 | 2 | **0.6270** | 0.6391 | 0.9946 | 0.9999 | 0.6767 | 0.6690 | 0.6736 | 0.6583 |
| 4 | 1 | **0.5142** | 0.5316 | 0.9935 | 1.0003 | 0.6288 | 0.6182 | 0.6245 | 0.6031 |
| 4 | 2 | **0.5745** | 0.5872 | 1.0008 | 1.0000 | 0.6501 | 0.6405 | 0.6462 | 0.6247 |
| | | | | T=500, S=10 | | | | | |
| 1 | 1 | **0.5046** | 0.5920 | 0.9969 | 0.9992 | 0.6132 | 0.6046 | 0.6098 | 0.5864 |
| 1 | 2 | **0.8317** | 0.9172 | 0.9992 | 1.0000 | 0.8743 | 0.8699 | 0.8726 | 0.8433 |
| 2 | 1 | **0.5597** | 0.6880 | 0.9959 | 0.9999 | 0.6723 | 0.6639 | 0.6690 | 0.6391 |
| 2 | 2 | **0.6908** | 0.7988 | 0.9996 | 1.0000 | 0.7542 | 0.7484 | 0.7519 | 0.7201 |
| 3 | 1 | **0.6106** | 0.7829 | 0.9968 | 0.9993 | 0.6818 | 0.6746 | 0.6789 | 0.6490 |
| 3 | 2 | **0.7300** | 0.8689 | 1.0003 | 1.0001 | 0.7729 | 0.7675 | 0.7707 | 0.7405 |
| 4 | 1 | **0.6258** | 0.8780 | 0.9982 | 1.0000 | 0.7149 | 0.7070 | 0.7118 | 0.6741 |
| 4 | 2 | 0.7030 | 0.9687 | 0.9993 | 1.0000 | 0.6813 | 0.6766 | 0.6794 | **0.6594** |

Notes:

(1) Eight methods include TVMA, IC in Gao et al. (2024b), BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) RMSPE is computed as RMSPE $= \frac{\sqrt{\sum_{d=1}^{D}\left\|\widehat{y}_{T+h,k}^{d}-y_{T+h,k}^{d}\right\|^2/D}}{\sqrt{\sum_{d=1}^{D}\left\|\widehat{y}_{T+h,k,\mathrm{AIC}}^{d}-y_{T+h,k}^{d}\right\|^2/D}}$, where $\widehat{y}_{T+h,k}^{d}$ and $y_{T+h,k}^{d}$ respectively denote the forecasts from competing method and the true value in the $d$th replication, $\widehat{y}_{T+h,k,\mathrm{AIC}}^{d}$ is the forecast from the AIC method, and $D = 1000$ is the number of replications.

