Econometric Theory, **42**, 2026, 101–134.
doi:10.1017/S0266466624000288

OPTIMAL MODEL AVERAGING FOR JOINT VALUE-AT-RISK AND EXPECTED SHORTFALL REGRESSION

Shoukun Jiao
University of Science and Technology of China

Wenchao Xu
Shanghai University of International Business and Economics

Wuyi Ye
University of Science and Technology of China

Xinyu Zhang
University of Science and Technology of China and Chinese Academy of Sciences

Since the implementation of the Basel III Accord, expected shortfall (ES) has gained increasing attention from regulators as a complement to value-at-risk (VaR). The problem of elicitability for ES makes jointly modeling VaR and ES a popular method to study ES. In this article, we develop model averaging for joint VaR and ES regression models that selects the two weight vectors by minimizing a jackknife criterion. We show the large sample properties of the estimators under potential model misspecification with increasing dimension of parameters and the asymptotic optimality of the selected weights in the sense of minimizing the out-of-sample excess final prediction error. Simulation studies and three empirical analyses reveal good finite sample performance.

1. INTRODUCTION

Value-at-risk (VaR), a tail quantile of the distribution of a financial return, is a standard measure of financial risk. However, VaR is completely blind to the behavior of the tail beyond the reference quantile. In contrast, expected shortfall (ES) depends on the whole tail beyond VaR, which was widely valued by regulators after the financial crisis in 2008. In addition, VaR is not a subadditive risk measure,

The authors would like to thank the Co-Editor, Viktor Todorov, and three anonymous referees for their guidance and insightful comments which substantially improved the article. S.J. and W.Y. thank research support by the National Natural Science Foundation of China (Grant No. 72371230) and the Anhui Provincial Natural Science Foundation (Grant No. 2208085J41). W.X. thanks research support by the National Natural Science Foundation of China (Grant No. 12101591). X.Z. thanks research support by the National Natural Science Foundation of China (Grant Nos. 71925007, 72091212, and 71988101). Address correspondence to Wuyi Ye, School of Management, University of Science and Technology of China, Hefei, China; e-mail: wyye@ustc.edu.cn.

S.J. and W.X. have contributed equally to this work and should be considered co-first authors.

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

but ES is. This feature means that the VaR for a portfolio may be less than the sum of the VaR of the components of the portfolio, whereas the ES is not lower than the sum of the ES of its components. Ever since the implementation of Basel III, risk managers have increasingly focused on ES. In particular, the "market discipline" aspects of Basel III require banks to regularly disclose ES and VaR. Compared with the large literature on VaR models, ES models are not as rich, because ES is not elicitable (Weber, 2006; Gneiting, 2011). That is, we cannot find a loss function such that ES is the solution that minimizes the expectation of the loss function. This makes it infeasible to construct an ES regression such as quantile regression, which can intuitively investigate the relationship between certain covariate variables and responses (VaR or ES). In a seminal article, Fissler and Ziegel (2016) explored the fact that ES is jointly elicitable with VaR under a set of suitable scoring functions and provided a family of scoring functions that can be minimized by the true VaR and ES. This work inspired a number of semiparametric joint models for VaR and ES. For example, Patton et al. (2019) discussed some dynamic semiparametric models for VaR and ES. Dimitriadis and Bayer (2019) attenuated the assumptions of Patton et al. (2019) and provided the regression framework. Taylor (2019) proposed a joint estimation based on the asymmetric Laplace distribution. With similar advantages as quantile regression, the joint VaR and ES regression model can be regarded as an extension of quantile regression, and has received increasing attention by researchers, especially as ES has gradually become a standard risk measure complementing the VaR measure. In many empirical analyses, to investigate specific VaR and ES, there may be many plausible candidate models. The single best model found by model selection is not stable because it is not a continuous process in terms of the data. That is, even if we slightly perturb one observation in the data set, the model selection method may give us a different model from the original chosen one. By compromising across the candidate models, model averaging provides a kind of insurance against selecting a poor model, thus reducing the risk of model misspecification and the instability of model selection. However, to the best of our knowledge, there are still few studies on joint VaR and ES regression models, especially on model averaging.

In econometrics and statistics, model selection and model averaging are two popular approaches when many candidate models exist. The aim of model selection is to choose the single best model from a list of candidate models. As an alternative to model selection, model averaging seeks to obtain a combined estimator or prediction by taking the weighted average of the estimators or predictions obtained from all candidate models. Thus model averaging is a smoothed extension of model selection and can potentially reduce risk relative to model selection (Yuan and Yang, 2005; Magnus et al., 2010; Peng and Yang, 2022). This article focuses on model averaging of the joint VaR and ES regression model.

Model averaging can be classified as Bayesian model averaging (BMA) and frequentist model averaging (FMA). BMA has been popular for a long time; see Hoeting et al. (1999) and Clyde and George (2004) for reviews. In the past two decades, FMA has also been actively developed. Buckland et al. (1997)

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

constructed model averaging weights based on the Akaike information criterion (AIC) or Bayesian information criterion (BIC). Hansen (2007) proposed selecting the weights by minimizing the Mallows criterion. Hansen and Racine (2012) proposed jackknife model averaging (JMA) for least squares regression with heteroscedastic errors which selects weights by minimizing a leave-one-out cross-validation criterion. Recently, JMA was further applied to dependent data (Zhang et al., 2013), quantile regression (Lu and Su, 2015; Wang et al., 2023), and expectile regression (Tu and Wang, 2020; Bai et al., 2022).

In this article, we propose model averaging for the joint VaR and ES regression framework of Dimitriadis and Bayer (2019) and Patton et al. (2019), where the model averaging weight vectors for VaR and ES are allowed to differ. The contributions of the article are as follows. First, we prove the consistency and asymptotic normality of parameter estimators for a single potentially misspecified joint model with a diverging number of parameters. To the best of our knowledge, there is no study on the joint VaR and ES model under potential model misspecification. Then, we derive the uniform consistency of the leave-one-out parameter estimators. Second, after selecting model averaging weights using the JMA approach, we prove asymptotic optimality with two model averaging weight vectors (which is new in literature) in the sense of minimizing the out-of-sample excess final prediction error (EFPE).

The remainder of the article is organized as follows: Section 2 introduces the joint VaR and ES regression model and its model averaging estimator. Section 3 presents the asymptotic properties of the joint VaR and ES model estimators and the asymptotic optimality of our jackknife-selected weight vector. Section 4 conducts Monte Carlo experiments to examine finite-sample performance of our JMA estimators. Section 5 investigates three empirical applications and illustrates the performance of the proposed model. Section 6 concludes the article. Proofs of technical results are given in the Appendix and Supplementary Material.

Notation. For an $m \times n$ real matrix $\mathbf{A}$, we denote its transpose as $\mathbf{A}^\top$ and its Frobenius norm as $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^\top)}$. When $\mathbf{A}$ is symmetric, we use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to denote its smallest and largest eigenvalues, respectively. All limiting processes are with respect to $n \to \infty$, unless stated otherwise. The notation $\xrightarrow{d}$ denotes convergence in distribution, and a.s. means almost surely.

2. MODEL AVERAGING ESTIMATOR

2.1. Joint VaR and ES Regression Model Averaging

Let $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ be a random sample, where y_i is a scalar dependent variable and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots)^\top$ is a countably infinite-dimensional covariate. Let $F_{y|\mathbf{x}}(\cdot|\mathbf{x}_i)$ denote the conditional distribution function of y_i given $\mathbf{x}_i$. For a fixed $\tau \in (0,1)$, the τth conditional VaR and ES of y_i given $\mathbf{x}_i$ are defined as $\text{VaR}_\tau(y_i|\mathbf{x}_i) = \inf\{y \in \mathbb{R} : F_{y|\mathbf{x}}(y|\mathbf{x}_i) \geq \tau\}$ and $\text{ES}_\tau(y_i|\mathbf{x}_i) = \mathbb{E}\{y_i | y_i \leq \text{VaR}_\tau(y_i|\mathbf{x}_i), \mathbf{x}_i\}$, respectively. As in Koenker and Bassett (1982) and Lu and Su (2015), we consider the

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

following data generating process (DGP):

$$y_i = \sum_{j=1}^{\infty} \beta_j x_{ij} + \left(\sum_{j=1}^{\infty} \alpha_j x_{ij} \right) \eta_i, \tag{2.1}$$

where β_j and α_j are unknown parameters with $\alpha_1 = 1$ and η_i is an independent and identically distributed (i.i.d.) unobservable disturbance and independent of $\mathbf{x}_i$. Without loss of generality, we assume that $x_{i1} = 1$. Under the condition that $\sum_{j=1}^{\infty} \alpha_j x_{ij} > 0$ a.s., the τth conditional VaR and ES of y_i given $\mathbf{x}_i$ are given by

$$\begin{cases} \mathrm{VaR}_\tau(y_i|\mathbf{x}_i) = \sum_{j=1}^{\infty} \{\beta_j + \alpha_j q_\tau(\eta_i)\} x_{ij} \equiv \sum_{j=1}^{\infty} \vartheta_j(\tau) x_{ij}, \\ \mathrm{ES}_\tau(y_i|\mathbf{x}_i) = \sum_{j=1}^{\infty} \{\beta_j + \alpha_j \mathrm{ES}_\tau(\eta_i)\} x_{ij} \equiv \sum_{j=1}^{\infty} \theta_j(\tau) x_{ij}, \end{cases} \tag{2.2}$$

where $q_\tau(\eta_i)$ and $\mathrm{ES}_\tau(\eta_i)$ denote the τth quantile and ES of η_i, respectively, $\vartheta_j(\tau) = \beta_j + \alpha_j q_\tau(\eta_i)$ and $\theta_j(\tau) = \beta_j + \alpha_j \mathrm{ES}_\tau(\eta_i)$. It follows that we have the following joint VaR and ES regression model:

$$\begin{cases} y_i = \sum_{j=1}^{\infty} \vartheta_j(\tau) x_{ij} + \varepsilon_i, \\ y_i = \sum_{j=1}^{\infty} \theta_j(\tau) x_{ij} + \epsilon_i, \end{cases} \tag{2.3}$$

where $\mathrm{VaR}_\tau(\varepsilon_i|\mathbf{x}_i) = 0$ and $\mathrm{ES}_\tau(\epsilon_i|\mathbf{x}_i) = 0$. For the sake of simplicity, we suppress the dependence of $\vartheta_j(\tau)$ and $\theta_j(\tau)$ on τ and thus write $\vartheta_j = \vartheta_j(\tau)$ and $\theta_j = \theta_j(\tau)$.

The goal is to estimate the unknown $\mathrm{VaR}_\tau(y_i|\mathbf{x}_i)$ and $\mathrm{ES}_\tau(y_i|\mathbf{x}_i)$. In many empirical analyses, since some components of $\mathbf{x}_i$ do not contribute significantly to y_i, especially when $\mathbf{x}_i$ is high-dimensional, model selection or model averaging is necessary. Hence, we consider a sequence of approximating models $m = 1, \ldots, M$, where the mth model uses k_m regressors belonging to $\mathbf{x}_i$ and M is allowed to grow to infinity with the sample size. Specifically, we write the mth approximating model as follows:

$$\begin{cases} y_i = \boldsymbol{x}_{i,m}^\top \boldsymbol{\vartheta}_m + \varepsilon_{i,m} = \sum_{j=1}^{k_m} \vartheta_{j,m} x_{ij,m} + \varepsilon_{i,m}, \\ y_i = \boldsymbol{x}_{i,m}^\top \boldsymbol{\theta}_m + \epsilon_{i,m} = \sum_{j=1}^{k_m} \theta_{j,m} x_{ij,m} + \epsilon_{i,m}, \end{cases} \tag{2.4}$$

where $\mathbf{x}_{i,m} = (x_{i1,m}, \ldots, x_{ik_m,m})^\top$ is a k_m-dimensional vector, $\boldsymbol{\vartheta}_m = (\vartheta_{1,m}, \ldots, \vartheta_{k_m,m})^\top$ and $\boldsymbol{\theta}_m = (\theta_{1,m}, \ldots, \theta_{k_m,m})^\top$ are the corresponding coefficients. Moreover, the error terms $\varepsilon_{i,m}$ and $\epsilon_{i,m}$ both consist of a random error term (ε_i/ϵ_i) and an approximation error term ($b_{i,m}$ /$d_{i,m}$), namely $\varepsilon_{i,m} = \varepsilon_i + b_{i,m}$ and $\epsilon_{i,m} = \epsilon_i + d_{i,m}$. The approximation error terms are defined as $b_{i,m} = \mathrm{VaR}_\tau(y_i|\mathbf{x}_i) - \sum_{j=1}^{k_m} \vartheta_{j,m} x_{ij,m}$ and $d_{i,m} = \mathrm{ES}_\tau(y_i|\mathbf{x}_i) - \sum_{j=1}^{k_m} \theta_{j,m} x_{ij,m}$. For notational simplicity, we set $\Theta_m = (\boldsymbol{\vartheta}_m^\top, \boldsymbol{\theta}_m^\top)^\top$, $v_{i,m}(\Theta_m) = \mathbf{x}_{i,m}^\top \boldsymbol{\vartheta}_m$, and $e_{i,m}(\Theta_m) = \mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m$. As discussed by Lu and Su (2015), k_m is permitted to diverge to infinity with n both in practice and in theory.

When $b_{i,m} = d_{i,m} \equiv 0$ (i.e., the model is correctly specified) and k_m is fixed, the joint model (2.4) has been studied by Patton et al. (2019) and Dimitriadis and Bayer

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

(2019). Following Patton et al. (2019), we define the loss function as follows:

$$L(y,v,e) = -\frac{1}{\tau e}\mathbb{I}\{y \le v\}(v-y) + \frac{v}{e} + \log(-e) - 1, \tag{2.5}$$

where $\mathbb{I}(\cdot)$ is an indicator function. To ensure that the loss function makes sense, we assume that $e < 0$. Note that, $\partial \mathbb{E}\{L(y,v,e)\}/\partial v = 0$ and $\partial \mathbb{E}\{L(y,v,e)\}/\partial e = 0$ if and only if $\tau - F(v) = 0$ and $e - \tau^{-1}\int_{-\infty}^{v} y dF(y) = 0$. Therefore, $(\mathrm{VaR}_\tau(y), \mathrm{ES}_\tau(y)) = \mathrm{argmin}_{(v,e)}\mathbb{E}\{L(y,v,e)\}$ under some trivial conditions, which indicates that the loss function (2.5) is reasonable.

