Composite Absolute Value and Sign Forecasts

André B.M. Souza*

November 9, 2020

Job Market Paper

[Click here for the latest version]

Abstract

This paper introduces composite absolute value and sign (CAVS) forecasts, a non-linear framework that combines forecasts of the sign and absolute value of a time series into conditional mean forecasts. In contrast to linear models, the proposed framework allows different predictors to separately impact the sign and absolute value of the target series. Among other results, I show that the conditional mean can be accurately approximated by the product of the mean squared error optimal sign and absolute value forecasts. An empirical application using the FRED-MD dataset shows that CAVS forecasts substantially outperform linear forecasts for series that exhibit persistent volatility dynamics, such as output and interest rates.

Keywords: Forecasting, Directional Predictability, Machine Learning.

JEL classification: C53, C22, C52, C58.

*Universitat Pompeu Fabra and Barcelona Graduate School of Economics.
e-mail: andre.souza@upf.edu, website: www.andrebmsouza.com.
I am very grateful to Christian Brownlees for the invaluable guidance and support. I would like to thank Kirill Evdokimov, Geert Meesters, Katerina Petrova and Barbara Rossi for providing numerous helpful comments. I would also like to thank the participants of the 8th SIDE Workshop for Econometrics and Empirical Economics (Bertinoro, 03-04 September 2020), in particular Francesco Ravazzolo for a detailed discussion of this paper. I acknowledge financial support from the Spanish Ministry of Science and Technology (FPI Grant BES-2017-080615).

1 Introduction

Forecasting macroeconomic and financial time series is a challenging task. A ubiquitous finding in the economic forecasting literature is that linear models for the conditional mean improve only marginally, if at all, over simple benchmarks (Stock and Watson, 2007; Goyal and Welch, 2008). In contrast, standard volatility models have shown considerable success in capturing volatility dynamics (Engle, 1982; Andersen *et al.*, 2006; Brownlees *et al.*, 2011) and there is evidence of directional predictability (Leung *et al.*, 2000; Diebold *et al.*, 2007; Nyberg, 2011). Several explanations are available for the unconvincing performance of linear forecasts. The predictable component of the target series may be small relative to the unpredictable error term, in which case even a correctly specified model will display only mild gains over simple benchmarks. Alternatively, this finding may be taken as evidence of model misspecification, implying that the class of models considered is not rich enough to exploit all available information.

In this work, I introduce composite absolute value and sign (CAVS) forecasts, a nonlinear forecasting framework that exploits predictability in signs and absolute values to generate conditional mean forecasts.[1] Based on the fact that any random variable Y_t can be written as $|Y_t|\text{sign}(Y_t)$, CAVS forecasts are defined as a function of mean squared error (MSE) optimal sign and absolute value forecasts. In contrast to linear models, in which conditional mean predictors must impact both the sign and absolute value, CAVS allows for different predictors to separately affect each of the components of the target series. In contrast to general nonlinear models, CAVS forecasts are simple to interpret and are designed to exploit two specific features that figure prominently in the macroeconomic and financial forecasting literature: volatility and sign predictability.

I introduce a framework to formalize CAVS forecasts and study its properties. The proposed framework is employed to establish three results. First, the conditional mean can be written as the product of MSE optimal forecasts of signs and absolute values as well as a covariance term that can be explicitly modeled. If the underlying data-

[1] I assume throughout that the sign of the target series is not constant, as would be the case for series expressed in growth rates.

generating process (DGP) is additive with symmetric shocks, the covariance term will be small relative to the variance of the shocks. Second, I provide an upper bound to the MSE of CAVS forecasts that scales with the losses in sign and absolute value forecasting. This result highlights that CAVS-based forecasts are particularly suited for series that exhibit persistent volatility dynamics, and hence absolute value predictability. Third, I study a nonlinear DGP in which variables may affect signs and absolute values differently. I show that, for this DGP, the MSE of the best linear predictor increases quadratically with the degree of nonlinearity. A simulation study highlights that departures from linearity generate substantial MSE gains for CAVS models.

The proposed framework is applied to the FRED-MD dataset (McCracken and Ng, 2016), which consists of 128 monthly financial and macroeconomic time series. I construct and evaluate 1, 3, 6, and 12 months ahead forecasts for the conditional mean, the absolute value and the sign of each series in the dataset. It is well-known that when the number of predictors is large, dimension reduction techniques may improve forecast accuracy (Stock and Watson, 2012; Ng, 2013; Kim and Swanson, 2014). I consider principal components regression (PCR), ridge and LASSO as the baseline linear models. CAVS forecasts are constructed as the product of sign and absolute value forecasts, and a number of specifications may be entertained. I consider CAVS-PCR, CAVS-Ridge, and CAVS-LASSO as the baseline CAVS models. Each baseline model is constructed using the same method to forecast both the sign and the absolute value of the target variable. In addition to the baseline CAVS models, I consider absolute value forecasts implied by a GARCH(1,1), sign forecasts obtained by a variety of machine learning algorithms, as well as combinations of sign and absolute value forecasts constructed using different methods to forecast each component. Detailed results are presented for the series considered in Kim and Swanson (2014). These include the unemployment rate, personal income less transfer payments, the 10-year treasury rate, the consumer price index, the producer price index, nonfarm payroll employment, industrial production, M2 money stock, and the S&P 500 index. In addition to the variables considered in Kim and Swanson (2014), I present detailed results for the federal funds rate, which may be of particular interest to private sector forecasters.

A number of findings emerge from the empirical application. First, CAVS-based directional and conditional mean forecasts outperform their linear counterparts for the majority of the selected series across all horizons considered. In particular, CAVS-based forecasts are substantially more accurate than linear forecasts for the federal funds rate, industrial production, nonfarm payroll employment, and the S&P 500 across all horizons. Among the linear models considered, ridge and LASSO display similar performance and outperform PCR, particularly for short forecast horizons. All baseline CAVS specifications considered perform similarly, with CAVS-Ridge modestly outperforming CAVS-LASSO and CAVS-PCR.