Table 3: Forecast evaluation of different methods during 1984Q2-2008Q4

| $h$ | | TVMA | IC | AIC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GDP | | | | |
| 1 | rMSPE | **0.0068** | 0.0119 | 0.0141 | 0.0141 | 0.0141 | 0.0094 | 0.0089 | 0.0092 | 0.0084 |
| | $p$-value | | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0001 | 0.0004 |
| 2 | rMSPE | **0.0065** | 0.0132 | 0.0169 | 0.0169 | 0.0169 | 0.0094 | 0.0087 | 0.0091 | 0.0081 |
| | $p$-value | | 0.0058 | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0009 | 0.0006 | 0.0047 |
| 3 | rMSPE | **0.0069** | 0.0184 | 0.0223 | 0.0223 | 0.0223 | 0.0112 | 0.0102 | 0.0107 | 0.0091 |
| | $p$-value | | 0.0336 | 0.0050 | 0.0050 | 0.0050 | 0.0032 | 0.0046 | 0.0036 | 0.0078 |
| 4 | rMSPE | **0.0079** | 0.0296 | 0.0336 | 0.0336 | 0.0336 | 0.0154 | 0.0137 | 0.0147 | 0.0121 |
| | $p$-value | | 0.1421 | 0.0647 | 0.0647 | 0.0647 | 0.0293 | 0.0266 | 0.0282 | 0.0264 |
| | | | | | | GDPD | | | | |
| 1 | rMSPE | **0.0019** | 0.0035 | 0.0041 | 0.0041 | 0.0041 | 0.0027 | 0.0026 | 0.0026 | 0.0024 |
| | $p$-value | | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | rMSPE | **0.0025** | 0.0049 | 0.0062 | 0.0062 | 0.0062 | 0.0037 | 0.0034 | 0.0036 | 0.0032 |
| | $p$-value | | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0003 |
| 3 | rMSPE | **0.0029** | 0.0069 | 0.0083 | 0.0083 | 0.0083 | 0.0044 | 0.0041 | 0.0043 | 0.0038 |
| | $p$-value | | 0.0098 | 0.0014 | 0.0014 | 0.0014 | 0.0050 | 0.0082 | 0.0061 | 0.0178 |
| 4 | rMSPE | **0.0026** | 0.0103 | 0.0133 | 0.0133 | 0.0133 | 0.0058 | 0.0051 | 0.0055 | 0.0046 |
| | $p$-value | | 0.0459 | 0.0136 | 0.0136 | 0.0136 | 0.0090 | 0.0119 | 0.0099 | 0.0186 |
| | | | | | | FF | | | | |
| 1 | rMSPE | **0.6097** | 1.2596 | 1.3865 | 1.3865 | 1.3865 | 0.9200 | 0.8683 | 0.8990 | 0.8226 |
| | $p$-value | | 0.0074 | 0.0022 | 0.0022 | 0.0022 | 0.0126 | 0.0138 | 0.0131 | 0.0170 |
| 2 | rMSPE | **0.8343** | 1.9881 | 2.2026 | 2.2026 | 2.2026 | 1.2717 | 1.1857 | 1.2364 | 1.1121 |
| | $p$-value | | 0.0680 | 0.0262 | 0.0262 | 0.0262 | 0.0454 | 0.0524 | 0.0479 | 0.0600 |
| 3 | rMSPE | **1.0337** | 3.3011 | 3.6663 | 3.6663 | 3.6663 | 1.8674 | 1.6984 | 1.7981 | 1.5472 |
| | $p$-value | | 0.1699 | 0.0916 | 0.0916 | 0.0916 | 0.1057 | 0.1123 | 0.1080 | 0.1209 |
| 4 | rMSPE | **1.0114** | 5.8045 | 6.1726 | 6.1726 | 6.1726 | 2.7507 | 2.4181 | 2.6151 | 2.1134 |
| | $p$-value | | 0.2458 | 0.1807 | 0.1807 | 0.1807 | 0.1814 | 0.1848 | 0.1825 | 0.1860 |

Notes:

(1) Nine methods include TVMA, IC in Gao et al. (2024b), AIC, BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) The out-of-sample forecast period begins from 1984Q2 and ends at 2008Q4.

(3) rMSPE is root mean squared prediction error.