Remark 1. Fissler and Ziegel (2016) proposed a family of suitable loss functions for jointly estimating VaR and ES as follows:

$$L_{\mathrm{FZ}}(y,v,e;a,G_1,G_2) = [\mathbb{I}\{y \le v\} - \tau]\left\{G_1(v) - G_1(y) + \frac{1}{\tau}G_2(e)v\right\} - G_2(e)\left(\frac{1}{\tau}\mathbb{I}\{y \le v\}y - e\right) - \mathcal{G}_2(e) + a(y), \tag{2.6}$$

where G_1, a, G_2, and $\mathcal{G}_2$ are functions satisfying that $\mathcal{G}_2' = G_2$, G_1 is increasing, and $\mathcal{G}_2$ is increasing and convex. By choosing $G_1(x) = 0$, $G_2(x) = -1/x$, and $a(x) = 0$, (2.6) reduces to (2.5), which is by far the most commonly used loss function and has certain properties such as being homogeneous of degree zero; see Patton et al. (2019) for detailed discussions. For simplicity, this article focuses on the special loss function (2.5). If we choose other loss functions (2.6), the results in this article need to be restudied. We leave this for future work.

Then, the unknown parameter Θ_m in the mth approximating model is estimated by[1]

$$\widehat{\Theta}_m = \left(\widehat{\boldsymbol{\vartheta}}_m^\top, \widehat{\boldsymbol{\theta}}_m^\top\right)^\top = \arg\min_{\Theta_m} \sum_{i=1}^{n} L(y_i, v_{i,m}(\Theta_m), e_{i,m}(\Theta_m)).$$

As a result, under the mth model, the estimators of $\mathrm{VaR}_\tau(y_i|\mathbf{x}_i)$ and $\mathrm{ES}_\tau(y_i|\mathbf{x}_i)$ are $\mathbf{x}_{i,m}^\top\widehat{\boldsymbol{\vartheta}}_m$ and $\mathbf{x}_{i,m}^\top\widehat{\boldsymbol{\theta}}_m$, respectively. Let $\boldsymbol{\varpi} = (\varpi_1, \ldots, \varpi_M)^\top$ and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_M)^\top$ be weight vectors belonging to

$$\mathcal{H} = \left\{(\boldsymbol{\varpi}, \boldsymbol{\omega}) \colon \boldsymbol{\varpi}, \boldsymbol{\omega} \in [0,1]^M, \sum_{m=1}^{M}\varpi_m = 1, \sum_{m=1}^{M}\omega_m = 1\right\}.$$

Then, the model averaging estimators of $\mathrm{VaR}_\tau(y_i|\mathbf{x}_i)$ and $\mathrm{ES}_\tau(y_i|\mathbf{x}_i)$ with weights $(\boldsymbol{\varpi}, \boldsymbol{\omega})$ are, respectively,

[1] The numerical search algorithm needs to be chosen carefully due to the presence of the indicator function $\mathbb{I}\{y \le v\}$ in the loss function. In this article, we follow the algorithm in Appendix C of Patton et al. (2019) and use the relevant MATLAB code provided by Patton et al. (2019).

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

$$\widehat{v}_i(\boldsymbol{\varpi}) = \sum_{m=1}^{M} \varpi_m \mathbf{x}_{i,m}^{\top} \widehat{\boldsymbol{\vartheta}}_m \quad \text{and} \quad \widehat{e}_i(\boldsymbol{\omega}) = \sum_{m=1}^{M} w_m \mathbf{x}_{i,m}^{\top} \widehat{\boldsymbol{\theta}}_m.$$

It is worth emphasizing that, unlike previous literature, we use two different weight vectors $\boldsymbol{\varpi}$ and $\boldsymbol{\omega}$ corresponding to VaR and ES.

The most critical step of model averaging is to select weights $(\boldsymbol{\varpi}, \boldsymbol{\omega})$. As discussed by Lu and Su (2015), the Mallows-type criterion of Hansen (2007) cannot be applied to the quantile regression model averaging since we do not have the usual bias–variance decomposition for the Mallows-type evaluation criterion. Therefore, the Mallows model averaging cannot be extended to the joint VaR and ES model (2.3), either. On the other hand, since it is difficult to define AIC and BIC for model (2.3), we could not obtain the corresponding model averaging method as in Lu and Su (2015) or Hansen (2007). For these reasons, below we follow JMA of Hansen and Racine (2012) and Lu and Su (2015) and propose a jackknife selection of $(\boldsymbol{\varpi}, \boldsymbol{\omega})$.

2.2. Weight Choice Criterion

Following Hansen and Racine (2012), Zhang et al. (2013), and Lu and Su (2015), we propose to use leave-one-out cross-validation to choose weights $(\boldsymbol{\varpi}, \boldsymbol{\omega})$. For $m = 1, \ldots, M$, let $\widehat{\Theta}_{-i,m} = (\widehat{\boldsymbol{\vartheta}}_{-i,m}^{\top}, \widehat{\boldsymbol{\theta}}_{-i,m}^{\top})^{\top}$ denote the jackknife estimator of Θ_m in the mth candidate model with the ith observation deleted. We define a cross-validation criterion function as follows:

$$\mathrm{CV}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^{n} L\left(y_i, \sum_{m=1}^{M} \varpi_m \mathbf{x}_{i,m}^{\top} \widehat{\boldsymbol{\vartheta}}_{-i,m}, \sum_{m=1}^{M} w_m \mathbf{x}_{i,m}^{\top} \widehat{\boldsymbol{\theta}}_{-i,m} \right). \tag{2.7}$$

Then, the jackknife choice of the weight vector $(\widehat{\boldsymbol{\varpi}}, \widehat{\boldsymbol{\omega}})$ is obtained by choosing $(\boldsymbol{\varpi}, \boldsymbol{\omega}) \in \mathcal{H}$ to minimize the above criterion function, that is,

$$(\widehat{\boldsymbol{\varpi}}, \widehat{\boldsymbol{\omega}}) = \arg\min_{(\boldsymbol{\varpi}, \boldsymbol{\omega}) \in \mathcal{H}} \mathrm{CV}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}).$$

Given $(\widehat{\boldsymbol{\varpi}}, \widehat{\boldsymbol{\omega}})$, we can obtain the JMA estimators of $\mathrm{VaR}_\tau(y_i|\mathbf{x}_i)$ and $\mathrm{ES}_\tau(y_i|\mathbf{x}_i)$ by

$$\widehat{v}_i(\widehat{\boldsymbol{\varpi}}) = \sum_{m=1}^{M} \widehat{\varpi}_m \mathbf{x}_{i,m}^{\top} \widehat{\boldsymbol{\vartheta}}_m \quad \text{and} \quad \widehat{e}_i(\widehat{\boldsymbol{\omega}}) = \sum_{m=1}^{M} \widehat{\omega}_m \mathbf{x}_{i,m}^{\top} \widehat{\boldsymbol{\theta}}_m, \tag{2.8}$$

respectively, where $\widehat{\boldsymbol{\varpi}} = (\widehat{\varpi}_1, \ldots, \widehat{\varpi}_M)^{\top}$ and $\widehat{\boldsymbol{\omega}} = (\widehat{\omega}_1, \ldots, \widehat{\omega}_M)^{\top}$.

3. MAIN RESULTS

In this section, we first investigate the asymptotic properties of $\widehat{\Theta}_m$ and $\widehat{\Theta}_{-i,m}$, and then establish the asymptotic optimality of jackknife weights $(\widehat{\boldsymbol{\varpi}}, \widehat{\boldsymbol{\omega}})$.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

3.1. Asymptotic Properties of $\widehat{\Theta}_m$ and $\widehat{\Theta}_{-i,m}$

Although there are some asymptotic results on the joint VaR and ES model in the literature (see, e.g., Dimitriadis and Bayer, 2019; Patton et al., 2019), most of them postulate a correctly specified parametric model and a fixed number of parameters. In practical economics problems, however, it is unrealistic to accurately identify all the covariates associated with ES, which leads to the possibility that all candidate models may be misspecified. In addition, due to the context of certain research questions or to avoid the omission of significant covariates, we need to consider high-dimensional covariates, that is, we allow diverging numbers of parameters in some or all candidate models. This means that the existing asymptotic results cannot be used. Therefore, in this section, we study the asymptotic properties of $\widehat{\Theta}_m$ and $\widehat{\Theta}_{-i,m}$ when the underlying model is potentially misspecified and the number of parameters in the model is permitted to diverge to infinity with the sample size n.

Let $f_{y|\mathbf{x}}(\cdot|\mathbf{x}_i)$ be the probability density function given y_i given $\mathbf{x}_i$. Define the pseudo-true parameter for the mth model as

$$\Theta_m^* = \left(\boldsymbol{\vartheta}_m^{*\top}, \boldsymbol{\theta}_m^{*\top}\right)^\top \equiv \arg\min_{\Theta_m} \mathbb{E}\{L(y_i, v_{i,m}(\Theta_m), e_{i,m}(\Theta_m))\}.$$

In this article, we always assume that, for each $m = 1, \ldots, M$, Θ_m^* exists and is unique, and the corresponding parameter space is compact. To investigate the consistency and asymptotic normality of $\widehat{\Theta}_m$, we define

$$\boldsymbol{g}_{i,m}(\Theta_m) = \frac{\left(\boldsymbol{x}_{i,m}; \mathbf{0}_{k_m \times 1}\right)^\top}{-e_{i,m}(\Theta_m)} \left[\frac{1}{\tau}\mathbb{I}\left\{y_i \le v_{i,m}(\Theta_m)\right\} - 1\right]$$
$$+ \frac{\left(\mathbf{0}_{k_m \times 1}; \boldsymbol{x}_{i,m}\right)^\top}{e_{i,m}(\Theta_m)^2} \left[\frac{1}{\tau}\mathbb{I}\left\{y_i \le v_{i,m}(\Theta_m)\right\}\left\{v_{i,m}(\Theta_m) - y_i\right\} - v_{i,m}(\Theta_m) + e_{i,m}(\Theta_m)\right]$$

and $$\Lambda_m(\Theta_0) = \left.\frac{\partial \mathbb{E}\{\mathbf{g}_{i,m}(\Theta_m)\}}{\partial \Theta_m}\right|_{\Theta_m = \Theta_0}$$

for all m and i, which are consistent with Patton et al. (2019). Let $\mathbf{D}_m = \Lambda_m(\Theta_m^*)$, $\mathbf{A}_m = \mathbb{E}\{\mathbf{g}_{i,m}(\Theta_m^*)\mathbf{g}_{i,m}(\Theta_m^*)^\top\}$, $\Sigma_m = \mathbf{D}_m^{-1}\mathbf{A}_m\mathbf{D}_m^{-1}$, and $\bar{k} = \max_{1 \le m \le M} k_m$.

The following assumption is needed to establish the asymptotic properties.

Assumption 3.1. (i) $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ are i.i.d. such that model (2.1) holds, and $\mathbb{E}|y_i| < \infty$.

(ii) $F_{y|\mathbf{x}}(\cdot|\mathbf{x}_i)$ has finite second moments and $f_{y|\mathbf{x}}(\cdot|\mathbf{x}_i)$ is uniformly bounded by a finite constant c_f and continuous over its support a.s.

(iii) $\mathrm{VaR}_\tau(\varepsilon_i|\mathbf{x}_i) = 0$ and $\mathrm{ES}_\tau(\epsilon_i|\mathbf{x}_i) = 0$ a.s.

(iv) For each $m = 1, \ldots, M$, there exists a neighborhood of Θ_m^*, $\mathcal{N}_m\left(\Theta_m^*\right)$, such that for all $i = 0, 1, \ldots, n$, and $\Theta \in \mathcal{N}_m(\Theta_m^*)$, $0 < |1/e_{i,m}(\Theta)| < K_2 < \infty$ and $e_{i,m}(\Theta_m^*) < v_{i,m}(\Theta_m^*) < 0$ hold uniformly, where K_2 is a constant that does not depend on n and k_m.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

(v) There exist constants $\underline{c}_{D(m)}$, $\bar{c}_{D(m)}$, $\underline{c}_{X(m)}$, and $\bar{c}_{X(m)}$ that may depend on k_m such that $0 < \underline{c}_{D(m)} \leq \lambda_{\min}(\mathbf{D}_m) \leq \lambda_{\max}(\mathbf{D}_m) \leq \bar{c}_{D(m)} < \infty$ and $0 < \underline{c}_{X(m)} \leq \lambda_{\min}\{\mathbb{E}(\mathbf{x}_{i,m}\mathbf{x}_{i,m}^\top)\} \leq \lambda_{\max}\{\mathbb{E}(\mathbf{x}_{i,m}\mathbf{x}_{i,m}^\top)\} \leq \bar{c}_{X(m)} < \infty$. Furthermore, $\bar{c}_{X(m)}/k_m = O(\underline{c}_{D(m)}^2)$.

Assumption 3.1(i) specifies the data are i.i.d. Assumption 3.1(ii) is similar to those in Dimitriadis and Bayer (2019), which imposes weak restrictions on $F_{y|\mathbf{x}}(\cdot|\mathbf{x})$ and $f_{y|\mathbf{x}}(\cdot|\mathbf{x})$. Assumption 3.1(iii) ensures that y_i has a unique conditional ES and VaR at level τ. Assumption 3.1(iv) corresponds to Assumptions 2(A) and 2(C)(i) of Patton et al. (2019) and guarantees that the loss function makes sense. Assumption 3.1(iv) is unlikely to be a binding constraint when we subtract y_i from its mean and only discuss the case where $\tau < 0.5$. Similar to Assumption A.2(ii) in Lu and Su (2015), Assumption 3.1(v) places restrictions on the range of eigenvalues of $\mathbf{D}_m$ and $\mathbb{E}(\mathbf{x}_{i,m}\mathbf{x}_{i,m}^\top)$, and allows these restrictions to depend on k_m. When $k_m \to \infty$, we allow $\bar{c}_{D(m)}$ and $\bar{c}_{X(m)}$ to diverge to infinity and $\underline{c}_{D(m)}$ and $\underline{c}_{X(m)}$ to converge to 0. Furthermore, $\bar{c}_{X(m)}/k_m = O(\underline{c}_{D(m)}^2)$ restricts the rate of convergence of $\bar{c}_{X(m)}/k_m$.

THEOREM 3.1. *Under Assumption 3.1, we have*

$$\left\|\widehat{\Theta}_m - \Theta_m^*\right\| = O_p\left(\sqrt{k_m/n}\right).$$

Theorem 3.1 establishes the rate of convergence of $\widehat{\Theta}_m$, which is a standard result and similar to that obtained by Lu and Su (2015) for quantile regression and Tu and Wang (2020) and Bai et al. (2022) for expectile regression.

Next, we establish the asymptotic distribution of $\widehat{\Theta}_m$, which is of interest per se. We make the following assumption.