Second, I compare the performances of CAVS-Ridge and PCR for all components of the FRED-MD dataset. CAVS-Ridge outperforms PCR for the majority of the FRED-MD components, and is particularly successful for series that exhibit persistent conditional volatility dynamics, such as interest rates, stocks, and output series. As the forecast horizon increases, the CAVS-Ridge performance gains become widespread, outperforming PCR for about 75% of the FRED-MD components. Notably, these findings remain qualitatively the same for any choice of CAVS specification and linear benchmark considered.

Finally, I explore a number of alternative CAVS specifications. In addition to combinations between signs and absolute values obtained by PCR, Ridge, and LASSO, I consider absolute value forecasts based on a GARCH(1,1) and sign forecasts obtained from random forests (Breiman, 2001), AdaBoost (Freund and Schapire, 1995), k-nearest neighbors (kNN; see Devroye *et al.*, 1996) and neural networks (NN; see White, 2006). All specifications perform similarly, with GARCH(1,1)-Ridge displaying modest gains over the baseline CAVS specifications. Among the machine learning algorithms, the most accurate forecasts are achieved by methods based on regression trees. In particular, random forests produce more accurate forecasts when compared to the baseline CAVS specifications for 20% of the selected series. Additionally, I consider forecast combinations and model selection strategies. In line with Timmerman (2006) and Diebold and Shin (2019), I find that forecast averaging performs better than model selection. In addition, I find that forecast averages that include CAVS forecasts outperform those that do not. Overall, the results support the view that exploiting nonlinearities in macroeconomics series improves forecast accuracy (see also Marcellino, 2002; Clements *et al.*, 2004; Terasvirta, 2006).

It is important to emphasize that this is not the first paper to consider the decomposition $Y_t = |Y_t|\text{sign}(Y_t)$. When applied to stock returns, Diebold and Christoffersen (2006) argue that both of the right-hand side components are predictable, yet their product is not. Anatolyev and Gerko (2005) apply this decomposition to develop a market timing test, whereas Rydberg and Shephard (2003) employ it to model the dynamics of trade-by-trade price movements in a market microstructure setting. Closely related to this work, Anatolyev and Gospodinov (2010) model excess returns by combining a multiplicative error model for the absolute values, a dynamic binary model for signs, and a copula for their interaction. This paper differs from Anatolyev and Gospodinov (2010) in that rather than modeling the joint likelihood of the decomposition model through the use of copulas, CAVS forecasts are defined as the solutions to two independent minimization problems — thus bypassing the need for a copula altogether. This paper is related to the macroeconomic forecasting literature, which includes work by Stock and Watson (2002, 2012),Kim and Swanson (2014), and Cheng and Hansen (2015), among others. Additionally, this work also relates to the literature on nonlinear forecasting methods in economics. This includes work by Clements *et al.* (2004), Terasvirta (2006), and White (2006), among others.

The remainder of this paper is structured as follows. Section 2 introduces the CAVS framework, Section 3 contains a simulation study, and Section 4 presents the empirical application. Concluding remarks follow in Section 5.

2 CAVS Forecasts

2.1 Notation and Definition

Let $\{Y_t\}$ denote a zero mean scalar time series of growth rates[2] and $\{X_t\}$ with $X_t \in \mathbb{R}^n$ a time series of $t-1$ measurable predictors. In this work, I consider a forecaster whose objective is to construct a forecast of Y_t given X_t that minimizes the MSE. It is well known (see Brockwell and Davis, 1991) that the MSE optimal forecast of Y_t given X_t is the conditional expectation

$$\mu(X_t) = \mathbb{E}\Big[Y_t \Big| X_t\Big] = \arg\min_{m \in \mathcal{M}} \mathbb{E}\Big[\big(Y_t - m(X_t)\big)^2\Big] \ ,$$

where $\mathcal{M}$ is the collection of measurable functions m of X_t having finite variance.

The methodology introduced in this paper builds on a decomposition of the conditional mean into the product of the conditional expectations of the absolute value and sign of the target variable. It is straightforward to verify that the identity $Y_t = |Y_t|\text{sign}(Y_t)$ implies that the conditional mean may be expressed as

$$\mu(X_t) = \mathbb{E}\Big[|Y_t|\Big|X_t\Big]\mathbb{E}\Big[\text{sign}(Y_t)\Big|X_t\Big] + \mathbb{C}\text{ov}\Big(|Y_t|, \text{sign}(Y_t)\Big|X_t\Big) \ . \tag{1}$$

The representation in (1) highlights that the optimal forecast for Y_t given X_t can be written as the product of MSE optimal absolute value and sign forecasts and a conditional covariance term. This decomposition motivates the introduction of CAVS forecasts, a nonlinear forecasting framework that exploits componentwise predictability in the absolute value and sign to approximate the conditional mean.

Formally, denoting by $\mathcal{C}_A$ and $\mathcal{C}_S$ collections of functions for the absolute value and the sign of Y_t, CAVS forecasts are defined as

$$\mu_{\text{CAVS}}(X_t) = m_A^*(X_t) m_S^*(X_t) + c(X_t) \ , \tag{2}$$

[2] I assume that $\mathbb{E}[Y_t^2] < \infty$ throughout.

where

$$m^*_A(X_t) = \arg\min_{m \in \mathcal{C}_A} \mathbb{E}\Big[\Big(|Y_t| - m(X_t)\Big)^2\Big] \ , \tag{3}$$

$$m^*_S(X_t) = \arg\min_{m \in \mathcal{C}_S} \mathbb{E}\Big[\Big(\text{sign}(Y_t) - m(X_t)\Big)^2\Big] \ , \tag{4}$$

$$c(X_t) = \mathbb{C}\text{ov}\Big(|Y_t|, \text{sign}(Y_t)\Big|X_t\Big) \ . \tag{5}$$

It is important to emphasize that CAVS forecasts do not generally coincide with the conditional mean. In particular, if $\mathcal{C}_A$ and $\mathcal{C}_S$ do not include the optimal forecasts $\mu_A(X_t) = \mathbb{E}\big[|Y_t|\big|X_t\big]$ and $\mu_S(X_t) = \mathbb{E}\big[\text{sign}(Y_t)\big|X_t\big]$ respectively, the CAVS forecast will differ from the conditional mean.