Table 4: Forecast evaluation of different methods during 1984Q2-2015Q4

| $h$ | | TVMA | IC | AIC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GDP | | | | |
| 1 | rMSPE | **0.0067** | 0.0186 | 0.0213 | 0.0212 | 0.0213 | 0.0118 | 0.0109 | 0.0115 | 0.0103 |
| | $p$-value | | 0.1348 | 0.0471 | 0.0484 | 0.0471 | 0.0569 | 0.0576 | 0.0572 | 0.0727 |
| 2 | rMSPE | **0.0069** | 0.0260 | 0.0296 | 0.0292 | 0.0296 | 0.0144 | 0.0129 | 0.0138 | 0.0120 |
| | $p$-value | | 0.2129 | 0.1046 | 0.1138 | 0.1046 | 0.1354 | 0.1433 | 0.1383 | 0.1635 |
| 3 | rMSPE | **0.0078** | 0.0544 | 0.0573 | 0.0560 | 0.0573 | 0.0227 | 0.0198 | 0.0215 | 0.0182 |
| | $p$-value | | 0.2653 | 0.2208 | 0.2423 | 0.2208 | 0.2105 | 0.2133 | 0.2114 | 0.2245 |
| 4 | rMSPE | **0.0095** | 0.1405 | 0.1407 | 0.1406 | 0.1407 | 0.0534 | 0.0456 | 0.0502 | 0.0412 |
| | $p$-value | | 0.2901 | 0.2900 | 0.2909 | 0.2900 | 0.2831 | 0.2829 | 0.2830 | 0.2855 |
| | | | | | | GDPD | | | | |
| 1 | rMSPE | **0.0021** | 0.0045 | 0.0053 | 0.0054 | 0.0053 | 0.0033 | 0.0031 | 0.0032 | 0.0029 |
| | $p$-value | | 0.0035 | 0.0072 | 0.0062 | 0.0072 | 0.0109 | 0.0096 | 0.0104 | 0.0122 |
| 2 | rMSPE | **0.0026** | 0.0051 | 0.0066 | 0.0062 | 0.0066 | 0.0038 | 0.0035 | 0.0037 | 0.0033 |
| | $p$-value | | 0.0000 | 0.0001 | 0.0000 | 0.0001 | 0.0005 | 0.0010 | 0.0006 | 0.0043 |
| 3 | rMSPE | **0.0028** | 0.0087 | 0.0106 | 0.0086 | 0.0106 | 0.0052 | 0.0047 | 0.0050 | 0.0045 |
| | $p$-value | | 0.0187 | 0.0159 | 0.0004 | 0.0159 | 0.0233 | 0.0228 | 0.0231 | 0.0467 |
| 4 | rMSPE | **0.0034** | 0.0154 | 0.0178 | 0.0178 | 0.0178 | 0.0081 | 0.0072 | 0.0077 | 0.0067 |
| | $p$-value | | 0.0794 | 0.0258 | 0.0254 | 0.0258 | 0.0864 | 0.0960 | 0.0901 | 0.1292 |
| | | | | | | FF | | | | |
| 1 | rMSPE | **0.6040** | 1.2457 | 1.4886 | 1.4847 | 1.4886 | 0.9195 | 0.8618 | 0.8959 | 0.8166 |
| | $p$-value | | 0.0012 | 0.0002 | 0.0002 | 0.0002 | 0.0021 | 0.0027 | 0.0023 | 0.0033 |
| 2 | rMSPE | **0.8319** | 1.9793 | 2.2976 | 2.2655 | 2.2976 | 1.2762 | 1.1794 | 1.2364 | 1.1052 |
| | $p$-value | | 0.0263 | 0.0075 | 0.0094 | 0.0075 | 0.0144 | 0.0187 | 0.0158 | 0.0217 |
| 3 | rMSPE | **0.9999** | 3.7415 | 3.9728 | 3.8207 | 3.9728 | 2.0193 | 1.8142 | 1.9355 | 1.6793 |
| | $p$-value | | 0.0418 | 0.0200 | 0.0293 | 0.0200 | 0.0202 | 0.0222 | 0.0209 | 0.0204 |
| 4 | rMSPE | **1.4120** | 7.1808 | 7.2800 | 7.2669 | 7.2800 | 3.3600 | 2.9694 | 3.2002 | 2.7214 |
| | $p$-value | | 0.0720 | 0.0560 | 0.0565 | 0.0560 | 0.0604 | 0.0627 | 0.0612 | 0.0634 |

Notes:

(1) Nine methods include TVMA, IC in Gao et al. (2024b), AIC, BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) The out-of-sample forecast period begins from 1984Q2 and ends at 2015Q4.

(3) rMSPE is root mean squared prediction error.