Assumption 3.2. (i) There exist constants $\underline{c}_{A(m)}$ and $\bar{c}_{A(m)}$ that may depend on k_m such that $0 < \underline{c}_{A(m)} \leq \lambda_{\min}(\mathbf{A}_m) \leq \lambda_{\max}(\mathbf{A}_m) \leq \bar{c}_{A(m)} < \infty$.

(ii) $\sup_{j\geq 1}\mathbb{E}(x_{ij}^8) \leq c_{\mathbf{x}}$ and $\mathbb{E}(y_i^8) \leq c_{\mathbf{x}}$, where $c_{\mathbf{x}}$ is a strictly positive constant that does not depend on n and k_m.

(iii) $\bar{k}^5/(n\underline{c}_A^2) \to 0$, $\bar{k}^5/(n\underline{c}_A) \to 0$ and $\bar{k}^4/n^3 \to 0$ where $\underline{c}_A = \min_{1\leq m\leq M}\underline{c}_{A(m)}$.

Assumption 3.2(i) is similar to Assumption 3.1(v), and we also allow $\bar{c}_{A(m)}$ to diverge to infinity and $\underline{c}_{A(m)}$ to converge to 0 as $k_m \to \infty$. The first requirement in Assumption 3.2(ii) can be found in Lu and Su (2015), which implies that $\mathbb{E}\|\mathbf{x}_{i,m}\|^4 \leq c_{\mathbf{x}}k_m^2$ and $\mathbb{E}\|\mathbf{x}_{i,m}\|^8 \leq c_{\mathbf{x}}k_m^4$. The rates of convergence of n, $\bar{k}$, and $\underline{c}_{A(m)}$ are restricted by Assumption 3.2(iii). Assumption 3.2(iii) implies $\bar{k}^3\log n/(\underline{c}_{A(m)}n^{5/2}) \to 0$, which will be used in the proof.

THEOREM 3.2. *Let $\mathbf{C}_m$ denote an $l_m \times 2k_m$ matrix such that $\mathbf{C}_0 \equiv \lim_{n\to\infty}\mathbf{C}_m\mathbf{C}_m^\top$ exists and is positive definite, where $l_m \in [1, 2k_m]$ is a fixed integer. Then, under Assumptions 3.1 and 3.2, we have*

$$\sqrt{n}\mathbf{C}_m\Sigma_m^{-1/2}\left(\widehat{\Theta}_m - \Theta_m^*\right) \xrightarrow{d} N(0, \mathbf{C}_0).$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Theorem 3.2 establishes the asymptotic normality for any arbitrary linear combinations of elements of the joint VaR and ES regression estimator when the model can be misspecified and the number of parameters diverges to infinity. In the special case that the joint model is correctly specified and the number of parameters is fixed, the result of Theorem 3.2 reduces to Theorem 2 in Patton et al. (2019) or Theorem 2.7 in Dimitriadis and Bayer (2019). For the correctly specified model, $\mathrm{ES}_\tau(y_i|\mathbf{x}_i) = \tau^{-1}\int_0^\tau \mathrm{VaR}_q(y_i|\mathbf{x}_i)\,dq$, which is not satisfied for the conditional VaR and ES estimators proposed in this article. In addition, $F_{y|\mathbf{x}}(v_{i,m}(\widehat{\Theta}_m)|\mathbf{x}_i) = \tau$ does not hold for all i and m. For these reasons, the theorems of Patton et al. (2019) and Dimitriadis and Bayer (2019) are not applicable here. For fixed k_m, the asymptotic normality of the parameters with a misspecified model can be derived from the sketch of the proof in Patton et al. (2019) which builds on Lemma A.1 of Weiss (1991). However, the relevant theorems for Patton et al. (2019) and Weiss (1991) cannot hold when $k_m \to \infty$.

Next, we study the uniform convergence property of $\widehat{\Theta}_m$ and $\widehat{\Theta}_{-i,m}$. We make the following assumption.

Assumption 3.3. Let $\underline{c}_D = \min_{1\le m\le M}\underline{c}_{D(m)}$, $\bar{c}_D = \max_{1\le m\le M}\bar{c}_{D(m)}$, and $\bar{c}_A = \max_{1\le m\le M}\bar{c}_{A(m)}$.

(i) $\bar{c}_A\bar{c}_D\bar{k}/(n\underline{c}_D^3) \to 0$.

(ii) $nMn^{-0.5L\bar{k}\underline{c}_D^3/(\bar{c}_A\bar{c}_D)} = o(1)$ for a sufficiently large constant L.

Assumption 3.3 is similar to Assumption A.3 of Lu and Su (2015), which imposes restrictions on the largest dimension of the models ($\bar{k}$), the potential number of the candidate models (M), and the constants $\bar{c}_A$, $\bar{c}_D$, and $\underline{c}_D$. Assumption 3.3(ii) suggests that M is allowed to grow at a polynomial rate with n.

THEOREM 3.3. *Under Assumptions 3.1 and 3.3, we have:*

(i) $\max_{1\le m\le M}\left\|\widehat{\Theta}_m - \Theta_m^*\right\| = O_p\left(\sqrt{n^{-1}\bar{k}\log n}\right)$.

(ii) $\max_{1\le i\le n}\max_{1\le m\le M}\left\|\widehat{\Theta}_{-i,m} - \Theta_m^*\right\| = O_p\left(\sqrt{n^{-1}\bar{k}\log n}\right)$.

Theorem 3.3 shows that the rate of uniform convergence of $\widehat{\Theta}_m$ and $\widehat{\Theta}_{-i,m}$ to Θ_m^* depends on the sample size n and the largest dimension of regressors $\bar{k}$ in all candidate models. The same uniform convergence rate is established by Lu and Su (2015) and Tu and Wang (2020) for quantile regressions and expectile regressions, respectively.

3.2. Asymptotic Optimality of Jackknife Weights

Let $\mathcal{D}_n = \{(y_i, \mathbf{x}_i)\}_{i=1}^n$, $(y_0, \mathbf{x}_0)$ be an independent copy of $\mathcal{D}_n$, $\mathbf{x}_{0,m} = (x_{01,m}, \ldots, x_{0k_m,m})^\top$ correspond to the k_m regressors in the mth model and $x_{0j,m}, j = 1, \ldots, k_m$, denote the variables in $\mathbf{x}_0$. For $i = 0, 1, \ldots, n$, let (v_i, e_i) denote the true VaR and ES,

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

that is, $(v_i, e_i) \equiv (\mathrm{VaR}_\tau(y_i|\mathbf{x}_i), \mathrm{ES}_\tau(y_i|\mathbf{x}_i))$. We use the EFPE to evaluate prediction performance:

$$\mathrm{EFPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) = \mathrm{FPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) - \mathbb{E}\{L(y_0, v_0, e_0)\}, \tag{3.1}$$

where

$$\mathrm{FPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) = \mathbb{E}\left\{ L\left(y_0, \sum_{m=1}^{M} \varpi_m \mathbf{x}_{0,m}^\top \widehat{\boldsymbol{\vartheta}}_m, \sum_{m=1}^{M} \omega_m \mathbf{x}_{0,m}^\top \widehat{\boldsymbol{\theta}}_m \right) \middle| \mathcal{D}_n \right\}.$$

Lu and Su (2015) and Tu and Wang (2020) used similar FPEs in model averaging studies of quantile regression and expectile regression, respectively. Bai et al. (2022) constructed the asymptotic optimality of model averaging estimator within the expectile regression framework, in the sense of minimizing the corresponding EFPE. Given the properties of the loss function, it is evident that $\mathrm{EFPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) \geq 0$; please refer to Theorem 5.2 in Fissler and Ziegel (2016) and its proof for details.

In this subsection, we establish the asymptotic optimality of $(\widehat{\boldsymbol{\varpi}}, \widehat{\boldsymbol{\omega}})$ in the sense of achieving the lowest $\mathrm{EFPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega})$. Define

$$\mathrm{EFPE}_n^*(\boldsymbol{\varpi}, \boldsymbol{\omega}) = \mathbb{E}\left\{ L\left(y_0, \sum_{m=1}^{M} \varpi_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\vartheta}_m^*, \sum_{m=1}^{M} \omega_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\theta}_m^* \right) \right\} - \mathbb{E}\{L(y_0, v_0, e_0)\}.$$

Let $\gamma_n = n \inf_{(\boldsymbol{\varpi}, \boldsymbol{\omega}) \in \mathcal{H}} \mathrm{EFPE}_n^*(\boldsymbol{\varpi}, \boldsymbol{\omega})$. In the discussion following (3.1), we show that the EFPE is no-negative almost surely. Similarly, it is evident that γ_n is also nonnegative. We make an assumption as follows.

Assumption 3.4. $n/\{\bar{k}^2 M (\log M)^4\} \to \infty$ and $\gamma_n^{-1} n^{1/2} \bar{k} (\log n)^{1/2} \to 0$.

The first part of Assumption 3.4 imposes restrictions on the growth rates of $\bar{k}$ and M. The second part of Assumption 3.4 requires that γ_n grows at a rate no slower than $n^{1/2}\bar{k}(\log n)^{1/2}$, which is similar to the condition (21) of Zhang et al. (2013), condition (7) of Ando and Li (2014), and condition C.4 of Bai et al. (2022). In addition, when there is at least one true model included in the set of candidate models, $\gamma_n = 0$. Therefore, the second part of Assumption 3.4 implies that all candidate models are misspecified.

THEOREM 3.4. *If Assumptions 3.1 and 3.4 are satisfied, we have*

$$\frac{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}}, \widehat{\boldsymbol{\omega}})}{\inf_{(\boldsymbol{\varpi}, \boldsymbol{\omega}) \in \mathcal{H}} \mathrm{EFPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega})} = 1 + o_p(1).$$

Theorem 3.4 implies that the out-of-sample EFPE using the jackknife weight vector is asymptotically equal to the excess prediction error using the infeasible optimal weight vector.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

4. MONTE CARLO SIMULATIONS

In this section, we conduct simulations to evaluate the finite sample performance of the proposed model averaging estimator, and compare it with three other model selection or model averaging methods.

4.1. Data Generating Processes

For comparison, we consider two types of simulations. In type I simulations, the observed $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$ are i.i.d.; whereas in type II simulations, we allow the data to be non-i.i.d.

4.1.1. *I.I.D. Cases.* Here, we consider five data generating processes, each of which contains both homoscedasticity and heteroscedasticity cases. DGP1 and DGP2 are based on the same setup as that of Lu and Su (2015), while DGP1 is a common setting used in JMA model averaging, for example, Hansen (2007), Hansen and Racine (2012), and Zhang et al. (2013). DGP1 and DGP2 correspond to linear and nonlinear relationships between y_i and $\mathbf{x}_i$, respectively. The true model is included in the candidate models of DGP3, but not in the other DGPs. The error term in DGP4 accounts for a heavy-tailed distribution, whereas DGP5 specifies interdependencies among the explanatory variables. In addition, the candidate models of DGP3 is a non-nested setup such as Li et al. (2021).

DGP1:

$$y_i = \alpha \sum_{j=1}^{1000} j^{-1} x_{ij} + \xi_i,$$

where $x_{i1} = -1$ and $x_{ij}, j = 2, 3, \ldots$ are each i.i.d. $N(0,1)$ and are mutually independent of each other. We consider the following two situations for the term ξ_i: (i) $\xi_i \sim N(0,1)$ and independent of x_{ij}; (ii) $\xi_i = (\sum_{i=2}^{6} x_{ij}^2)\zeta_i$ and $\zeta_i \sim N(0,1)$ and independent of x_{ij}, representing homoscedastic and heteroscedastic errors, respectively. We consider different choices of α such that the population $R^2 = \{\text{var}(y_i) - \text{var}(\xi_i)\}/\text{var}(y_i) = 0.1, 0.2, \ldots, 0.9$. We consider three sample sizes $n = 100$, 200, and 400. The number of candidate models is given by $M = [3n^{1/3}]$ where $[\cdot]$ represents the nearest integer of $\cdot$. Furthermore, we consider nested candidate models, that is, the first model contains the first variable x_{i1}, the second model contains the first and second variables x_{i1}, x_{i2}, and so forth.

DGP2:

$$y_i = \alpha \left\{ x_{i1} + \sum_{j=2}^{25} j^{-1} \Phi(x_{ij}) \right\} + \xi_i,$$

where $x_{i1} = -1, x_{ij}, j = 2, 3, \ldots$ are each i.i.d. $N(0,1)$ and are mutually independent of each other, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. We consider $\xi_i = \zeta_i$ and $\xi_i = (0.01 + \sum_{i=2}^{11} x_{ij}^2)\zeta_i$ for the homoscedasticity and

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

heteroscedasticity cases where ζ_i is i.i.d. $N(0,1)$ and independent of x_{ij}'s. Except that we fix $M = 20$ for each case, the values and the setting of α are the same as those in DGP1.

DGP3:

$$y_i = \alpha \sum_{j=1}^{6} \beta_j x_{ij} + \xi_i,$$

where $x_{i1} = -1$ and $x_{ij}, j = 2, 3, \ldots$ are each i.i.d. $N(0,1)$ and are mutually independent of each other. The coefficients $\beta_3 = 0$ and $\beta_4 = 0.1$, and the other coefficients are 1. The first two covariates x_{i1} and x_{i2} are included in all candidate models. All possible combinations of the last four regressors constitute the candidate models (i.e., $M = 2^4 = 16$). The other settings are the same as DGP1, except that $\xi_i = (\sum_{i=2, i\neq 3}^{6} x_{ij}^2)\zeta_i$ in the heteroscedasticity case.

DGP4: The settings of DGP4 mirror those of DGP1, differing only in $\xi_i \sim$ skew-$t(0, 1, 5, -0.5)$ and $\zeta_i \sim$ skew-$t(0, 1, 5, -0.5)$ for the homoscedasticity and heteroscedasticity cases, respectively. Here, skew-$t(0, 1, 5, -0.5)$ denote the standardized skew t distribution of Hansen (1994), with degrees of freedom and skewness parameters set to $(5, -0.5)$.

DGP5:

$$y_i = \alpha \sum_{j=1}^{25} j^{-1} x_{ij} + \xi_i,$$

where $x_{i1} = -1$, and $x_{ij}, j = 2, 3, \ldots$ follow $N(0,1)$ but are not mutually independent. We set $\text{corr}(x_{ij}, x_{i,j+1}) = 0.7$, $\text{corr}(x_{ij}, x_{i,j+2}) = 0.5$, $\text{corr}(x_{ij}, x_{i,j+3}) = 0.3$, $\text{corr}(x_{ij}, x_{i,j+4}) = 0.1$, and all other correlation coefficients are 0. Except that we fix $M = 20$, the other settings are the same as DGP1.