It is widely documented in the economic forecasting literature that linear models for the conditional mean may exhibit poor out-of-sample performance (Stock and Watson, 2007; Rossi, 2013). Nonlinear models are typically able to approximate arbitrary functions (White, 1990). This flexibility comes at the expense of increased computational complexity, heightened risks of overfitting, and difficulties of interpretation (White, 2006), and the evidence regarding the performance of nonlinear models in macroeconomic and financial time series is inconclusive (Clements *et al.*, 2004; Terasvirta, 2006). In contrast to general nonlinear models, the CAVS framework is designed to exploit two specific features that figure prominently in the macroeconomic and financial forecasting literature: volatility and sign predictability (see Diebold and Christoffersen, 2006).

The interpretation of the CAVS forecast is straightforward: it is the expected magnitude of the next realization weighted by the probability that it will be positive (or negative). In fact, the CAVS forecast is the MSE optimal forecast given knowledge of either the sign or absolute value of the next realization. Computationally, the construction of CAVS forecasts requires the estimation of models for the absolute value and sign of the target series. The results from the empirical application show that CAVS forecasts based on generalized linear models — which are simple to compute — add substantial flexibility when compared to linear models for the conditional mean.

To make the CAVS definition operational, the forecaster must specify appropriate functions for the absolute value, the sign, and the covariance term $c(X_t)$.

The absolute value of the target series may be modeled as generalized linear functions of the predictors. Alternatively, building on the empirical success of GARCH models (Engle, 1982; Brownlees *et al.*, 2011), absolute value forecasts may be constructed based on conditional volatility forecasts.

A number of possibilities are available to construct sign forecasts. First, one may employ standard generalized linear models, such as logit and probit, to forecast the signs. Alternatively, the machine learning literature has put forward a number of methods to forecast binary random variables and their associated probabilities. In the empirical application, generalized linear models as well as nonparametric methods are used to construct probability forecasts. Note that the sign forecast used in the construction of CAVS forecasts is based on the MSE (of signs) and hence is a probabilistic, rather than binary, forecast. The MSE is a natural choice in this setting as it elicits the (componentwise) conditional expectation. Alternative combinations of componentwise loss functions may be explored, particularly for conditional mean forecasting under different loss functions.

Finally, I remark that the $c(X_t)$ term is often negligible in practice. In fact, Proposition 3 shows that, for additive models with symmetric shocks, $c(X_t)$ is small relative to the irreducible MSE and may therefore be ignored for forecasting purposes. Alternatively, one may explicitly model the joint distribution of signs and absolute values by means of a copula function, as in Anatolyev and Gospodinov (2010).

2.2 Theoretical Results

This section provides a number of theoretical results for CAVS forecasts. Throughout this section, I express Y_t as the sum of the conditional mean and an unpredictable error term:

$$Y_t = \mu(X_t) + \sigma_u u_t \ , \ u_t \overset{i.i.d}{\sim} \mathcal{D}(0, 1) \ , \tag{6}$$

where $\mathbb{E}[Y_t] = 0$, $\mathbb{E}[Y_t^2] < \infty$ and $\mathcal{D}$ is a distribution with mean zero and unit variance. All proofs are presented in the Appendix.

Bounding the loss of CAVS forecasts. Forecasts of economic series are typically based on misspecified approximations to the conditional mean (see White, 2006). The following proposition provides an upper bound for the MSE of CAVS forecasts that depends the approximating properties of $\mathcal{C}_A$ and $\mathcal{C}_S$.

Proposition 1. *The MSE of the CAVS forecast given in* (2) *is such that*

$$\mathbb{E}\Big[\Big(Y_t - \mu_{CAVS}(X_t)\Big)^2\Big] \leq \sigma_u^2 + a_1\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big] \ ,$$

where $a_1 > 0$ is a constant that depends on $\mathbb{E}[Y_t^2]$.

Proposition 1 shows that the accuracy of CAVS forecasts depends on the accuracy of the approximations to $\mu_A(X_t)$ and $\mu_S(X_t)$. If $\mu_A(X_t) \in \mathcal{C}_A$ and $\mu_S(X_t) \in \mathcal{C}_S$, the CAVS forecast is equivalent to the conditional mean. A large body of research has documented the good forecasting performance of standard volatility models (see Andersen *et al.*, 2006; Brownlees *et al.*, 2011). This observation suggests that, for series with persistent volatility dynamics, $\mu_A(X_t)$ may be accurately approximated. In addition, Diebold and Christoffersen (2006) show that conditional volatility dynamics implies directional predictability, and Leung *et al.* (2000) document evidence of sign predictability in excess of that captured by linear models for the conditional mean. Taken together, these observations suggest that componentwise approximations to the conditional mean may be able to leverage volatility and sign predictability into accurate conditional mean forecasts.

Proposition 1 assumes that the forecaster knows $c(X_t)$, the conditional covariance of signs and absolute values. In practice, this is typically not the case. The forecaster may explicitly model this term. Alternatively, the next proposition describes the MSE of CAVS forecasts when $c(X_t)$ is unknown and set to 0 in (2).

Proposition 2. *The MSE of the forecast constructed as $m_A^*(X_t)m_S^*(X_t)$ is such that*

$$\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \le \sigma_u^2 + a_2\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big] + 2\mathbb{E}\Big[c(X_t)^2\Big] ,$$

where $a_2 > 0$ is a constant that depends on $\mathbb{E}[Y_t^2]$ and $c(X_t)$.