Table 5: rMSPE of different methods for government spending forecasts

| $h$ | | TVMA | IC | AIC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | begin time: 1985Q1 | | | | |
| 1 | rMSPE | **0.0021** | 0.0023 | 0.0050 | 0.0050 | 0.0050 | 0.0030 | 0.0029 | 0.0029 | 0.0028 |
| | $p$-value | | 0.2038 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 |
| 2 | rMSPE | **0.0035** | 0.0042 | 0.0099 | 0.0099 | 0.0099 | 0.0057 | 0.0056 | 0.0057 | 0.0053 |
| | $p$-value | | 0.0068 | 0.0006 | 0.0006 | 0.0006 | 0.0054 | 0.0061 | 0.0057 | 0.0070 |
| 3 | rMSPE | **0.0047** | 0.0057 | 0.0196 | 0.0196 | 0.0196 | 0.0095 | 0.0092 | 0.0094 | 0.0086 |
| | $p$-value | | 0.0073 | 0.0128 | 0.0128 | 0.0128 | 0.0223 | 0.0231 | 0.0226 | 0.0220 |
| 4 | rMSPE | **0.0058** | 0.0073 | 0.0402 | 0.0402 | 0.0402 | 0.0157 | 0.0150 | 0.0154 | 0.0138 |
| | $p$-value | | 0.0129 | 0.0849 | 0.0849 | 0.0849 | 0.0379 | 0.0380 | 0.0379 | 0.0349 |
| | | | | | | begin time: 1990Q1 | | | | |
| 1 | rMSPE | **0.0019** | 0.0022 | 0.0046 | 0.0046 | 0.0046 | 0.0028 | 0.0027 | 0.0027 | 0.0026 |
| | $p$-value | | 0.0432 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0002 | 0.0002 | 0.0004 |
| 2 | rMSPE | **0.0032** | 0.0039 | 0.0086 | 0.0086 | 0.0086 | 0.0051 | 0.0050 | 0.0051 | 0.0049 |
| | $p$-value | | 0.0112 | 0.0005 | 0.0005 | 0.0005 | 0.0084 | 0.0092 | 0.0087 | 0.0112 |
| 3 | rMSPE | **0.0043** | 0.0054 | 0.0138 | 0.0138 | 0.0138 | 0.0078 | 0.0076 | 0.0077 | 0.0073 |
| | $p$-value | | 0.0177 | 0.0102 | 0.0102 | 0.0102 | 0.0263 | 0.0274 | 0.0268 | 0.0312 |
| 4 | rMSPE | **0.0054** | 0.0067 | 0.0225 | 0.0225 | 0.0225 | 0.0114 | 0.0111 | 0.0113 | 0.0105 |
| | $p$-value | | 0.0291 | 0.0343 | 0.0343 | 0.0343 | 0.0382 | 0.0393 | 0.0387 | 0.0439 |
| | | | | | | begin time: 1995Q1 | | | | |
| 1 | rMSPE | **0.0018** | 0.0020 | 0.0039 | 0.0039 | 0.0039 | 0.0025 | 0.0024 | 0.0025 | 0.0024 |
| | $p$-value | | 0.0906 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0003 | 0.0003 | 0.0004 |
| 2 | rMSPE | **0.0031** | 0.0037 | 0.0076 | 0.0076 | 0.0076 | 0.0045 | 0.0045 | 0.0045 | 0.0044 |
| | $p$-value | | 0.0072 | 0.0019 | 0.0019 | 0.0019 | 0.0169 | 0.0184 | 0.0175 | 0.0205 |
| 3 | rMSPE | **0.0041** | 0.0053 | 0.0131 | 0.0131 | 0.0131 | 0.0071 | 0.0069 | 0.0070 | 0.0067 |
| | $p$-value | | 0.0187 | 0.0305 | 0.0305 | 0.0305 | 0.0335 | 0.0346 | 0.0339 | 0.0359 |
| 4 | rMSPE | **0.0051** | 0.0068 | 0.0226 | 0.0226 | 0.0226 | 0.0106 | 0.0103 | 0.0105 | 0.0100 |
| | $p$-value | | 0.0505 | 0.1132 | 0.1132 | 0.1132 | 0.0508 | 0.0499 | 0.0504 | 0.0488 |
| | | | | | | begin time: 2000Q1 | | | | |
| 1 | rMSPE | **0.0019** | 0.0021 | 0.0034 | 0.0034 | 0.0034 | 0.0024 | 0.0024 | 0.0024 | 0.0024 |
| | $p$-value | | 0.2648 | 0.0000 | 0.0000 | 0.0000 | 0.0015 | 0.0018 | 0.0016 | 0.0023 |
| 2 | rMSPE | **0.0032** | 0.0039 | 0.0071 | 0.0071 | 0.0071 | 0.0045 | 0.0044 | 0.0045 | 0.0044 |
| | $p$-value | | 0.0305 | 0.0073 | 0.0073 | 0.0073 | 0.0266 | 0.0283 | 0.0273 | 0.0305 |
| 3 | rMSPE | **0.0045** | 0.0056 | 0.0118 | 0.0118 | 0.0118 | 0.0068 | 0.0067 | 0.0067 | 0.0065 |
| | $p$-value | | 0.0279 | 0.0330 | 0.0330 | 0.0330 | 0.0340 | 0.0345 | 0.0342 | 0.0351 |
| 4 | rMSPE | **0.0060** | 0.0076 | 0.0186 | 0.0186 | 0.0186 | 0.0097 | 0.0095 | 0.0097 | 0.0093 |
| | $p$-value | | 0.0555 | 0.0630 | 0.0630 | 0.0630 | 0.0436 | 0.0432 | 0.0434 | 0.0429 |
| | | | | | | begin time: 2005Q1 | | | | |
| 1 | rMSPE | **0.0020** | 0.0022 | 0.0032 | 0.0032 | 0.0032 | 0.0024 | 0.0024 | 0.0024 | 0.0024 |
| | $p$-value | | 0.4717 | 0.0006 | 0.0006 | 0.0006 | 0.0077 | 0.0094 | 0.0083 | 0.0116 |
| 2 | rMSPE | **0.0034** | 0.0042 | 0.0069 | 0.0069 | 0.0069 | 0.0045 | 0.0045 | 0.0045 | 0.0044 |
| | $p$-value | | 0.1851 | 0.0499 | 0.0499 | 0.0499 | 0.1231 | 0.1275 | 0.1249 | 0.1325 |
| 3 | rMSPE | **0.0049** | 0.0059 | 0.0113 | 0.0113 | 0.0113 | 0.0064 | 0.0063 | 0.0064 | 0.0062 |
| | $p$-value | | 0.0651 | 0.0954 | 0.0954 | 0.0954 | 0.1087 | 0.1088 | 0.1087 | 0.1099 |
| 4 | rMSPE | **0.0067** | 0.0079 | 0.0172 | 0.0172 | 0.0172 | 0.0090 | 0.0089 | 0.0090 | 0.0087 |
| | $p$-value | | 0.0401 | 0.0781 | 0.0781 | 0.0781 | 0.0797 | 0.0798 | 0.0797 | 0.0800 |