4.1.2. *Non-i.i.d. Cases.* Although we restrict the data to be i.i.d in Assumption 3.1(i), it is necessary to discuss the application of our method to non-i.i.d data, since VaR and ES are widely used as risk measures on financial time series data.

Considering the generalizability of the autoregressive moving average-generalized autoregressive conditional heteroskedastic (ARMA–GARCH) model, we suggest utilizing ARMA–GARCH for the original time series as the initial step, and subsequently applying our proposed method, designed for i.i.d. data, to model the innovation. Specifically, for the time series data $\{z_t\}_{t=1}^n$, we consider the following ARMA(p,q)–GARCH(r,s) model:

$$z_t = \phi_0 + \sum_{i=1}^{p} \phi_i z_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t,$$

$$\varepsilon_t = \sigma_t \eta_t,$$

$$\sigma_t^2 = w + \sum_{i=1}^{r} b_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} c_j \sigma_{t-j}^2.$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

By quasi-maximum likelihood estimation, we can obtain the parameter estimator $(\widehat{\phi}_0, \widehat{\phi}_1, \ldots, \widehat{\phi}_p, \widehat{\psi}_1, \ldots, \widehat{\psi}_q, \widehat{w}, \widehat{b}_1, \ldots, \widehat{b}_r, \widehat{c}_1, \ldots, \widehat{c}_s)$ as well as the estimator $\{\widehat{\eta}_t\}_{t=1}^n$ for the innovation $\{\eta_t\}_{t=1}^n$. Then, the JMA estimators for the conditional VaR and conditional ES of $\widehat{\eta}_t$ are obtained by modeling the innovation $\widehat{\eta}_t$ using the JMA method in Section 2. In other words, by treating $\widehat{\eta}_t$ as y_i in our model, we obtain the JMA estimator $\widehat{\mathrm{VaR}}_\tau(\widehat{\eta}_t|\boldsymbol{x}_t)$ and $\widehat{\mathrm{ES}}_\tau(\widehat{\eta}_t|\boldsymbol{x}_t)$ for $\mathrm{VaR}_\tau(\widehat{\eta}_t|\mathbf{x}_t)$ and $\mathrm{ES}_\tau(\widehat{\eta}_t|\mathbf{x}_t)$. Finally, the conditional VaR and conditional ES of the time series $\{z_t\}_{t=1}^n$ are calculated as

$$\widehat{\mathrm{VaR}}_\tau(z_t|\mathcal{F}_{t-1}) = \widehat{\phi}_0 + \sum_{i=1}^{p} \widehat{\phi}_i z_{t-i} + \sum_{j=1}^{q} \widehat{\psi}_j \varepsilon_{t-j} + \sqrt{\widehat{w} + \sum_{i=1}^{r} \widehat{b}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \widehat{c}_j \sigma_{t-j}^2}\, \widehat{\mathrm{VaR}}_\tau(\widehat{\eta}_t|\boldsymbol{x}_t),$$

$$\widehat{\mathrm{ES}}_\tau(z_t|\mathcal{F}_{t-1}) = \widehat{\phi}_0 + \sum_{i=1}^{p} \widehat{\phi}_i z_{t-i} + \sum_{j=1}^{q} \widehat{\psi}_j \varepsilon_{t-j} + \sqrt{\widehat{w} + \sum_{i=1}^{r} \widehat{b}_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \widehat{c}_j \sigma_{t-j}^2}\, \widehat{\mathrm{ES}}_\tau(\widehat{\eta}_t|\boldsymbol{x}_t).$$

This section will also investigate the performance of $\mathrm{VaR}_\tau(z_t|\mathcal{F}_{t-1})$ and $\mathrm{ES}_\tau(z_t|\mathcal{F}_{t-1})$ estimators obtained from the above discussion, using the following data generation process.

DGP6:

$$z_t = 0.01 + 0.8 z_{t-1} + \varepsilon_t - 0.9\varepsilon_{t-j},$$
$$\varepsilon_t = \sigma_t \eta_t,$$
$$\sigma_t^2 = 0.02 + 0.15\varepsilon_{t-1}^2 + 0.8\sigma_{t-1}^2$$
$$\eta_t = \frac{1}{\alpha^2 \sum_{j=1}^{50} j^{-2} + 1}\left(\alpha \sum_{j=1}^{50} \frac{x_{tj}}{j} + \xi_t\right),$$

where $x_{tj}, j = 1, 2, \ldots$ are each i.i.d. $N(0,1)$ and are mutually independent of each other. $\xi_t \sim N(0,1)$ and independent of x_{tj}. $\left(\alpha^2 \sum_{j=1}^{50} j^{-2} + 1\right)^{-1}$ is a standardized coefficient, ensuring that the variance of η_t is 1. In line with DGP1–3, we explore various α values to cover a range of population $R^2 = \{\mathrm{var}(\eta_t) - \mathrm{var}((\alpha^2 \sum_{j=1}^{50} j^{-2} + 1)^{-1}\xi_t)\}/\mathrm{var}(\eta_t)$ from 0.1 to 0.9. Additionally, the number of candidate models M is the same as that in DGP1. The sample sizes are $n = 100, 200,$ and 400, and the mth candidate model has regressors $(1, x_{t1}, \ldots, x_{t,m-1})$. Furthermore, for consistency with the notation in (2.8), $\widehat{\mathrm{VaR}}_\tau(z_t|\mathcal{F}_{t-1})$ and $\widehat{\mathrm{ES}}_\tau(z_t|\mathcal{F}_{t-1})$ are treated as $\widehat{v}_i(\widehat{\boldsymbol{\varpi}})$ and $\widehat{e}_i(\widehat{\boldsymbol{\omega}})$, respectively.

4.2. Alternative Methods

We compare the proposed JMA estimator with the following three estimators:

1. Largest model (Largest): The largest model refers to the model with the largest dimension of covariates among the candidate models. Specifically, the largest model uses regressors $(x_{i1}, \ldots, x_{iM})$, with the exception of the largest model

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

for DGP3 using regressors $(x_{i1}, x_{i2}, x_{i4}, x_{i5}, x_{i6})$ since the coefficient of x_{i3} is equal to 0.

2. Cross-validation model selection (CVMS): Choose the candidate model that minimizes cv_m as the result of CVMS, where $\text{cv}_m = n^{-1}\sum_{i=1}^{n} L(y_i, \mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\vartheta}}_{-i,m}, \mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_{-i,m})$.
3. Scoring-weighted model averaging (SWMA): The relative score combining method proposed by Taylor (2020). The combined estimators are obtained by $\widehat{v}_i(\boldsymbol{\varpi}) = \sum_{m=1}^{M} w_{\text{SW},m}\boldsymbol{x}_{i,m}^{\top}\widehat{\boldsymbol{\vartheta}}_m$ and $\widehat{e}_i(\boldsymbol{\omega}) = \sum_{m=1}^{M} w_{\text{SW},m}\boldsymbol{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_m$, where

$$w_{\text{SW},m} = \frac{\exp\left(-\lambda\sum_{i=1}^{n} L(y_i, \boldsymbol{x}_{i,m}^{\top}\widehat{\boldsymbol{\vartheta}}_{i,m}, \boldsymbol{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_{i,m})\right)}{\sum_{j=1}^{M}\exp\left(-\lambda\sum_{i=1}^{n} L(y_i, \boldsymbol{x}_{ij}^{\top}\widehat{\boldsymbol{\vartheta}}_{ij}, \boldsymbol{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_{ij})\right)}$$

with the tuning parameter $\lambda > 0$ being the value that minimises the sum of the in-sample values of the loss function.

For each replication r, we calculated the EFPE with 100 out-of-sample observations as

$$\text{EFPE}(r) = \frac{1}{100}\sum_{s=1}^{100} L\left(y_s, \widehat{\text{VaR}}_{\tau}(y_s|\boldsymbol{x}_s), \widehat{\text{ES}}_{\tau}(y_s|\boldsymbol{x}_s)\right) - \frac{1}{100}\sum_{s=1}^{100} L(y_s, v_s, e_s), \quad \textbf{(4.1)}$$

where $\widehat{\text{VaR}}_{\tau}(y_s|\boldsymbol{x}_s)$ and $\widehat{\text{ES}}_{\tau}(y_s|\boldsymbol{x}_s)$ are estimators of $\text{VaR}_{\tau}(y_s|\boldsymbol{x}_s)$ and $\text{ES}_{\tau}(y_s|\boldsymbol{x}_s)$ by different methods, and $(v_s, e_s) \equiv (\text{VaR}_{\tau}(y_s|\boldsymbol{x}_s), \text{ES}_{\tau}(y_s|\boldsymbol{x}_s))$. Let $\text{EFPE}_{\text{out}}(r)$ be the infeasible EFPE (in-sample EFPE) calculated by 100 out-of-sample observations, that is, the $\widehat{\boldsymbol{\vartheta}}_m$, $\widehat{\boldsymbol{\theta}}_m$, $\widehat{\boldsymbol{\varpi}}_m$ and $\widehat{\boldsymbol{\omega}}_m$ in $\text{EFPE}(r)$ calculated from the out-of-sample data instead of the training data. Then, we average the out-of-sample prediction error over $R = 100$ replications, $\text{EFPE} = R^{-1}\sum_{r=1}^{R}\{\text{EFPE}(r) - \text{EFPE}_{\text{out}}(r)\}$. The smaller the EFPE is, the better the method in terms of the out-of-sample quantile prediction error. To facilitate comparisons, the EFPE values of all estimators are normalized by the EFPE of the infeasible optimal single model, as in Hansen (2007) and Lu and Su (2015).

4.3. Simulation Results

We only investigate the performance with $\tau = 0.1$ and 0.05, due to the limitation of Assumption 3.1(iv). Also because of Assumption 3.1(iv), we first normalize $\{y_i\}_{i=1}^{n}$ before conducting the simulation experiments of DGP1–5. Subsequently, we shift the normalized $\{y_i\}_{i=1}^{n}$ to ensure that the maximum true VaR of in-sample as well as out-of-sample is -2. As a rule of thumb, this treatment ensures that the in-sample estimator and the out-of-sample predictor of VaR and ES are always smaller than zero. For DGP6, for the same reason, we estimate $\text{VaR}_{\tau}(\widehat{\eta}_t|\mathbf{x}_t)$ and $\text{ES}_{\tau}(\widehat{\eta}_t|\mathbf{x}_t)$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

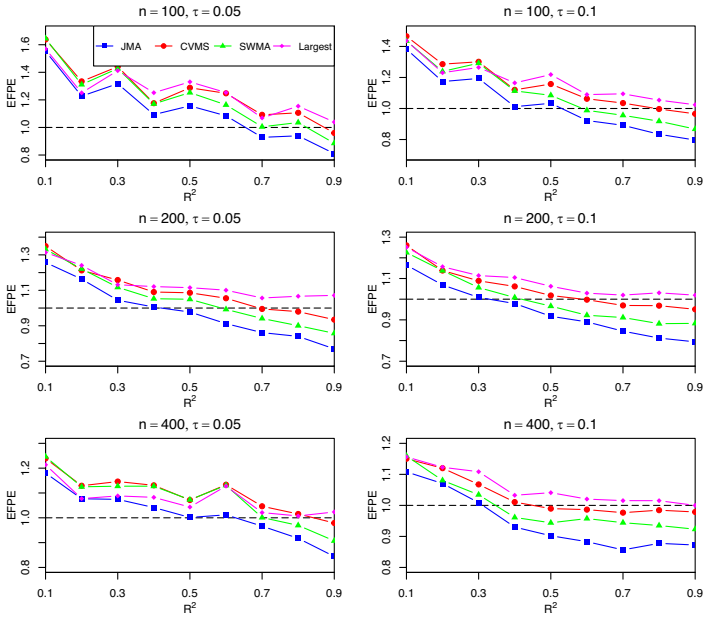


FIGURE 1. Out-of-sample performance: DGP1, homoscedasticity.

by first computing the estimators of VaR and ES for $\{\widehat{\eta}_i - \max_{1\le j\le n}\widehat{\eta}_j\}_{i=1}^n$, and then considering $\widehat{\mathrm{VaR}}_\tau(\widehat{\eta}_i - \max_{1\le j\le n}\widehat{\eta}_j|\boldsymbol{x}_t) + \max_{1\le j\le n}\widehat{\eta}_j$ and $\widehat{\mathrm{ES}}_\tau(\widehat{\eta}_i - \max_{1\le j\le n}\widehat{\eta}_j|\boldsymbol{x}_t) + \max_{1\le j\le n}\widehat{\eta}_j$ as the estimators of $\widehat{\mathrm{VaR}}_\tau(\widehat{\eta}_t|\boldsymbol{x}_t)$ and $\widehat{\mathrm{ES}}_\tau(\widehat{\eta}_t|\boldsymbol{x}_t)$.

The results are presented in Figures 1–11. Figures 1 and 2 show the results for DGP1 with homoscedastic and heteroscedastic errors, respectively. From Figures 1 and 2, we observe that the JMA estimator yields better performance than the other three estimators for all n and R^2. In most cases, especially when R^2 is large, the SWMA and the largest model seem to be the second best and worst methods, respectively. Moreover, there is no evidence that the gap between JMA and SWMA changes with an increasing R^2. That is, the dominance of JMA relative to SWMA does not diminish with the change in R^2.

For the results of DGP2 reported in Figures 3 and 4, the model averaging estimator dominates the other estimators, although the JMA estimator outperforms all other estimators only for small R^2. When R^2 is large, the performance of SWMA approaches that of JMA. In addition, there is always a significant gap between the performance of SVMS and Largest and that of JMA.

The results of DGP3 are presented in Figures 5 and 6. Figure 5 shows that, in the case of homoscedasticity, the JMA estimator outperforms all other estimators for large R^2. Figure 6 shows that for the heteroscedastic DGP3, the JMA estimator weakly dominates the other estimators when R^2 is large. Moreover, for all cases of DFP3, SWMA is better than CVMS and Largest when n is not small.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

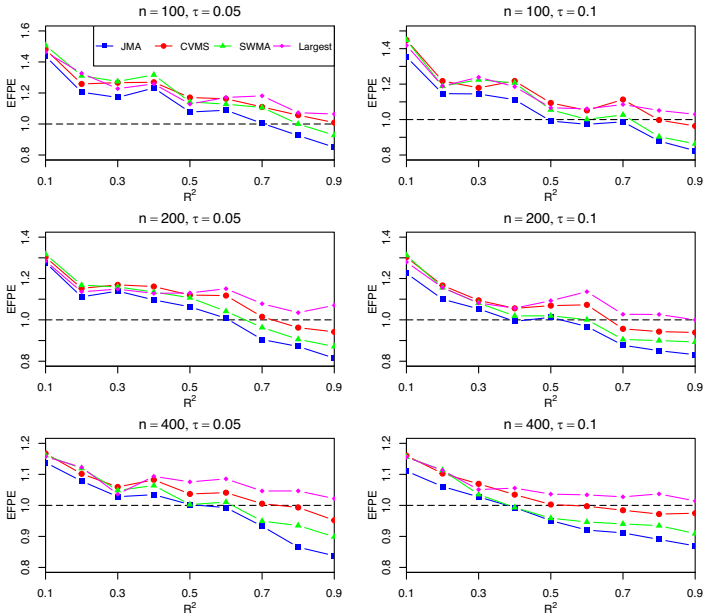


FIGURE 2. Out-of-sample performance: DGP1, heteroscedasticity.