Proposition 2 shows that the excess risk, in MSE terms, incurred due to setting $c(X_t)$ to 0 in the construction of the CAVS forecast is additive and proportional to $\mathbb{E}[c(X_t)^2]$. Next, I show that $\mathbb{E}[c(X_t)^2]$ is small relative to σ_u^2, the irreducible uncertainty.

Proposition 3. *Considering the representation given in* (6) *and assuming $\mathcal{D}$ is symmetric about 0, then*

$$\mathbb{E}\Big[\Big(Y_t - \mu_A(X_t)\mu_S(X_t)\Big)^2\Big] \le (1+\gamma)\mathbb{E}\Big[\Big(Y_t - \mu(X_t)\Big)^2\Big] ,$$

with $0 < \gamma \le 1/2$. In addition, if $\mathcal{D}$ is the Gaussian distribution, then $\gamma \approx 0.04187$.

Proposition 3 provides an upper bound to the contribution of the conditional covariance term — the approximation error obtained from a componentwise approximation of $\mu(X_t)$ — to the overall MSE. Combining this result with Proposition 2 yields the following corollary.

Corollary 1. *Consider the representation given in* (6) *and assume $\mathcal{D}$ is symmetric about 0. Then, the MSE of the forecast constructed as $m_A^*(X_t)m_S^*(X_t)$ is such that*

$$\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \le 2\sigma_u^2 + a_2\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big] ,$$

where a_2 is the same as in Proposition 2.

Corollary 1 shows that the MSE of the CAVS forecast constructed by neglecting $c(X_t)$ is driven by the losses in sign and absolute value forecasting. Taken together, Propositions 1 – 3 highlight that CAVS forecasts may provide suitable approximations to the conditional mean, especially when the componentwise forecasts are accurate approximations of their targets. Next, I present a DGP in which accurate approximations are available for each of the components, yet linear models for the conditional mean are not able to exploit all available information.

Linear forecasts in a nonlinear setting. Consider the following DGP:

$$Y_t = \big|X_t'(\beta + \delta e_1)\big|\mathrm{sign}(X_t'\beta) + u_t, \qquad u_t \overset{iid}{\sim} \mathcal{D}(0, \sigma_u^2)\ , \tag{7}$$

where $X = (x_{1,t}, x_{2\,t})'$, $e_1 = (1, 0)$, and $\delta \in \mathbb{R}$. This choice of functional form allows x_1 to affect the sign and absolute value of the target variable differently. Note that linear models are obtained by setting $\delta = 0$. In contrast, as $|\delta|$ increases, the resulting model exhibits weak linear conditional mean predictability with strong sign and absolute value predictability. In addition, $\kappa = \frac{\beta_1+\delta}{\beta_1}$ controls the size of δ relative to β_1, thus parameterizing departures from linearity in this model. Both the absolute value and the sign of Y_t may be accurately approximated by generalized linear models, yet linear models for the conditional mean will have poor performance as $|\kappa - 1|$ increases.

Proposition 4. *Consider the model given in* (7) *with* $\beta_1 = \beta_2 = \beta$, $cov(x_{1\,t}, x_{2\,t}) = 0$, *and* $var(x_{i\,t}) = \sigma_i^2$ *for* $i = 1, 2$. *The MSE of the best linear predictor is given by*

$$\begin{aligned}
\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] = {} & \sigma_u^2 + \beta^2\kappa^2\Big(\sigma_1^2 - \frac{\mathbb{E}[x_{1\,t}^2 S(\kappa)]^2}{\sigma_1^2} - \frac{\mathbb{E}[x_{1\,t}x_{2\,t}S(\kappa)]^2}{\sigma_2^2}\Big) \\
& - 2\beta^2\kappa\mathbb{E}[x_{1\,t}x_{2\,t}S(\kappa)]\Big(\frac{\mathbb{E}[x_{1\,t}^2 S(\kappa)]}{\sigma_1^2} + \frac{\mathbb{E}[x_{2\,t}^2 S(\kappa)]}{\sigma_2^2}\Big) \\
& + \beta^2\Big(\sigma_2^2 - \frac{\mathbb{E}[x_{2\,t}^2 S(\kappa)]^2}{\sigma_2^2} - \frac{\mathbb{E}[x_{1\,t}x_{2\,t}S(\kappa)]^2}{\sigma_1^2}\Big)\ ,
\end{aligned}$$

where $\beta^* = \arg\min_\beta \mathbb{E}\Big[\Big(Y_t - X_t'\beta\Big)^2\Big]$ *and* $S(\kappa) = \mathrm{sign}(x_{1\,t}^2\kappa + x_{1\,t}x_{2\,t}(1+\kappa) + x_{2\,t}^2)$.

Proposition 4 provides an exact formula for the MSE of the best linear predictor in a model where variables impact the sign and absolute value of the target differently. Although no closed formula solution is available for the expectations involved, they can be evaluated by numerical methods. In addition, the order of the MSE of the best linear predictor can be computed for large κ.

Corollary 2. *There exists a κ_0 such that for all $\kappa > \kappa_0$,*

$$\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] \geq a_3\kappa^2 \ ,$$

where $a_3 > 0$ is a constant that depends on $\mathbb{E}[x_{1t}^2 S(\kappa)], \mathbb{E}[x_{2t}^2 S(\kappa)]$ and $\mathbb{E}[x_{1t}x_{2t}S(\kappa)]$.

Corollary 2 shows that, for large enough κ, the MSE of the best linear predictor grows with κ^2. The next section contains detailed simulation results for this model.