Notes:

(1) Nine methods include TVMA, IC in Gao et al. (2024b), AIC, BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) The out-of-sample forecast period begins from the quarter denoted as “begin time”, and ends at 2015Q4.

(3) rMSPE is root mean squared prediction error.

Table 6: rMSPE of different methods for the real per capita GDP

| $h$ | | TVMA | IC | AIC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | begin time: 1985Q1 | | | | | |
| 1 | rMSPE | **0.0075** | 0.0084 | 0.0176 | 0.0176 | 0.0176 | 0.0100 | 0.0097 | 0.0099 | 0.0094 |
| | $p$-value | | 0.0131 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 |
| 2 | rMSPE | **0.0123** | 0.0151 | 0.0327 | 0.0327 | 0.0327 | 0.0178 | 0.0173 | 0.0176 | 0.0166 |
| | $p$-value | | 0.0034 | 0.0002 | 0.0002 | 0.0002 | 0.0053 | 0.0069 | 0.0059 | 0.0096 |
| 3 | rMSPE | **0.0176** | 0.0227 | 0.0509 | 0.0509 | 0.0509 | 0.0293 | 0.0284 | 0.0289 | 0.0271 |
| | $p$-value | | 0.0108 | 0.0030 | 0.0030 | 0.0030 | 0.0224 | 0.0240 | 0.0230 | 0.0232 |
| 4 | rMSPE | **0.0244** | 0.0328 | 0.0705 | 0.0705 | 0.0705 | 0.0435 | 0.0423 | 0.0430 | 0.0403 |
| | $p$-value | | 0.0418 | 0.0034 | 0.0034 | 0.0034 | 0.0278 | 0.0295 | 0.0285 | 0.0287 |
| | | | | | begin time: 1990Q1 | | | | | |
| 1 | rMSPE | **0.0080** | 0.0083 | 0.0160 | 0.0160 | 0.0160 | 0.0100 | 0.0099 | 0.0100 | 0.0096 |
| | $p$-value | | 0.4324 | 0.0000 | 0.0000 | 0.0000 | 0.0010 | 0.0017 | 0.0012 | 0.0035 |
| 2 | rMSPE | **0.0127** | 0.0150 | 0.0309 | 0.0309 | 0.0309 | 0.0185 | 0.0181 | 0.0183 | 0.0175 |
| | $p$-value | | 0.0218 | 0.0003 | 0.0003 | 0.0003 | 0.0105 | 0.0125 | 0.0112 | 0.0159 |
| 3 | rMSPE | **0.0175** | 0.0221 | 0.0488 | 0.0488 | 0.0488 | 0.0296 | 0.0288 | 0.0293 | 0.0276 |
| | $p$-value | | 0.0204 | 0.0126 | 0.0126 | 0.0126 | 0.0366 | 0.0391 | 0.0376 | 0.0423 |
| 4 | rMSPE | **0.0225** | 0.0292 | 0.0808 | 0.0808 | 0.0808 | 0.0437 | 0.0424 | 0.0432 | 0.0400 |
| | $p$-value | | 0.0302 | 0.0593 | 0.0593 | 0.0593 | 0.0549 | 0.0568 | 0.0556 | 0.0583 |
| | | | | | begin time: 1995Q1 | | | | | |
| 1 | rMSPE | **0.0073** | 0.0077 | 0.0135 | 0.0135 | 0.0135 | 0.0092 | 0.0090 | 0.0091 | 0.0089 |
| | $p$-value | | 0.3131 | 0.0000 | 0.0000 | 0.0000 | 0.0015 | 0.0021 | 0.0017 | 0.0031 |