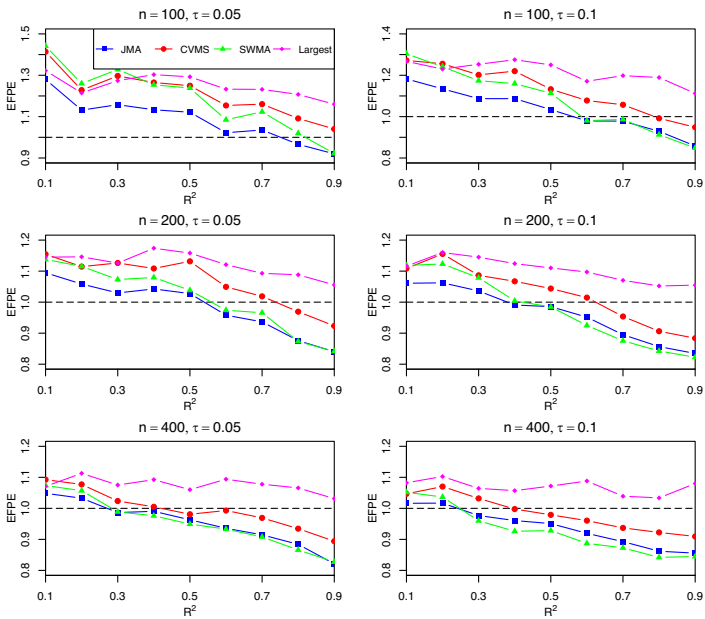


FIGURE 3. Out-of-sample performance: DGP2, homoscedasticity.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

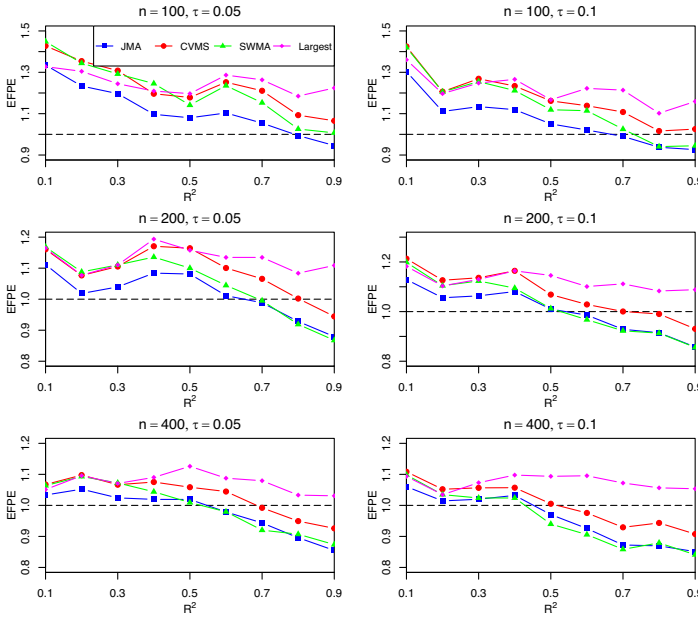


FIGURE 4. Out-of-sample performance: DGP2, heteroscedasticity.

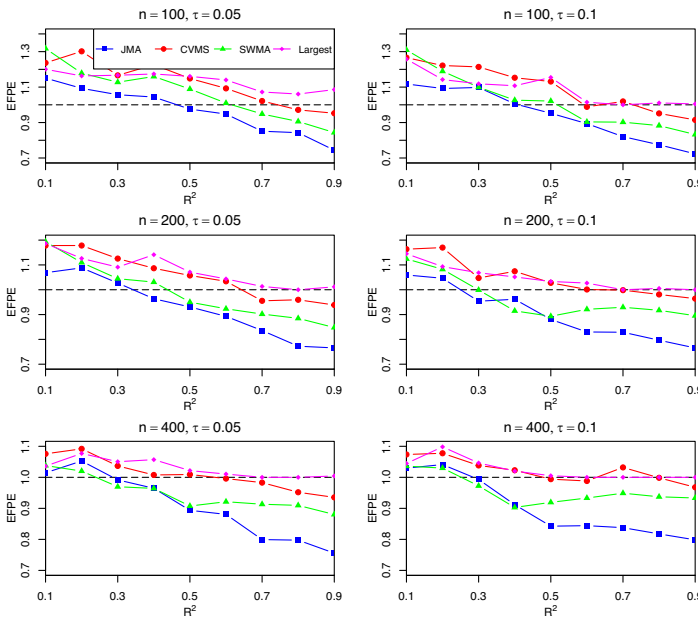


FIGURE 5. Out-of-sample performance: DGP3, homoscedasticity.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

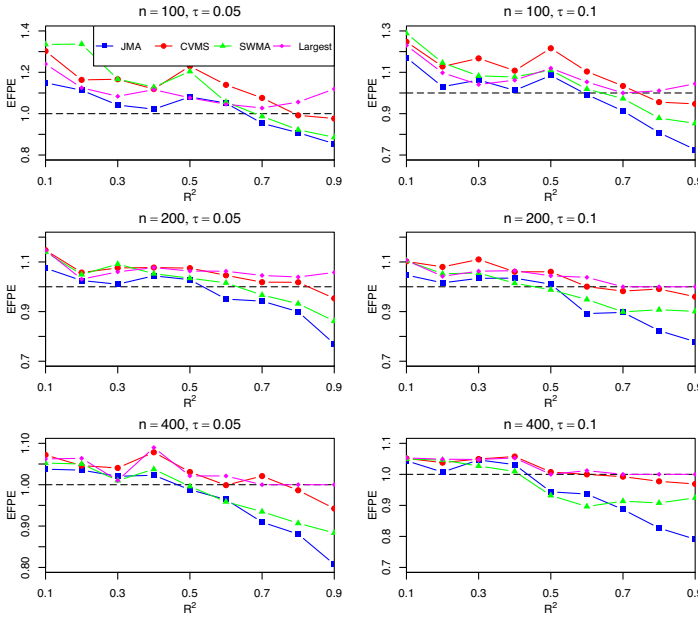


FIGURE 6. Out-of-sample performance: DGP3, heteroscedasticity.

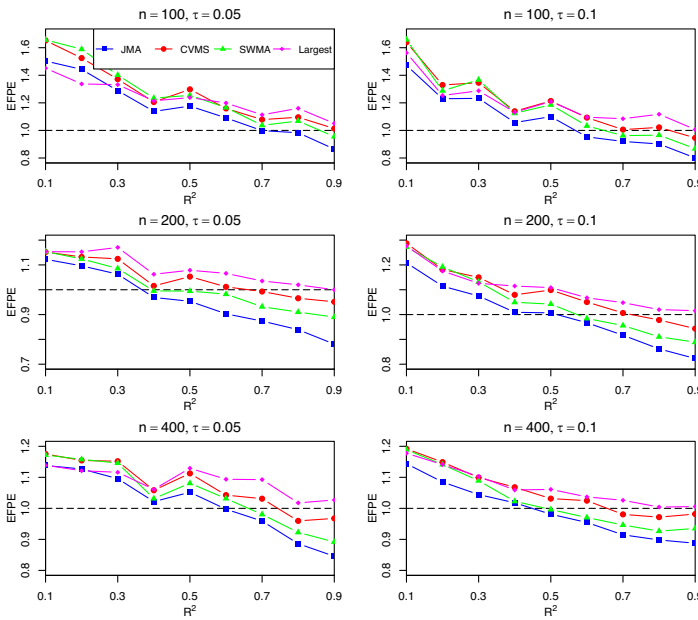


FIGURE 7. Out-of-sample performance: DGP4, homoscedasticity.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

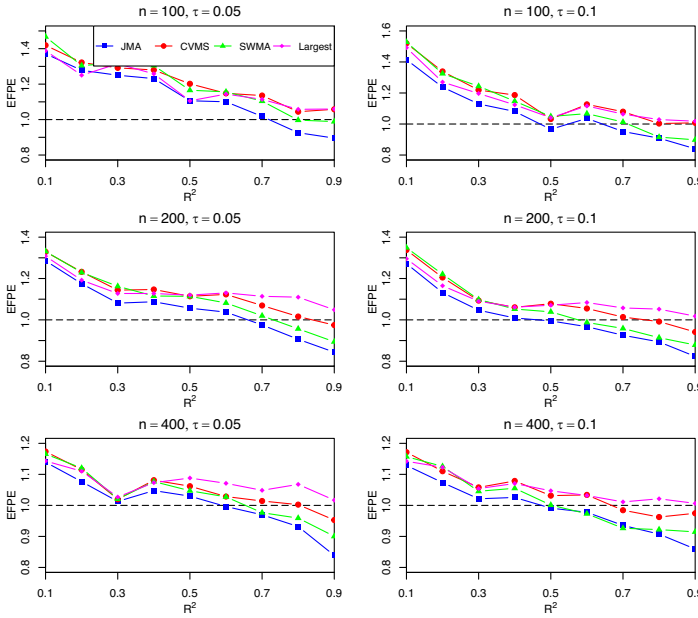


FIGURE 8. Out-of-sample performance: DGP4, heteroscedasticity.

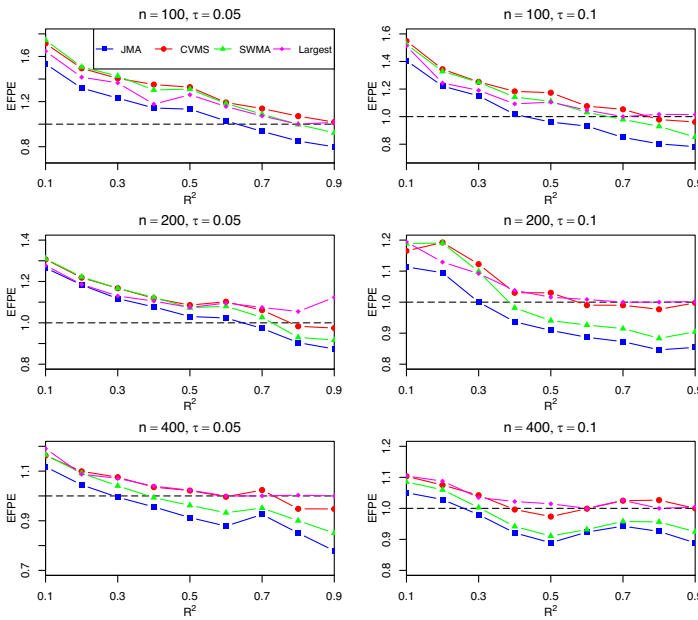


FIGURE 9. Out-of-sample performance: DGP5, homoscedasticity.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

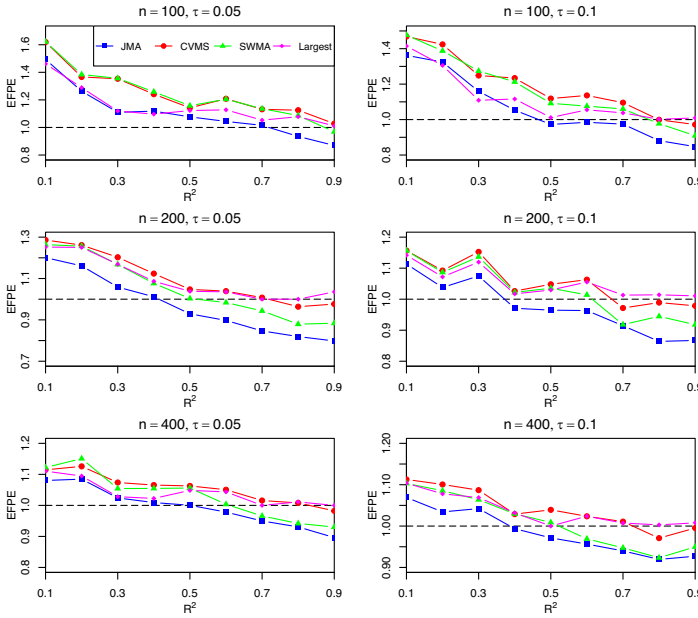


FIGURE 10. Out-of-sample performance: DGP5, heteroscedasticity.

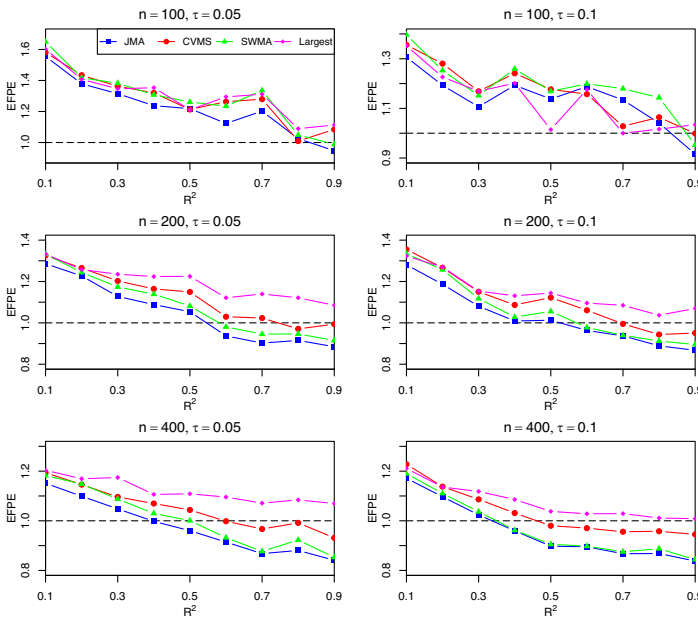


FIGURE 11. Out-of-sample performance: DGP6.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Figures 7–10 present the results for DGP4 and DGP5. These figures show that the JMA estimator outperforms other estimators in most cases. In addition, the dominance of JMA relative to other estimators does not diminish with an increasing R^2 or n. When R^2 is large, SWMA generally has suboptimal performance, while the largest model is the worst.

Figure 11 shows the results for DGP6, which is the non -i.i.d. data. We find that JMA outperforms the other methods across almost all values of n and R^2. SWMA methods always have suboptimal performance. As n increases, the performance of SWMA approaches that of JMA. For $n = 200$ and 400, there is a clear ordering of the performance of the four methods.