3 Simulation Study

In this section, I carry out a simulation study to numerically evaluate the performance of CAVS forecasts. I consider two simulation settings. In addition to the model given in (7), I simulate from the following model:

$$Y_t = \big|X_t'\beta\big|\text{sign}(X_t'(\beta + \delta e_1)) + u_t, \qquad u_t \overset{iid}{\sim} N(0,1) \ . \tag{8}$$

Both models are nonlinear in x_1 yet approximately linear in x_2, and nest a linear model when $\kappa = 1$. The nonlinearity in model (7) arises due to increases in the coefficient attached to x_1 in the absolute value component. This implies that the variance of the target variable increases with $\kappa = \frac{\beta_1+\delta}{\beta_1}$. In contrast, the nonlinearity in model (8) arises on the sign component, and hence does not influence the variance of the target variable. In both models, departures from linearity are obtained by varying κ. Each predictor is an i.i.d draw from a standard Gaussian distribution. Three forecasting strategies are compared. First, I consider the linear forecast obtained by ordinary least squares regression of Y_t on X_t. Second, I consider the CAVS forecast in (2) constructed by setting $c(X_t)$ to 0. Finally, I consider a CAVS forecast where $c(X_t) = 0$ and sign forecasts are obtained from a probit regression of $\mathbb{1}\{Y_t > 0\}$ on X_t.

FIGURE 1 ABOUT HERE

Figure 1 illustrates the MSE ratios of each forecasting strategy relative to the best linear predictor for the two simulation settings considered. Values below 1 indicate that the MSE of a given strategy is smaller than that of the linear forecast. Solid lines represent the CAVS forecasts constructed by setting $c(X_t) = 0$, whereas dashed lines represent the CAVS forecasts constructed with $c(X_t) = 0$ and sign forecasts based on a probit regression. The results from both simulation settings are similar. For large values of κ or for $\kappa \approx 0$, CAVS forecasts substantially outperform the linear forecast. In particular, for large κ, CAVS forecasts display the strongest performance gains on the DGP in which nonlinearities arise in the absolute value of the target series. In contrast, for $\kappa \approx 0$, CAVS forecasts display the largest gains on the DGP in which nonlinearities arise in the sign of the target series. For $\kappa \approx 1$, the benefits of exploiting the existing nonlinearities are outperformed by the cost of neglecting the conditional covariance term by setting $c(X_t)$ to 0, and linear models that ignore the nonlinearities will perform better than a CAVS forecast that does not model the conditional covariance term. Overall, the simulations highlight that deviations from linearity — such as different predictors impacting different components of the target series — generate sizable gains for CAVS forecasts.

4 Empirical Application

I employ the framework introduced in this work to forecast the components of the FRED-MD dataset (McCracken and Ng, 2016), which consists of 128 monthly financial and macroeconomic series. As in McCracken and Ng (2016), series are arranged in eight groups: (1) output and income; (2) labor market; (3) housing; (4) consumption, orders and inventories; (5) money and credit; (6) interest and exchange rates; (7) prices; and (8) stock market. Detailed results are presented for a subset of the selected series considered in Kim and Swanson (2014). Additionally, I also report results for the Fed Funds rate, which may be particularly of interest to private sector forecasters.

All series are transformed as suggested in McCracken and Ng (2016). In particular, some series are not expressed in growth rates.[3] Series that do not change signs over the whole sample are excluded from the forecast comparison, but are kept as predictors

[3]For example, series in the housing group are the logs of housing starts, and hence are always positive.

to forecast the remaining variables. Additionally, the suggested transformations assume price series are integrated of order two, implying that the transformed series are twice differenced. This may hamper sign predictability. Details of the data and their transformations can be found in McCracken and Ng (2016). Five series with more than 10 missing observations are dropped from the dataset.[4] The remaining missing values are replaced by the unconditional mean of the series computed over the whole sample. After all transformations, the panel consists of data for 123 series from March 1959 to January 2020, corresponding to 731 months.

There is a large literature on macroeconomic forecasting. Stock and Watson (2002, 2007, 2012) introduce diffusion index models and document the good performance of PCR for macroeconomic forecasting. Kim and Swanson (2014) compare the forecasting performance of a variety of dimension reduction techniques, and find that PCR is improved when combined with other shrinkage-based techniques. Cheng and Hansen (2015) consider forecast averaging methods, and find that model averaging improves over PCR at longer horizons and performs on par with other shrinkage methods at shorter forecast horizons. Most of these studies focus on linear models and emphasize the issues related to the dimensionality of the data.

In contrast, Stock and Watson (1999), White and Swanson (1997), Marcellino (2002), and Bai and Ng (2008) examine nonlinear forecasts of economic variables. There is mixed evidence regarding the performance of nonlinear models. Stock and Watson (1999) find limited evidence of univariate nonlinear models improving upon linear forecasts, and that the best nonlinear models are typically tightly parameterized. White and Swanson (1997) find that flexible nonlinear models are particularly suitable for forecasting at longer horizons and document evidence of nonlinearities in nine macroeconomic series. Marcellino (2002) documents substantial evidence of exploitable nonlinearities in a number of macroeconomic series in the European Monetary Union, and Bai and Ng (2008) find that allowing for nonlinearities in the construction of the factors may improve the

[4]These are: New orders for consumer goods (ACOGNO), New orders for nondefense capital goods (ANDENOx), Trade weighted U.S. Dollar Index: Major currencies (TWEXMMTH), consumer sentiment index (UMCSENTx), and the VXO volatility index (VXOCLSx).

accuracy of PCR forecasts. See Clements *et al.* (2004), White (2006), and Terasvirta (2006) for extended discussions on forecasting economic variables with nonlinear models.

4.1 Forecasting Methodology

I carry out a pseudo out-of-sample forecasting exercise on the FRED-MD dataset. I construct and evaluate conditional mean, absolute value and sign forecasts for $h = 1, 3, 6$ and 12 months ahead for 113 series.[5] Following Stock and Watson (2012) and Boot and Nibbering (2019) all models include 4 lags of the dependent variable on the predictive regression, and dimension reduction techniques are applied to the remaining predictors. Forecasts are produced recursively starting from January 1985 until the end of the sample.