| 2 | rMSPE | **0.0121** | 0.0140 | 0.0257 | 0.0257 | 0.0257 | 0.0164 | 0.0162 | 0.0163 | 0.0158 |
| | $p$-value | | 0.0795 | 0.0124 | 0.0124 | 0.0124 | 0.0134 | 0.0142 | 0.0137 | 0.0147 |
| 3 | rMSPE | **0.0162** | 0.0204 | 0.0469 | 0.0469 | 0.0469 | 0.0256 | 0.0250 | 0.0254 | 0.0244 |
| | $p$-value | | 0.0546 | 0.0532 | 0.0532 | 0.0532 | 0.0376 | 0.0384 | 0.0379 | 0.0378 |
| 4 | rMSPE | **0.0201** | 0.0255 | 0.0849 | 0.0849 | 0.0849 | 0.0369 | 0.0357 | 0.0364 | 0.0344 |
| | $p$-value | | 0.0839 | 0.1270 | 0.1270 | 0.1270 | 0.0563 | 0.0551 | 0.0558 | 0.0533 |
| | | | | | begin time: 2000Q1 | | | | | |
| 1 | rMSPE | **0.0078** | 0.0085 | 0.0133 | 0.0133 | 0.0133 | 0.0094 | 0.0093 | 0.0094 | 0.0092 |
| | $p$-value | | 0.1241 | 0.0002 | 0.0002 | 0.0002 | 0.0093 | 0.0110 | 0.0099 | 0.0134 |
| 2 | rMSPE | **0.0140** | 0.0165 | 0.0255 | 0.0255 | 0.0255 | 0.0178 | 0.0176 | 0.0177 | 0.0174 |
| | $p$-value | | 0.0270 | 0.0128 | 0.0128 | 0.0128 | 0.0224 | 0.0234 | 0.0228 | 0.0251 |
| 3 | rMSPE | **0.0205** | 0.0259 | 0.0423 | 0.0423 | 0.0423 | 0.0278 | 0.0274 | 0.0276 | 0.0270 |
| | $p$-value | | 0.0224 | 0.0407 | 0.0407 | 0.0407 | 0.0339 | 0.0332 | 0.0336 | 0.0323 |
| 4 | rMSPE | **0.0275** | 0.0349 | 0.0662 | 0.0662 | 0.0662 | 0.0385 | 0.0379 | 0.0383 | 0.0372 |
| | $p$-value | | 0.1093 | 0.0503 | 0.0503 | 0.0503 | 0.0274 | 0.0265 | 0.0271 | 0.0258 |
| | | | | | begin time: 2005Q1 | | | | | |
| 1 | rMSPE | **0.0082** | 0.0087 | 0.0125 | 0.0125 | 0.0125 | 0.0093 | 0.0093 | 0.0093 | 0.0092 |
| | $p$-value | | 0.3357 | 0.0076 | 0.0076 | 0.0076 | 0.1336 | 0.1446 | 0.1380 | 0.1600 |
| 2 | rMSPE | **0.0156** | 0.0186 | 0.0251 | 0.0251 | 0.0251 | 0.0187 | 0.0185 | 0.0186 | 0.0184 |
| | $p$-value | | 0.0659 | 0.0239 | 0.0239 | 0.0239 | 0.1259 | 0.1328 | 0.1287 | 0.1420 |
| 3 | rMSPE | **0.0242** | 0.0309 | 0.0408 | 0.0408 | 0.0408 | 0.0297 | 0.0294 | 0.0296 | 0.0291 |
| | $p$-value | | 0.0225 | 0.0537 | 0.0537 | 0.0537 | 0.0655 | 0.0630 | 0.0645 | 0.0610 |
| 4 | rMSPE | **0.0346** | 0.0445 | 0.0623 | 0.0623 | 0.0623 | 0.0432 | 0.0428 | 0.0430 | 0.0423 |
| | $p$-value | | 0.1255 | 0.0361 | 0.0361 | 0.0361 | 0.0026 | 0.0014 | 0.0021 | 0.0008 |