Overall, the performance of JMA is quite satisfactory. In most cases, JMA outperforms CV model selection and Largest model. In some special cases, such as small n or R^2, the SWMA method may show better performance. In general, the JMA estimator outperforms all other estimators when R^2 is large. The performance of Largest is poor in most cases, which implies the necessity of model selection and model averaging when constructing statistical models.

5. EMPIRICAL APPLICATIONS

5.1. Stock Returns

The data we used are taken from Welch and Goyal (2008) and updated to 2020. Many model averaging studies have analyzed similar data sets; see, for example, Elliott et al. (2013), Lu and Su (2015), Liu and Kuo (2016), and Sun et al. (2021). We select monthly data in the data set from January 1950 to December 2020, for a total of 852 observations. The dependent variable y represents stock returns which are measured as the difference between the continuously compounded return on the S&P 500 index, including dividends, and the Treasury bill rate.[2] The 10 regressors used in the data set are default return spread, inflation, dividend yield, treasury bill, long-term return, book-to-market ratio, earnings-to-price ratio, dividend-to-price ratio, net equity expansion, and default yield spread; see Welch and Goyal (2008) for a detailed description of the data. We sort the 10 regressions based on the absolute value of their correlation with the dependent variable and then construct 11 candidate nested models with regressors $\{1\}, \{1, x_1\}, \{1, x_1, x_2\}, \ldots, \{1, x_1, x_2, \ldots, x_{12}\}$. The mth candidate models used here are $\mathrm{VaR}_\tau(y_{t+1}|\mathbf{x}_t) = \sum_{j=1}^{m} \vartheta_j x_{tj}$ and $\mathrm{ES}_\tau(y_{t+1}|\mathbf{x}_t) = \sum_{j=1}^{m} \theta_j x_{tj}$. We construct one-period-ahead forecasts of the conditional ES and VaR of stock returns using various fixed in-sample sizes, $T_1 = 100$, 200, and 400.

The JMA, SWMA, CVMS, and largest model methods, which are introduced in the previous section, will be used for prediction. Unlike the simulation experiments in Section 4, since this is a one-period-ahead forecasts problem, we compare the forecast methods based on the out-of-sample

[2]To ensure that Assumption 3.2(iv) is satisfied, we use $y_i - \max_{1 \le j \le n} y_j$ as the dependent variable in our model. Sections 5.2 and 5.3 also does a similar treatment for the dependent variable.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Table 1. Out-of-sample EFPE for the stock returns data.

τ	T_1	JMA	SWMA	CVMS	Largest
0.1	100	**0.1196**	0.1226	0.1316	0.1210
0.1	200	**0.0977**	0.1055	0.1072	0.1045
0.1	400	**0.1002**	0.1031	0.1023	0.1157
0.05	100	**0.1231**	0.1297	0.1352	0.1449
0.05	200	**0.1258**	0.1352	0.1357	0.1409
0.05	400	**0.1347**	0.1387	0.1451	0.1357

Note: Lower values indicate better out-of-sample performance. The lowest values in each row are highlighted in bold.

EFPE $= \left[\sum_{t=T_1+1}^{T} \{L(y_t, \widehat{v}_t, \widehat{e}_t) - L_{\min}(y_t, \widehat{v}_t, \widehat{e}_t)\}\right] / (T - T_1)$, where $L_{\min}$ denotes the minimum loss in JMA, SWMA, CVMS and largest model methods, and all candidate models[3]. Furthermore, τ=0.05 and 0.1 are considered.

Table 1 reports the out-of-sample EFPE for JMA, SWMA, CVMS, and largest model estimators. From the table, model averaging outperforms the other methods. For $T_1 = 100$, 200, and 400, JMA clearly dominates all other methods. SWMA exhibited suboptimal results at $T_1 = 200$ and 400, highlighting the strength of the model averaging approach. The results from this real data analysis are consistent with those obtained from the simulation study.

5.2. Wage

In this subsection, we conduct the ES and VaR forecast exercises using Current Population Survey data for 1975.[4] The sample size is $n = 526$, and we use the logarithm of the average hourly wage as the dependent variable. According to the absolute value of the correlation between 20 regressors and the dependent variable, we choose 10 candidate nested models using the first 10 regressors; see, for example, Lu and Su (2015). These 10 regressors are professional occupation, education, female, tenure, service occupation, married, trade, SMSA, services, and clerk occupation. The detailed explanations of the 10 regressors in the data set are shown in Wooldridge (2015). We randomly divide the sample into a testing set of size $n - n_1$ and a training set of size n_1. The data from the training set and the testing set will be used to obtain the estimators and calculate the out-of-sample EFPE $= \left[\sum_{t\in\text{testing set}} \{L(y_t, \widehat{v}_t, \widehat{e}_t) - L_{\min}(y_t, \widehat{v}_t, \widehat{e}_t)\}\right] / (n - n_1)$ where $L_{\min}$ defined in Section 5.1, respectively. We consider different estimation sample sizes, $n_1 = 100$, 200, and 400, and we repeat this sample splitting 200 times and

[3]Since $\text{EFPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) = \text{FPE}_n(\boldsymbol{\varpi}, \boldsymbol{\omega}) - \mathbb{E}(L(y_0, v_0, e_0))$, we know that $\text{EFPE} - \text{EFPE}_{\min} = \left[\sum_{t=T_1+1}^{T} \{L(y_t, \widehat{v}_t, \widehat{e}_t) - L_{\min}(y_t, \widehat{v}_t, \widehat{e}_t)\}\right] / (T - T_1)$. That is, the out-of-sample EFPE in this section is the difference between the EFPE of each model and the $\text{EFPE}_{\min}$, where $\text{EFPE}_{\min}$ denotes the minimum EFPE in JMA, SWMA, CVMS, Largest, and all candidate models.

[4]The data can be found in the `wooldridge` package in R.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

TABLE 2. Out-of-sample EFPE for the wage data.

τ	n_1	JMA	SWMA	CVMS	Largest
0.1	100	**0.0069**	0.0088	0.0127	0.0089
0.1	200	**0.0031**	0.0099	0.0071	0.0065
0.1	400	**0.0048**	0.0146	0.0062	0.0070
0.05	100	**0.0134**	0.0136	0.0225	0.0135
0.05	200	**0.0058**	0.0182	0.0116	0.0093
0.05	400	**0.0059**	0.0233	0.0091	0.0095

Note: Lower values indicate better out-of-sample performance. The lowest values in each row are highlighted in bold.

TABLE 3. Dependent variable and regressors for the WTI data.

	Description	Correlation with y
y	Nominal West Texas Intermediate Spot Price	1
x_1	CRB Raw Materials Index	0.8833
x_2	MSCI World Index	0.3238
x_3	S&P 500 Index	0.2840
x_4	S&P GSCI Non-Energy Index	0.2737
x_5	CBOE Volatility Index	−0.2302
x_6	U.S. Dollar Index	−0.1625
x_7	Henry Hub Natural Gas Spot Price	0.0412
x_8	Baltic Exchange Dirty Tanker Index	0.0328
x_9	Three-month U.S. Treasury bill rate	0.0253
x_{10}	The total number of confirmed deaths, encompassing both victims and attackers, resulting from terrorist attacks in the Middle East and North Africa.	0.0006

calculate the average of the out-of-sample EFPE. Furthermore, $\tau = 0.05$ and 0.1 are considered.

Table 2 reports the EFPE of each method. Whether $\tau = 0.05$ or $\tau = 0.1$, it is clear that the JMA estimator dominates all other model averaging and model selection methods for all in-sample sizes n_1. The results of other methods have obvious gaps with JMA.

5.3. WTI Returns

In contrast to the empirical data in the previous two subsections that satisfy the assumption of i.i.d., we analyze non-i.i.d. financial time series data here. The data are daily from 1 January 2014 to 31 December 2019 with total number of observations $T = 973$. The dependent variable is WTI returns. Detailed descriptions of the dependent variable and 10 regressors are given in Table 3. We employ the identical model framework as utilized in DGP4. Moreover, as discussed in

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Table 4. Out-of-sample EFPE for the WTI returns data.

τ	T_1	JMA	SWMA	CVMS	Largest
0.1	100	**0.0598**	0.0650	0.0620	0.0624
0.1	200	**0.0482**	0.0485	0.0487	0.0514
0.1	400	**0.0255**	0.0262	0.0266	0.0264
0.05	100	**0.0798**	0.0818	0.0868	0.0848
0.05	200	**0.0663**	0.0696	0.0688	0.0744
0.05	400	**0.0384**	0.0385	0.0397	0.0448

Note: Lower values indicate better out-of-sample performance. The lowest values in each row are highlighted in bold.

Section 5.1, we conduct a one-period-ahead forecast by regressing η_t on x_{t-1}. The candidate models are constructed as Section 5.1, that is, the 11 candidate models with regressors $\{1\}, \{1, x_1\}, \ldots, \{1, x_1, x_2, \ldots, x_{10}\}$. Other settings, including out-of-sample EFPE, remain consistent with those in Section 5.1.

Table 4 shows the results of out-of-sample EFPE for JMA, SWMA, CVMS, and largest model estimators. Similarly, JMA significantly outperforms all other methods across all T_1 values. SWMA has suboptimal performance for $T_1 = 400$. The largest model at $T_1 = 200$ and 400 performs the poorest. These findings demonstrate that, as the sample size T_1 increases, the model selection method will select a more appropriate model, and the model averaging method will more closely approximate the true model.

6. CONCLUSION

In this article, we propose a JMA method for joint VaR and ES regression. We show that the model averaging estimator with weights chosen by leave-one-out cross-validation is asymptotically optimal in terms of minimizing the out-of-sample EFPE. The prediction performance in the simulation studies and empirical applications is satisfactory.

There are a few interesting extensions of this article. First, while this article focused on linear relationships between VaR and ES of y and $\mathbf{x}$, it would be highly desirable to study dynamic models, such as the generalized autoregressive score model in Patton et al. (2019). Another possible extension would be to extend the proposed method to high-dimensional cases, just as Wang et al. (2023) extended the JMA for quantile regressions by Lu and Su (2015) to high-dimensional quantile regression. In addition, the asymptotic properties of non-i.i.d. cases require further exploration. We leave these topics for future research.

APPENDIX

Proof of Theorem 3.4. We prove Theorem 3.4 in this section and relegate the proofs of other theorems to the Supplementary Material.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Let $\widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) = \mathrm{CV}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - n^{-1}\sum_{i=1}^{n} L(y_i, v_i, e_i)$. Note that $(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) \equiv \arg\min_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \mathrm{CV}_n(\boldsymbol{\varpi},\boldsymbol{\omega})$, and we also can see that $(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) = \arg\min_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega})$. To prove Theorem 3.4, we first show that

$$\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right), \tag{A.1}$$

and

$$\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}{\mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}\right| = o_p(1). \tag{A.2}$$

If (A.1) holds, combining with Assumption 3.4, we have

$$\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{\widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}{\mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}\right| = o_p(1). \tag{A.3}$$

Then, if (A.2) holds, combining (A.3) and Assumption 3.4,

$$\begin{aligned}
&\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{\widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega})}{\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega})}\right| \\
&\le \left(\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{\widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega})}{\mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}\right| + \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}{\mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}\right|\right) \\
&\qquad \times \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{\mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})}{\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega})}\right| \\
&= o_p(1).
\end{aligned}$$

From the definition of the infimum, there exist a nonnegative series $a_n \to 0$ and a vector $(\boldsymbol{\varpi}_n^*,\boldsymbol{\omega}_n^*) \in \mathcal{H}$ such that $\inf_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) + a_n = \mathrm{EFPE}_n(\boldsymbol{\varpi}_n^*,\boldsymbol{\omega}_n^*)$ and $a_n/\inf_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) = o_p(1)$. Then,

$$\frac{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\boldsymbol{\varpi}_n^*,\boldsymbol{\omega}_n^*)}\left(\frac{\widetilde{\mathrm{CV}}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) - \mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})} + 1\right) \le \frac{\widetilde{\mathrm{CV}}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) - \mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})} + 1. \tag{A.4}$$

By (A.3), (A.4), and Assumption 3.4, we have

$$\begin{aligned}
&\frac{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\boldsymbol{\varpi}_n^*,\boldsymbol{\omega}_n^*)} - 1 \\
&\le \left(\frac{\widetilde{\mathrm{CV}}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) - \mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})} - \frac{\widetilde{\mathrm{CV}}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) - \mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}\right) \\
&\qquad \times \left(\frac{\widetilde{\mathrm{CV}}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}}) - \mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})} + 1\right)^{-1} \\
&= o_p(1).
\end{aligned}$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Then,

$$
\begin{aligned}
0 &\le \frac{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\inf_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega})} - 1 \\
&\le \frac{\mathrm{EFPE}_n(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})}{\mathrm{EFPE}_n(\boldsymbol{\varpi}_n^*,\boldsymbol{\omega}_n^*)}\left(1+\frac{a_n}{\inf_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega})}\right) - 1 \\
&= o_p(1).
\end{aligned}
$$

Therefore, the asymptotic optimality in Theorem 3.4 is established for $(\widehat{\boldsymbol{\varpi}},\widehat{\boldsymbol{\omega}})$.