4.1.1 Linear Forecasts

Principal components regression For each series in the panel and at each out-of-sample period T, PCR forecasts are constructed using

$$\widehat{Y}_{T+h}^{\mathrm{PCR}}(r) = \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + \sum_{i=1}^{r} \widehat{\lambda}_i \widehat{F}_{iT} \ ,$$

where $\widehat{F}_{iT}$ is the i-th principal component of $\{X_t\}_{t=1}^{T}$, and X_t denotes a vector of predictors. The predictive regression is estimated by ordinary least squares, and forecasts are constructed for $r = 1, \ldots, 100$. Following Boot and Nibbering (2019), for each out-of-sample period T, I choose,

$$r^* = \arg \min_{1,\ldots,100} \sum_{t=T-60}^{T} \left(Y_t - \widehat{Y}_t(r) \right)^2 ,$$

and take the PCR forecast to be $\widehat{Y}_{T+h}^{\mathrm{PCR}}(r^*)$.

[5]These are the 123 series considered with the exception of those that do not change sign in the evaluation period.

Penalized regression For each series in the panel and at each out-of-sample period T, penalized forecasts are constructed using the model:

$$\widehat{Y}_{T+h}(\lambda_l, l) = \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + X_T' \widehat{\beta} \ ,$$

where

$$(\widehat{\rho}_1, \ldots, \widehat{\rho}_4, \widehat{\beta}) = \arg\min \frac{1}{T} \sum_{t=4}^{T} \Big(Y_{t+h} - \sum_{p=1}^{4} \widehat{\rho}_p Y_{t-p+1} + X_t' \widehat{\beta} \Big)^2 + \lambda g_l(\beta)$$

and where λ is a tuning parameter, $g_1(\beta) = 2\sum_{i=1}^{n} |\beta_i|$ is the LASSO penalty, and $g_2(\beta) = \sum_{i=1}^{n} \beta_i^2$ is the ridge penalty. Following Boot and Nibbering (2019), LASSO forecasts are constructed for $\log(\lambda) \in \{-30, -29.7, \ldots, 0\}$, whereas ridge forecasts are constructed for $\log(\lambda) \in \{-15, -14.7, \ldots, 15\}$. I then select, for each out-of-sample period T and choice of l,

$$\lambda_l^* = \arg\min_{\lambda} \sum_{t=T-60}^{T} \Big(Y_t - \widehat{Y}_t(\lambda, l) \Big)^2 \ ,$$

and the LASSO forecasts are given by $\widehat{Y}_{T+h}^{\text{LASSO}}(\lambda_1^*, 1)$, and ridge forecasts by $\widehat{Y}_{T+h}^{\text{Ridge}}(\lambda_2^*, 2)$.

4.1.2 CAVS Forecasts

Absolute value forecasts Absolute value forecasts are obtained using the same baseline methods employed to construct linear conditional mean forecasts, with minor adjustments. In particular, the target variable is $|Y_{T+h}|$ rather than Y_{T+h}, and a larger set of predictors $W_t = (X_t', |X_t'|)$ is considered.[6] In addition to PCR, ridge, and LASSO, I consider absolute value forecasts implied by an AR(4)-GARCH(1,1) model:

$$\begin{aligned} Y_{T+1} &= \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + \sigma_{T+1|T} u_{T+1}, \qquad u_{T+1} \overset{i.i.d}{\sim} \mathcal{D}(0,1) \\ \sigma_{T+1|T}^2 &= \omega + \alpha \Big(Y_T - \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p} \Big)^2 + \beta \sigma_T^2 \ , \end{aligned}$$

[6]The absolute value is taken coordinate wise.

where $\mathcal{D}$ is a distribution with zero mean and unit variance. Forecasts are constructed as:

$$|Y_{T+1}|^{\mathrm{G}} = \frac{1}{B} \sum_{b=1}^{B} \left| \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + \sigma_{T+1} \widehat{u}_{T+1}^{b} \right| ,$$

where $\widehat{u}^b$ is a draw from the GARCH filtered residuals (Barone-Adesi *et al.*, 2008). Forecasts for $h > 1$ are obtained by iterating the model forwards (Brownlees and Souza, 2020).

Sign forecasts Similarly to absolute value forecasts, sign forecasts are constructed employing the baseline methods to forecast the signs of the target variable, with a few adjustments. First, the target variable is $Z_{T+h} = \mathbb{1}\{Y_{T+h} > 0\}$. For Sign-PCR, a logit regression based on the same factors $(\widehat{F}_T)$ created to forecast Y_{T+h} is estimated. Sign-Ridge and Sign-LASSO are estimated by penalized maximum likelihood.[7] In addition to PCR, ridge and LASSO, I consider a number of machine learning algorithms to construct sign forecasts. In particular, I consider random forests (Breiman, 2001), AdaBoost (Freund and Schapire, 1995), k-nearest neighbors (Devroye *et al.*, 1996) and neural networks (White, 2006). All tuning parameters are selected on the basis of past predictive performance.

Baseline CAVS forecasts CAVS forecasts are constructed as the product of sign and absolute value forecasts, and a number of specifications may be entertained. I consider a set of baseline CAVS specifications where the same method is used to forecast signs and absolute values, and denote them as CAVS-PCR, CAVS-Ridge, and CAVS-LASSO.

4.1.3 Forecast combinations

In addition to baseline linear and CAVS models, I explore the performance of forecast combinations. First, I consider model selection based on past predictive performance for both CAVS and linear forecasts. The linear forecast is based on the model (PCR, ridge,

[7]The grid for λ is constructed based on the whole sample and is given by $\epsilon, \ldots, \lambda_{\max}$, where $\epsilon \approx 0$ and λ_{max} is the λ value such that all coefficients in the model are zero (see Friedman *et al.*, 2010).

or LASSO) that minimizes the MSE over the last 60 months. In contrast, model selection in the CAVS framework amounts to choosing the best forecasting strategy for each component. Component-specific models are chosen by their predictive performance in forecasting the appropriate target variable, where performance is measured by the MSE. Second, I consider equally weighted forecast combinations based exclusively on linear or CAVS forecasts. I consider 12 CAVS specifications to average over: the permutations of PCR, Ridge and LASSO, in addition to GARCH-based absolute value forecasts. In contrast, there are 3 linear specifications to average over. Finally, I consider a hybrid forecast given by the average of the two previously constructed equal weighted forecasts.