Notes:

(1) Nine methods include TVMA, IC in Gao et al. (2024b), AIC, BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) The out-of-sample forecast period begins from the quarter denoted as “begin time”, and ends at 2015Q4.

(3) rMSPE is root mean squared prediction error.

Table 7: Forecast evaluation of different methods for government spending forecasts

| $h$ | | TVMA | IC | AIC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | “Great Moderation” 1983Q1-2006Q4 | | | | | |
| 1 | rMSPE | **0.0022** | 0.0024 | 0.0056 | 0.0056 | 0.0056 | 0.0031 | 0.0030 | 0.0031 | 0.0029 |
| | $p$-value | | 0.1494 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0003 |
| 2 | rMSPE | **0.0032** | 0.0042 | 0.0095 | 0.0095 | 0.0095 | 0.0055 | 0.0053 | 0.0054 | 0.0051 |
| | $p$-value | | 0.0326 | 0.0054 | 0.0054 | 0.0054 | 0.0076 | 0.0082 | 0.0079 | 0.0092 |
| 3 | rMSPE | **0.0039** | 0.0051 | 0.0187 | 0.0187 | 0.0187 | 0.0080 | 0.0077 | 0.0079 | 0.0073 |
| | $p$-value | | 0.0744 | 0.0172 | 0.0173 | 0.0172 | 0.0284 | 0.0286 | 0.0285 | 0.0309 |
| 4 | rMSPE | **0.0047** | 0.0064 | 0.0327 | 0.0328 | 0.0327 | 0.0121 | 0.0115 | 0.0118 | 0.0107 |
| | $p$-value | | 0.0882 | 0.0536 | 0.0536 | 0.0536 | 0.0746 | 0.0761 | 0.0752 | 0.0816 |
| | | | | | “Post Great Moderation” 2007Q1-2015Q4 | | | | | |
| 1 | rMSPE | **0.0021** | 0.0023 | 0.0035 | 0.0035 | 0.0035 | 0.0026 | 0.0026 | 0.0026 | 0.0025 |
| | $p$-value | | 0.4569 | 0.0007 | 0.0007 | 0.0007 | 0.0097 | 0.0117 | 0.0105 | 0.0137 |
| 2 | rMSPE | **0.0037** | 0.0045 | 0.0075 | 0.0075 | 0.0075 | 0.0049 | 0.0048 | 0.0049 | 0.0047 |
| | $p$-value | | 0.1983 | 0.0492 | 0.0492 | 0.0492 | 0.1310 | 0.1358 | 0.1329 | 0.1413 |
| 3 | rMSPE | **0.0052** | 0.0064 | 0.0119 | 0.0119 | 0.0119 | 0.0069 | 0.0068 | 0.0068 | 0.0067 |
| | $p$-value | | 0.0890 | 0.0963 | 0.0963 | 0.0963 | 0.1280 | 0.1289 | 0.1284 | 0.1284 |
| 4 | rMSPE | **0.0071** | 0.0085 | 0.0180 | 0.0180 | 0.0180 | 0.0095 | 0.0094 | 0.0095 | 0.0092 |
| | $p$-value | | 0.0455 | 0.0779 | 0.0779 | 0.0779 | 0.0872 | 0.0874 | 0.0873 | 0.0877 |

Notes:

(1) Nine methods include TVMA, IC in Gao et al. (2024b), AIC, BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) For the first out of sample forecast, the estimation sample starts from 1954Q1 and ends at 1982Q4-h+1.