The remainder of this proof is to prove (A.1) and (A.2). Using (2.5), we observe that

$$
\begin{aligned}
&L(y_i, v_{i,m}(\Theta_m), e_{i,m}(\Theta_m)) \\
&\quad= -\frac{1}{\tau e_{i,m}(\Theta_m)}\mathbb{I}\left\{y_i \le v_{i,m}(\Theta_m)\right\}\{v_{i,m}(\Theta_m) - y_i\} + \frac{v_{i,m}(\Theta_m)}{e_{i,m}(\Theta_m)} + \log\{-e_{i,m}(\Theta_m)\} - 1 \\
&\quad= \frac{-\rho_\tau\left(y_i - v_{i,m}(\Theta_m)\right)}{\tau e_{i,m}(\Theta_m)} + \frac{y_i}{e_{i,m}(\Theta_m)} + \log\{-e_{i,m}(\Theta_m)\} - 1, \qquad \textbf{(A.5)}
\end{aligned}
$$

where $\rho_\tau(u) = u\{\tau - \mathbb{I}(u \le 0)\}$ is the check function. By using (2.7), (3.1), and (A.5), we have the following decomposition:

$$
\widetilde{\mathrm{CV}}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega}) = \widetilde{\mathrm{CV}}_{1n}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \widetilde{\mathrm{CV}}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \widetilde{\mathrm{CV}}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \widetilde{\mathrm{CV}}_{4n}(\boldsymbol{\varpi},\boldsymbol{\omega}),
$$

where

$$
\begin{aligned}
\widetilde{\mathrm{CV}}_{1n}(\boldsymbol{\varpi},\boldsymbol{\omega}) &= \frac{1}{n}\sum_{i=1}^n L(y_i, v_i, e_i) - \mathbb{E}\{L(y_0, v_0, e_0)\}, \\
\widetilde{\mathrm{CV}}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega}) &= \frac{1}{n}\sum_{i=1}^n \frac{y_i}{\sum_{m=1}^M \omega_m \boldsymbol{x}_{i,m}^\top \widehat{\boldsymbol{\theta}}_{-i,m}} - \mathbb{E}\left(\frac{y_0}{\sum_{m=1}^M \omega_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\theta}_m^*}\right), \\
\widetilde{\mathrm{CV}}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega}) &= \frac{1}{n}\sum_{i=1}^n \log\left(-\sum_{m=1}^M \omega_m \boldsymbol{x}_{i,m}^\top \widehat{\boldsymbol{\theta}}_{-i,m}\right) - \mathbb{E}\left\{\log\left(-\sum_{m=1}^M \omega_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\theta}_m^*\right)\right\}, \\
\widetilde{\mathrm{CV}}_{4n}(\boldsymbol{\varpi},\boldsymbol{\omega}) &= \frac{1}{n}\sum_{i=1}^n \frac{-\rho_\tau(y_i - \sum_{m=1}^M \varpi_m \boldsymbol{x}_{i,m}^\top \widehat{\boldsymbol{\vartheta}}_{-i,m})}{\tau\sum_{m=1}^M \omega_m \boldsymbol{x}_{i,m}^\top \widehat{\boldsymbol{\theta}}_{-i,m}} - \mathbb{E}\left\{\frac{-\rho_\tau(y_0 - \sum_{m=1}^M \varpi_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\vartheta}_m^*)}{\sum_{m=1}^M \omega_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\theta}_m^*}\right\}.
\end{aligned}
$$

Obviously, the proof of (A.1) is completed by showing that (i) $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_{1n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p(n^{-1/2}) = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$; (ii) $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$; (iii) $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$; (iv) $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_{4n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$. Note that (i) follows by the central limit theorem, so we only show (ii)–(iv) below.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

We first show (ii). We decompose $\widetilde{\mathrm{CV}}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \widetilde{\mathrm{CV}}_{2n,1}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \widetilde{\mathrm{CV}}_{2n,2}(\boldsymbol{\varpi},\boldsymbol{\omega})$, where

$$\widetilde{\mathrm{CV}}_{2n,1}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_{-i,m}} - \frac{y_i}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}\right),$$

$$\widetilde{\mathrm{CV}}_{2n,2}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*} - \mathbb{E}\left(\frac{y_0}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*}\right).$$

By Assumption 3.1(iv) and Theorem 3.3, we have

$$\begin{aligned}
&\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_{2n,1}(\boldsymbol{\varpi},\boldsymbol{\omega})| \\
&= \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{1}{n}\sum_{i=1}^{n}\frac{y_i\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}(\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_{-i,m})}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_{-i,m}\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}\right| \\
&\le \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} K_2^2\frac{1}{n}\sum_{i=1}^{n}\left|y_i\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\left(\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_{-i,m}\right)\right| \\
&\le K_2^2 \max_{1\le i\le n}\max_{1\le m\le M}\left\|\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_{-i,m}\right\| \max_{1\le m\le M}\frac{1}{n}\sum_{i=1}^{n}\|y_i\mathbf{x}_{i,m}\| \\
&= O_p\left(n^{-1/2}\bar{k}^{1/2}(\log n)^{1/2}\right)O_p(\bar{k}^{1/2}).
\end{aligned}$$

Since the first moment of y_i exists,

$$\mathbb{E}\left|\frac{y_i}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}\right| \le K_2\mathbb{E}|y_i| < \infty,$$

for all $(\boldsymbol{\varpi},\boldsymbol{\omega}) \in \mathcal{H}$. By the uniform law of large numbers (see, e.g., Newey and McFadden, 1994, Lemma 2.4), we have $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}|\widetilde{\mathrm{CV}}_{2n,2}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p(n^{-1/2})$. Consequently, $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}|\widetilde{\mathrm{CV}}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$.

(iii) Observe that $\widetilde{\mathrm{CV}}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \widetilde{\mathrm{CV}}_{3n,1}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\varpi},\boldsymbol{\omega}))$, where

$$\widetilde{\mathrm{CV}}_{3n,1}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n}\log\left(-\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\theta}}_{-i,m}\right) - \frac{1}{n}\sum_{i=1}^{n}\log\left(-\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right),$$

$$\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \frac{1}{n}\sum_{i=1}^{n}\log\left(-\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right) - \mathbb{E}\left\{\log\left(-\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*\right)\right\}.$$

By Assumption 3.1(iv) and Theorem 3.3, we have

$$\begin{aligned}
&\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\widetilde{\mathrm{CV}}_{3n,1}(\boldsymbol{\varpi},\boldsymbol{\omega})| \\
&= \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\frac{1}{n}\sum_{i=1}^{n}\log\left\{1 + \frac{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}(\widehat{\boldsymbol{\theta}}_{-i,m} - \boldsymbol{\theta}_m^*)}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}\right\}\right|
\end{aligned}$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

$$
\begin{aligned}
&\le \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}(\widehat{\boldsymbol{\theta}}_{-i,m}-\boldsymbol{\theta}_m^*)}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}\right| \\
&\le \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \frac{K_2}{n}\sum_{i=1}^{n}\left|\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\left(\widehat{\boldsymbol{\theta}}_{-i,m}-\boldsymbol{\theta}_m^*\right)\right| \\
&\le K_2 \max_{1\le i\le n}\max_{1\le m\le M}\left\|\widehat{\boldsymbol{\theta}}_{-i,m}-\boldsymbol{\theta}_m^*\right\| \max_{1\le m\le M}\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_{i,m}\| \\
&= O_p\left(n^{-1/2}\bar{k}^{1/2}(\log n)^{1/2}\right)O_p(\bar{k}^{1/2}),
\end{aligned}
$$

where the last line used the fact that

$$
\begin{aligned}
\max_{1\le m\le M}\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_{i,m}\| &\le \max_{1\le m\le M}\frac{1}{n}\sum_{i=1}^{n}E\|\mathbf{x}_{i,m}\| + \max_{1\le m\le M}\left|\frac{1}{n}\sum_{i=1}^{n}(E\|\mathbf{x}_{i,m}\|-\|\mathbf{x}_{i,m}\|)\right| \\
&= O(\bar{k}^{1/2})+o_p(1) = O_p(\bar{k}^{1/2}).
\end{aligned}
$$

Since $\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\varpi},\boldsymbol{\omega})$ is unrelated to $\boldsymbol{\varpi}$, we rewrite it as $\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega})$ for simplicity. Next, we show $\sup_{\boldsymbol{\omega}\in\mathcal{W}}|\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$, where $\mathcal{W} = \{\boldsymbol{\omega}\in[0,1]^M : \sum_{m=1}^{M}\omega_m = 1\}$. Define the metric $|\cdot|_1$ on $\mathcal{W}$ where $|\boldsymbol{\omega}-\bar{\boldsymbol{\omega}}|_1 = \sum_{m=1}^{M}|\omega_m-\bar{\omega}_m|$ for any $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_m)^{\top}\in\mathcal{W}$ and $\bar{\boldsymbol{\omega}} = (\bar{\omega}_1,\ldots,\bar{\omega}_m)^{\top}\in\mathcal{W}$. When M and $\bar{k}$ are finite, using the similar arguments in the proofs of Theorem 3.3 in Lu and Su (2015) and Theorem 3 in Wang et al. (2023), we can apply the Glivenko–Cantelli Theorem (e.g., Thm. 2.4.1 in Van der Vaart and Wellner, 1996) to show that $\sup_{\boldsymbol{\omega}\in\mathcal{W}}|\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$. When either $M\to\infty$ or $\bar{k}\to\infty$ as $n\to\infty$, let $h_n = 1/(\bar{k}n)$. We create grids using regions of the form $W_j = \{\boldsymbol{\omega} : |\boldsymbol{\omega}-\boldsymbol{\omega}_j|_1 \le h_n\}$. By selecting $\boldsymbol{\omega}_j = (\omega_{j1},\ldots,\omega_{jM})^{\top}$ to lay on a grid, $\mathcal{W}$ can be covered with $N = O(1/h_n^{M-1})$ regions $W_j, j = 1,\ldots,N$. Observe that

$$
\begin{aligned}
&\sup_{\boldsymbol{\omega}\in W_j}|\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega})-\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega}_j)| \\
&= \sup_{\boldsymbol{\omega}\in W_j}\left|\frac{1}{n}\sum_{i=1}^{n}\log\left(-\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right)-\frac{1}{n}\sum_{i=1}^{n}\log\left(-\sum_{m=1}^{M}\omega_{jm} \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right)\right| \\
&\le \sup_{\boldsymbol{\omega}\in W_j}\frac{1}{n}\sum_{i=1}^{n}\left|\log\left(-\sum_{m=1}^{M}\omega_m \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right)-\log\left(-\sum_{m=1}^{M}\omega_{jm} \mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right)\right| \\
&= \sup_{\boldsymbol{\omega}\in W_j}\frac{1}{n}\sum_{i=1}^{n}\left|\log\left\{1+\frac{\sum_{m=1}^{M}(\omega_m-\omega_{jm})\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}{\sum_{m=1}^{M}\omega_{jm}\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*}\right\}\right| \\
&\le \sup_{\boldsymbol{\omega}\in W_j}\frac{K_2}{n}\sum_{i=1}^{n}\left|\sum_{m=1}^{M}(\omega_m-\omega_{jm})\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right|
\end{aligned}
$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

$$\leq c_\Theta K_2 \max_{1\leq m\leq M} \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_{i,m}\| \sup_{\boldsymbol{\omega}\in W_j} \sum_{m=1}^{M} |\omega_m - \omega_{jm}|$$
$$\leq c_\Theta O_p(\bar{k}^{1/2}) h_n K_2 = O_p(n^{-1})$$

holds uniformly in j, where $c_\Theta = \max_{1\leq m\leq M} \|\boldsymbol{\theta}_m^*\| = O(\bar{k}^{1/2})$. Therefore

$$\begin{aligned}\sup_{\boldsymbol{\omega}\in\mathcal{W}} |\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega})| &= \max_{1\leq j\leq N} \sup_{\boldsymbol{\omega}\in\mathcal{W}} |\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega})| \\ &\leq \max_{1\leq j\leq N} |\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega}_j)| + \max_{1\leq j\leq N} \sup_{\boldsymbol{\omega}\in W_j} |\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega}) - \widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega}_j)| \\ &= \max_{1\leq j\leq N} |\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega}_j)| + O_p(n^{-1}).\end{aligned}$$

Let $\mu_i(\boldsymbol{\omega}_j) = \log(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m^*) - \mathbb{E}\{\log(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{0,m}^\top \boldsymbol{\theta}_m^*)\}$ and $e_n = (Mn\bar{k}^2)^{1/4}$. For any $\varepsilon > 0$, we have

$$\begin{aligned}\mathbb{P}\left\{\max_{1\leq j\leq N} |\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\omega}_j)| \geq 2\varepsilon\right\} &\leq \mathbb{P}\left[\max_{1\leq j\leq N}\left|\frac{1}{n}\sum_{i=1}^{n} \mu_i(\boldsymbol{\omega}_j)\mathbb{I}\{|\mu_i(\boldsymbol{\omega}_j)| \leq e_n\}\right| \geq \varepsilon\right] \\ &\quad + \mathbb{P}\left[\max_{1\leq j\leq N}\left|\frac{1}{n}\sum_{i=1}^{n} \mu_i(\boldsymbol{\omega}_j)\mathbb{I}\{|\mu_i(\boldsymbol{\omega}_j)| > e_n\}\right| \geq \varepsilon\right].\end{aligned}$$

Note that $\max_{1\leq m\leq M} \mathbb{E}(\mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m^*)^2 \leq \bar{k}\bar{\sigma}^2$ for some $\bar{\sigma}^2 < \infty$ and all $i = 0, 1, \ldots, n$. Let $(X)_+ = \max(X, 0)$ and $(X)_- = \min(X, 0)$. For any j, we have

$$\begin{aligned}&\mathrm{Var}\left[\mu_i(\boldsymbol{\omega}_j)\mathbb{I}\{|\mu_i(\boldsymbol{\omega}_j)| \leq e_n\}\right] \\ &\quad\leq \mathbb{E}\{|\mu_i(\boldsymbol{\omega}_j)|\}^2 \\ &\quad\leq 2\left\{\mathbb{E}\left|\log\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{0,m}^\top \boldsymbol{\theta}_m^*\right)\right|\right\}^2 + 2\mathbb{E}\left|\log\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m^*\right)\right|^2 \\ &\quad= 2\left[\mathbb{E}\left\{\log\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{0,m}^\top \boldsymbol{\theta}_m^*\right)_+\right\}\right]^2 + 2\left[\mathbb{E}\left\{\log\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{0,m}^\top \boldsymbol{\theta}_m^*\right)_-\right\}\right]^2 \\ &\qquad + 2\mathbb{E}\left\{\log\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m^*\right)_+\right\}^2 + 2\mathbb{E}\left\{\log\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m^*\right)_-\right\}^2 \\ &\quad\leq 2\left\{\mathbb{E}\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{0,m}^\top \boldsymbol{\theta}_m^*\right)\right\}^2 + 2\mathbb{E}\left(-\sum_{m=1}^{M} \omega_{jm}\mathbf{x}_{i,m}^\top \boldsymbol{\theta}_m^*\right)^2 + 4\log(K_2)^2 \\ &\quad\leq 4\bar{k}\bar{\sigma}^2 + 4\log(K_2)^2.\end{aligned}$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