Forecast evaluation As is standard in the forecasting literature, I evaluate conditional mean forecasts by their pseudo out-of-sample MSE, defined as

$$MSE_{im} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{Y}_{it}(m) - Y_{it}\right)^2 ,$$

where $i = 1, \ldots, n$ denotes the target series, T the number of pseudo out-of-sample observations, and $\widehat{Y}_{it}(m)$ is the forecast for Y_{it} based on method m. Diebold-Mariano tests (Diebold and Mariano, 1995) of superior predictive ability are carried out to assess whether strategies improve forecast accuracy relative to the PCR benchmark. In particular, denoting by $\epsilon_{t+h}(m)$ model m's prediction error, the null hypothesis of the DM test considered is $H_0 : \mathbb{E}[\epsilon^2_{t+h}(m)] < \mathbb{E}[\epsilon^2_{t+h}(\text{PCR})]$. The test statistic is constructed as the sample analog of $\mathbb{E}[\epsilon^2_{t+h}(m)] - \mathbb{E}[\epsilon^2_{t+h}(\text{PCR})]$, scaled by a heteroskedasticity and autocorrelation robust estimator of its standard deviation.

Directional forecasts are evaluated by the proportion of incorrect sign forecasts, defined as

$$DL_{im} = 1 - \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}\{\text{sign}(Y_{it}) = \text{sign}(\widehat{Y}_{it}(m))\} .$$

Assuming $H_{it}(m) = \mathbb{1}\{\text{sign}(Y_{it}) = \text{sign}(\widehat{Y}_{it}(m))\} \sim Ber(p_m)$, interest lies in verifying whether $p_m = p_{PCR}$ for the remaining forecasting strategies m. In other words, I test whether model m has the same probability of correctly classifying the sign of the next realization relative to the PCR benchmark. Following Christoffersen (1998), a likelihood ratio test is conducted to test the null hypothesis of $p_m = p_{PCR}$.

In addition to comparing forecasts across models, forecasts are compared against a benchmark constructed from the unconditional distribution of the target variable. If a model outperforms the unconditional benchmark, we say there is evidence of predictability, that is, the conditioning set improves forecast accuracy. Conditional mean predictability is assessed by DM tests of each strategy relative to the recursively estimated unconditional mean of each variable.

4.2 Empirical Results

4.2.1 Results for Selected Series

TABLE 1 ABOUT HERE

Table 1 reports the ratio of the MSE of each forecasting strategy relative to PCR for each selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability are carried out and stars denote significance levels. CAVS-based forecasts are the MSE best for the majority of series across all forecast horizons considered, substantially outperforming linear forecasts for the Federal funds, industrial production, nonfarm payroll employment, and the S&P 500 uniformly across forecast horizons. In particular, for the Federal funds, the best CAVS specification displays MSE reductions of about 15, 18, 6, and 3% relative to the best linear model at $h = 1, 3, 6$, and 12 months ahead, respectively. CAVS-Ridge is the best performing specification, selected as the MSE best across all series and forecast horizon combinations 25% of the time, followed by CAVS-LASSO, PCR, and CAVS-PCR, which are the MSE best 20, 17, and 15% of the time, respectively. For $h = 1$ month ahead forecasting, Ridge is the best performing linear model for 5 out of the 10 selected series, followed by LASSO and PCR, the best performing linear models for 4 and 1 series, respectively. CAVS-Ridge is the best performing CAVS specification for 5 out of the 10 selected series, followed by CAVS-PCR and CAVS-LASSO, the best performing CAVS specifications for 3 and 2 series, respectively. This ranking remains largely unchanged for $h = 3$ and $h = 6$ months ahead forecasts. For $h = 12$ months ahead, Ridge is the best performing linear model for 7 out of 10 series, followed by PCR, the best performing model for 3 series. LASSO is not the best linear model for any series. Among CAVS specifications, CAVS-Ridge dominates CAVS-PCR and is the best CAVS specification for 7 out of 10 series. CAVS-LASSO is the best performing CAVS specification for the remaining 3 series.

Overall, the good performance of the baseline CAVS forecasts highlights that exploiting directional and volatility predictability yields more accurate macroeconomic forecasts for the selected series. In particular, CAVS-Ridge forecasts are, on average, 6.5% more accurate than PCR forecasts, displaying the largest gains among all the baseline models considered.

4.2.2 Results for all FRED-MD components

Next, I provide a comparison of CAVS-Ridge and PCR for all FRED-MD components.

FIGURE 2 ABOUT HERE

Figure 2 reports the ratios of the MSE of CAVS-Ridge relative to PCR for all FRED-MD components, sorted by groups. Values below 1 indicate that CAVS-Ridge forecasts outperforms PCR. Colors indicate significance at the 10% level (gray), 5% level (dark-gray), or 1% level (black), based on a DM test of superior predictive ability. A number of findings emerge from inspection of Figure 2. First, CAVS-Ridge outperforms PCR for the majority of series and across all horizons, with particularly strong performance for series that exhibit persistent volatility dynamics, such as interest rates, output, and stocks series. Second, as the forecast horizon increases, CAVS-Ridge performance gains become widespread, outperforming PCR for 53, 80, 67, and 75% of series for $h = 1, 3, 6,$ and 12 months ahead, respectively. Finally, CAVS-Ridge — and CAVS forecasts in general — display poor performance for series that are not expressed in growth rates. In particular, none of the series that are in the top 90% quantile of MSE ratios — *i.e.*, series for which CAVS provides MSE increases greater than 12.6% — are expressed in growth rates. This is the case for all price series, most money, and a few labor series, all of which are twice differenced. Additionally, some series in the interest and exchange rate groups are expressed as spreads. Because spreads rarely change signs, CAVS forecasts generally display poor performance for these series — in stark contrast to its good performance for their growth rates counterparts. For example, CAVS-Ridge provides 33% more accurate forecasts of the first differences of the 3-month treasury bill when compared to PCR. By contrast, PCR provides 5% more accurate forecasts of the spread constructed as the 3-month treasury bill minus the federal funds rate relative to CAVS-Ridge. Overall, the results highlight that CAVS-Ridge compares favorably to PCR, particularly for series that display persistent volatility dynamics.