(3) rMSPE is root mean squared prediction error.

Table 8: Forecast evaluation of different methods for the real per capita GDP

| $h$ | | TVMA | IC | AIC | BIC | HQ | sAIC | sBIC | sHQ | SA |
|---|---|---|---|---|---|---|---|---|---|---|
| "Great Moderation" 1983Q1-2006Q4 | | | | | | | | | | |
| 1 | rMSPE | **0.0085** | 0.0097 | 0.0231 | 0.0231 | 0.0231 | 0.0124 | 0.0120 | 0.0123 | 0.0115 |
| | $p$-value | | 0.1913 | 0.0000 | 0.0000 | 0.0000 | 0.0069 | 0.0106 | 0.0082 | 0.0185 |
| 2 | rMSPE | **0.0133** | 0.0185 | 0.0431 | 0.0431 | 0.0431 | 0.0207 | 0.0200 | 0.0204 | 0.0190 |
| | $p$-value | | 0.0196 | 0.0023 | 0.0023 | 0.0023 | 0.0178 | 0.0194 | 0.0185 | 0.0201 |
| 3 | rMSPE | **0.0184** | 0.0264 | 0.0583 | 0.0583 | 0.0583 | 0.0307 | 0.0297 | 0.0303 | 0.0284 |
| | $p$-value | | 0.0548 | 0.0031 | 0.0031 | 0.0031 | 0.0069 | 0.0071 | 0.0070 | 0.0073 |
| 4 | rMSPE | **0.0288** | 0.0340 | 0.0885 | 0.0884 | 0.0885 | 0.0482 | 0.0465 | 0.0475 | 0.0442 |
| | $p$-value | | 0.4001 | 0.0068 | 0.0068 | 0.0068 | 0.0221 | 0.0262 | 0.0236 | 0.0343 |
| "Post Great Moderation" 2007Q1-2015Q4 | | | | | | | | | | |
| 1 | rMSPE | **0.0088** | 0.0094 | 0.0128 | 0.0128 | 0.0128 | 0.0098 | 0.0098 | 0.0098 | 0.0097 |
| | $p$-value | | 0.3433 | 0.0162 | 0.0162 | 0.0162 | 0.2277 | 0.2424 | 0.2336 | 0.2634 |
| 2 | rMSPE | **0.0171** | 0.0204 | 0.0254 | 0.0254 | 0.0254 | 0.0198 | 0.0197 | 0.0198 | 0.0195 |
| | $p$-value | | 0.0613 | 0.0369 | 0.0369 | 0.0369 | 0.2018 | 0.2096 | 0.2050 | 0.2238 |
| 3 | rMSPE | **0.0265** | 0.0340 | 0.0409 | 0.0409 | 0.0409 | 0.0312 | 0.0310 | 0.0312 | 0.0308 |
| | $p$-value | | 0.0231 | 0.0493 | 0.0493 | 0.0493 | 0.1135 | 0.1107 | 0.1125 | 0.1087 |
| 4 | rMSPE | **0.0378** | 0.0491 | 0.0628 | 0.0628 | 0.0628 | 0.0453 | 0.0450 | 0.0452 | 0.0446 |
| | $p$-value | | 0.1197 | 0.0259 | 0.0259 | 0.0259 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |

Notes:

(1) Nine methods include TVMA, IC in Gao et al. (2024b), AIC, BIC, HQ, sAIC, sBIC, sHQ and SA.

(2) For the first out of sample forecast, the estimation sample starts from 1954Q1 and ends at 2006Q4-h+1.

(3) rMSPE is root mean squared prediction error.