Therefore, by Bernstein's inequality and Assumption 3.4, we have

$$
\begin{aligned}
&\mathbb{P}\left[\max_{1\le j\le N}\left|\frac{1}{n}\sum_{i=1}^{n}\mu_i(\boldsymbol{\omega}_j)\mathbb{I}\left\{\left|\mu_i(\boldsymbol{\omega}_j)\right|\le e_n\right\}\right|\ge\varepsilon\right]\\
&\quad\le\mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n}\max_{1\le j\le N}\mu_i(\boldsymbol{\omega}_j)\mathbb{I}\left\{\left|\mu_i(\boldsymbol{\omega}_j)\right|\le e_n\right\}\right|\ge\varepsilon\right]\\
&\qquad+\mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n}\min_{1\le j\le N}\mu_i(\boldsymbol{\omega}_j)\mathbb{I}\left\{\left|\mu_i(\boldsymbol{\omega}_j)\right|\le e_n\right\}\right|\ge\varepsilon\right]\\
&\quad\le 4\exp\left(-\frac{n\varepsilon^2}{8\bar{k}\bar{\sigma}^2+8\log(K_2)^2+2e_n\varepsilon/3}\right)\\
&\quad=4\exp\left\{-\frac{n\varepsilon^2}{8\bar{k}\bar{\sigma}^2+8\log(K_2)^2+2\varepsilon\left(M\bar{k}^2n\right)^{1/4}/3}\right\}\\
&\quad=O\left(\exp\left(-\left(\frac{n^3}{M\bar{k}^2}\right)^{1/4}\right)\right)\\
&\quad=o_p\left(\left(\frac{n^3}{M\bar{k}^2}\right)^{-1/4}\right).
\end{aligned}
$$

On the other hand, by Boole's inequalities,

$$
\begin{aligned}
&\mathbb{P}\left[\max_{1\le j\le N}\left|\frac{1}{n}\sum_{i=1}^{n}\mu_i(\boldsymbol{\omega}_j)\mathbb{I}\left\{\left|\mu_i(\boldsymbol{\omega}_j)\right|>e_n\right\}\right|\ge\varepsilon\right]\\
&\quad\le\mathbb{P}\left\{\max_{1\le j\le N}\max_{1\le i\le n}\left|\mu_i(\boldsymbol{\omega}_j)\right|>e_n\right\}\\
&\quad\le\mathbb{P}\left\{\max_{1\le j\le N}\max_{1\le i\le n}\left|\log\left(-\sum_{m=1}^{M}\omega_{jm}\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right)\right|>e_n-\log(K_2)\right\}\\
&\quad\le\mathbb{P}\left\{\max_{1\le j\le N}\max_{1\le i\le n}\left|-\sum_{m=1}^{M}\omega_{jm}\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right|>e_n-\log(K_2)\right\}\\
&\quad\le\mathbb{P}\left\{\max_{1\le m\le M}\max_{1\le i\le n}\left|-\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right|>e_n-\log(K_2)\right\}\\
&\quad\le 1-\mathbb{P}\left(\left|-\boldsymbol{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right|>e_n-\log(K_2)\right)^{Mn},
\end{aligned}
$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

where the second inequality used the fact as follows

$$\mathbb{P}\left\{|\mu_i(\boldsymbol{\omega}_j)| > e_n\right\} = \mathbb{P}\left[\left|\log\left(-\sum_{m=1}^{M}\omega_{jm}\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right) - \mathbb{E}\left\{\log\left(-\sum_{m=1}^{M}\omega_{jm}\mathbf{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*\right)\right\}\right| > e_n\right]$$
$$\leq \mathbb{P}\left\{\left|\log\left(-\sum_{m=1}^{M}\omega_{jm}\mathbf{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right)\right| > e_n - \log(K_2)\right\}$$

hold when $e_n > \log(K_2) + \log(\bar{k}^{1/2}\bar{\sigma})$, that is, n is large enough. By Markov's inequalities and Lebesgue dominated convergence theorem,

$$\mathbb{P}\left(\left|-\boldsymbol{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right| < e_n - \log(K_2)\right) \leq \frac{\mathbb{E}\left[\left|-\boldsymbol{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right|^8 \mathbb{I}\left\{\left|-\boldsymbol{x}_{i,m}^{\top}\boldsymbol{\theta}_m^*\right| > e_n - \log(K_2)\right\}\right]}{\{e_n - \log(K_2)\}^8} = o_p\left(e_n^{-8}\right).$$

Then,

$$\mathbb{P}\left[\max_{1\leq j\leq N}\left|\frac{1}{n}\sum_{i=1}^{n}\mu_i(\boldsymbol{\omega}_j)\mathbb{I}\left\{|\mu_i(\boldsymbol{\omega}_j)| > e_n\right\}\right| \geq \varepsilon\right]$$
$$\leq 1 - \left(1 - o_p\left(\left(Mn\bar{k}^2\right)^{-2}\right)\right)^{Mn} \leq 1 - \left(1 - o_p\left((Mn)^{-2}\right)\right)^{Mn} = o_p\left((Mn)^{-1}\right).$$

It follows that $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}|\widetilde{\mathrm{CV}}_{3n,2}(\boldsymbol{\varpi},\boldsymbol{\omega})| = o_p\left(M^{1/4}\bar{k}^{1/2}n^{-3/4}\right) = o_p(n^{-1/2})$. Consequently, $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}|\widetilde{\mathrm{CV}}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$.

(iv) By Assumption 3.1(iv) and Theorem 3.3, we have

$$|\widetilde{\mathrm{CV}}_{4n}(\boldsymbol{\varpi},\boldsymbol{\omega})|$$
$$\leq K_2\left|\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - \sum_{m=1}^{M}\varpi_m\mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\vartheta}}_{-i,m}\right) - \mathbb{E}\left\{\rho_\tau\left(y_0 - \sum_{m=1}^{M}\varpi_m\mathbf{x}_{0,m}^{\top}\boldsymbol{\vartheta}_m^*\right)\middle|\mathcal{D}_n\right\}\right|.$$

Under Assumptions 3.1 and 3.3 and $n/\{\bar{k}^2M(\log M)^4\} \to \infty$, combining the similar technique in the proof of $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}|\widetilde{\mathrm{CV}}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$ and the arguments in the proof of Theorem 3.3 of Lu and Su (2015), it is easy to prove that

$$\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(y_i - \sum_{m=1}^{M}\varpi_m\mathbf{x}_{i,m}^{\top}\widehat{\boldsymbol{\vartheta}}_{-i,m}\right) - \mathbb{E}\left\{\rho_\tau\left(y - \sum_{m=1}^{M}\varpi_m\mathbf{x}_{0,m}^{\top}\boldsymbol{\vartheta}_m^*\right)\middle|\mathcal{D}_n\right\}\right|$$
$$= O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right).$$

Consequently, $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}}|\widetilde{\mathrm{CV}}_{4n}(\boldsymbol{\varpi},\boldsymbol{\omega})| = o_p(1)$ and we obtain (A.1).

Next, we are going to verify (A.2). We decompose $\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})$ as follows:

$$\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega}),$$
$$= \Delta\mathrm{EFPE}_{1n}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \Delta\mathrm{EFPE}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega}) + \Delta\mathrm{EFPE}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega}),$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

where

$$\Delta\text{EFPE}_{1n}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \mathbb{E}\left(\frac{y_0}{\sum_{m=1}^{M}\omega_m \boldsymbol{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*}\right) - \mathbb{E}\left(\frac{y_0}{\sum_{m=1}^{M}\omega_m \boldsymbol{x}_{0,m}^{\top}\widehat{\boldsymbol{\theta}}_m}\middle|\mathcal{D}_n\right),$$

$$\Delta\text{EFPE}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \mathbb{E}\left\{\log\left(-\sum_{m=1}^{M}\omega_m \boldsymbol{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*\right)\right\} - \mathbb{E}\left\{\log\left(-\sum_{m=1}^{M}\omega_m \boldsymbol{x}_{0,m}^{\top}\widehat{\boldsymbol{\theta}}_m\right)\middle|\mathcal{D}_n\right\},$$

$$\Delta\text{EFPE}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega}) = \mathbb{E}\left\{\frac{-\rho_\tau(y_0 - \sum_{m=1}^{M}\varpi_m \boldsymbol{x}_{0,m}^{\top}\boldsymbol{\vartheta}_m^*)}{\sum_{m=1}^{M}\omega_m \boldsymbol{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*}\right\} - \mathbb{E}\left\{\frac{-\rho_\tau(y_0 - \sum_{m=1}^{M}\varpi_m \boldsymbol{x}_{0,m}^{\top}\widehat{\boldsymbol{\vartheta}}_m)}{\sum_{m=1}^{M}\omega_m \boldsymbol{x}_{0,m}^{\top}\widehat{\boldsymbol{\theta}}_m}\middle|\mathcal{D}_n\right\}.$$

Then we have

$$\begin{aligned}
&\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\Delta\text{EFPE}_{1n}(\boldsymbol{\varpi},\boldsymbol{\omega})| \\
&= \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\mathbb{E}\left\{\frac{y_0\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m^*)}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\boldsymbol{\theta}_m^* \sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\widehat{\boldsymbol{\theta}}_m}\middle|\mathcal{D}_n\right\}\right| \\
&\le \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} K_2^2\mathbb{E}\left\{\left|y_0\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\left(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m^*\right)\right|\middle|\mathcal{D}_n\right\} \\
&\le K_2^2 \max_{1\le m\le M}\left\|\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m^*\right\| \max_{1\le m\le M}\|\mathbb{E}|y_0\mathbf{x}_{0,m}|\| \\
&= O_p\left(n^{-1/2}\bar{k}^{1/2}(\log n)^{1/2}\right)O_p(\bar{k}^{1/2}),
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\Delta\text{EFPE}_{2n}(\boldsymbol{\varpi},\boldsymbol{\omega})| \\
&= \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left|\mathbb{E}\left[\log\left\{1 - \frac{\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}(\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_m)}{\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*}\right\}\middle|\mathcal{D}_n\right]\right| \\
&\le \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \mathbb{E}\left\{\frac{\left|\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\left(\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_m\right)\right|}{|\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\boldsymbol{\theta}_m^*|}\middle|\mathcal{D}_n\right\} \\
&\le \sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} K_2\mathbb{E}\left\{\left|\sum_{m=1}^{M}\omega_m \mathbf{x}_{0,m}^{\top}\left(\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_m\right)\right|\middle|\mathcal{D}_n\right\} \\
&\le K_2 \max_{1\le m\le M}\left\|\boldsymbol{\theta}_m^* - \widehat{\boldsymbol{\theta}}_m\right\| \max_{1\le m\le M}\|E|\mathbf{x}_{0,m}|\| \\
&= O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right).
\end{aligned}$$

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

By Assumption 3.1(iv) and Theorem 3.3, we have

$$|\Delta\mathrm{EFPE}_{3n}(\boldsymbol{\varpi},\boldsymbol{\omega})| \le K_2 \left| \mathbb{E}\left\{ -\rho_\tau\Big(y_0 - \sum_{m=1}^{M} \varpi_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\vartheta}_m^*\Big) \right\} - \mathbb{E}\left\{ -\rho_\tau\Big(y_0 - \sum_{m=1}^{M} \varpi_m \boldsymbol{x}_{0,m}^\top \widehat{\boldsymbol{\vartheta}}_m\Big) \,\middle|\, \mathcal{D}_n \right\} \right|.$$

Under Assumptions 3.1 and 3.3 and $n/\{\bar{k}^2 M(\log M)^4\} \to \infty$, combining the arguments in the proof of Theorem 3.3 of Lu and Su (2015), it is easy to prove that

$$\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} \left| \mathbb{E}\left\{ -\rho_\tau\Big(y_0 - \sum_{m=1}^{M} \varpi_m \boldsymbol{x}_{0,m}^\top \boldsymbol{\vartheta}_m^*\Big) \right\} - \mathbb{E}\left\{ -\rho_\tau\Big(y_0 - \sum_{m=1}^{M} \varpi_m \boldsymbol{x}_{0,m}^\top \widehat{\boldsymbol{\vartheta}}_m\Big) \,\middle|\, \mathcal{D}_n \right\} \right| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right).$$

Consequently, $\sup_{(\boldsymbol{\varpi},\boldsymbol{\omega})\in\mathcal{H}} |\mathrm{EFPE}_n(\boldsymbol{\varpi},\boldsymbol{\omega}) - \mathrm{EFPE}_n^*(\boldsymbol{\varpi},\boldsymbol{\omega})| = O_p\left(n^{-1/2}\bar{k}(\log n)^{1/2}\right)$.

This completes the proof of Theorem 3.4. □

SUPPLEMENTARY MATERIAL

Jiao, S., Xu, W., Ye, W., and Zhang, X. (2024). Supplement to "Optimal Model Averaging for Joint Value-at-Risk and Expected Shortfall Regression". *Econometric Theory* Supplementary Material. To view, please visit https://doi.org/10.1017/S0266466624000288.

REFERENCES

Ando, T., & Li, K.-C. (2014). A model-averaging approach for high-dimensional regression. *Journal of the American Statistical Association*, 109(505), 254–265.

Bai, Y., Jiang, R., & Zhang, M. (2022). Optimal model averaging estimator for expectile regressions. *Journal of Statistical Planning and Inference*, 217, 204–223.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, 53(2), 603–618.

Clyde, M., & George, E. I. (2004). Model uncertainty. *Statistical Science*, 19(1), 81–94.

Dimitriadis, T., & Bayer, S. (2019). A joint quantile and expected shortfall regression framework. *Electronic Journal of Statistics*, 13(1), 1823–1871.

Elliott, G., Gargano, A., & Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2), 357–373.

Fissler, T., & Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *The Annals of Statistics*, 44(4), 1680–1707.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705–730.

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

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press

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

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4), 382–417.

Koenker, R., & Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50(1), 43–61.

Li, X., Zou, G., Zhang, X., & Zhao, S. (2021). Least squares model averaging based on generalized cross validation. *Acta Mathematicae Applicatae Sinica, English Series*, 37(3), 495–509.

Liu, C.-A., & Kuo, B.-S. (2016). Model averaging in predictive regressions. *The Econometrics Journal*, 19(2), 203–231.

Lu, X., & Su, L. (2015). Jackknife model averaging for quantile regressions. *Journal of Econometrics*, 188(1), 40–58.

Magnus, J. R., Powell, O., & Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154(2), 139–153.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. and Mcfadden, D., editors. *Handbook of Econometrics*, 4, 2111–2245.

Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2), 388–413.

Peng, J., & Yang, Y. (2022). On improvability of model selection by model averaging. *Journal of Econometrics*, 229(2), 246–262.

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

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1), 121–133.

Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting*, 36(2), 428–441.

Tu, Y., & Wang, S. (2020). Jackknife model averaging for expectile regressions in increasing dimension. *Economics Letters*, 197, 109607.

Van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer.

Wang, M., Zhang, X., Wan, A. T., You, K., & Zou, G. (2023). Jackknife model averaging for high-dimensional quantile regression. *Biometrics*, 79(1), 178–189.

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, 16(2), 419–441.

Weiss, A. A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory*, 7(1), 46–68.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455–1508.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage Learning.

Yuan, Z., & Yang, Y. (2005). Combining linear regression models: When and how? *Journal of the American Statistical Association*, 100(472), 1202–1214.

Zhang, X., Wan, A. T., & Zou, G. (2013). Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics*, 174(2), 82–94.

https://doi.org/10.1017/S0266466624000288 Published online by Cambridge University Press