4.2.3 Componentwise Forecast Accuracy

Directional forecasts Directional forecasts of macroeconomic series are objects of interest in their own right (see Pesaran and Timmerman, 1992; Sinclair *et al.*, 2010, among others). I compare the performance of directional forecasts implied by CAVS and linear forecasts for the selected series.

TABLE 2 ABOUT HERE

Table 2 reports the directional loss ratio of each forecasting strategy relative to the sign of the linear PCR forecast for each selected series and forecast horizon. Note that signs of CAVS forecasts directly target the sign of the series considered by construction. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. Likelihood ratio tests of superior predictive ability relative to the linear PCR benchmark, as described in Section 4.1, are carried out and stars denote significance levels. CAVS-based directional forecasts outperform their linear counterparts for the majority of series and across all horizons considered. In particular, CAVS-based directional forecasts are more accurate for the four series for which CAVS display the largest gains relative to linear forecasts. For most of the remaining series, CAVS-based directional forecasts perform on par with their linear counterparts. In particular, CAVS directional forecasts are outperformed by linear forecasts for the CPI and M2 series. This finding highlights that twice differencing may hamper sign predictability, particularly at short forecast horizons. Moreover, CAVS-based directional forecasts provide substantial improvements relative to their linear counterparts at intermediate forecast horizons. The performance of all CAVS specifications is similar. CAVS-Ridge is the best directional forecasting strategy for the most series across all horizons, but performance gains relative to CAVS-LASSO and CAVS-PCR are modest. Among directional forecasts based on linear models, Ridge and LASSO perform similarly at $h = 1$ month ahead forecasting. PCR is the best performing linear model for horizons greater than 1 month ahead. Overall, CAVS-based directional forecasts are more accurate than their linear counterparts, highlighting that there is sign predictability in excess of that implied by linear models for the conditional mean. In particular, the results show that directly targeting the sign yields more accurate directional forecasts than forecasts based on the sign of conditional mean forecasts.

Absolute value forecasts The performance of CAVS forecasts hinges on sign and absolute value predictability. This section reports the results for the performance of absolute value forecasts for the selected series.

TABLE 3 ABOUT HERE

Table 3 reports, for each forecasting strategy, selected series, and forecast horizon, the MSE of absolute value forecasts relative to a PCR benchmark. I report results for the absolute values of linear forecasts, as well as absolute value forecasts that directly target the absolute values of the series considered. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability are carried out and stars denote significance levels. Absolute value forecasts based on a standard AR(4)-GARCH(1,1) are selected as the MSE best forecast for the majority of series across all forecast horizons. For $h = 1$ month ahead, GARCH-based forecasts exhibit sizable performance gains relative to linear models for the absolute value. This finding highlights that standard time series models are able to accurately model conditional volatility dynamics for macroeconomic series, and that the added value of exogenous predictors is modest (see also Brownlees and Souza, 2020). As the forecast horizon increases, all models that target the absolute value perform similarly, with GARCH-based forecasts modestly outperforming their linear counterparts. Overall, Tables 2 and 3 show that there is componentwise predictability in excess of that implied by linear models.

4.2.4 Additional Results

Alternative CAVS specifications I explore whether alternative CAVS specifications can improve forecasting performance relative to the baseline CAVS models considered. In particular, given the good performance of the GARCH(1,1) in forecasting absolute values, I consider 4 absolute value forecasting strategies (PCR, ridge, LASSO and GARCH) and 3 sign forecasting strategies (PCR, ridge and LASSO), leading to 12 possible combinations of signs and absolute values.

TABLE 4 ABOUT HERE

Table 4 reports the MSE of CAVS-based forecasts relative to that of PCR, for each of the selected series and forecast horizons. The first and second rows denote the absolute value and sign forecasting strategy, respectively. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability relative to PCR are carried out, and stars denote significance levels. In line with Tables 2 and 3 , CAVS forecasts based on the standard GARCH (the best performing absolute value model), and ridge (the best performing sign model), are the MSE best for the most series and forecast horizon combinations considered. However, performance across all CAVS models is similar, suggesting that standard CAVS specifications are sufficient to exploit the relevant nonlinearities.

Machine learning sign forecasts. I consider whether CAVS forecasts can be improved upon by employing machine learning algorithms to forecast the signs.

TABLE 5 ABOUT HERE

Table 5 displays results for selected series. For each series and forecast horizon, I report the ratios of the MSE of the CAVS specification considered relative to PCR forecasts. Best performing methods according to each criteria are highlighted in boldface, and stars represent significance according to a DM test, evaluated at the 5% significance level. The most successful machine learning algorithm is random forests. CAVS forecasts where the sign forecasts are constructed from random forests outperform baseline CAVS forecasts for 2 out of the 10 series considered. Perhaps surprisingly, no other machine learning algorithm outperforms the baseline CAVS forecasts. It is important to emphasize that I consider standard implementations of the machine learning algorithms. It is well-known that training machine learning models requires a degree of experimentation (White, 2006). This implies, in particular, that careful tuning of each model's parameters could improve accuracy.

Models against an unconditional benchmark. I compare models in terms of their ability to outperform forecasts based on the recursively estimated unconditional mean.

TABLE 6 ABOUT HERE

Table 6 reports, for each FRED-MD group and forecast horizon, the percentage of series in each group for which each strategy outperforms the unconditional mean forecasts according to a DM test at the 5% significance level. Best performing methods are highlighted in boldface. For $h = 1$ month ahead, both CAVS-based and linear forecasts outperform the unconditional mean benchmark for the majority of the series. In particular, CAVS-based forecasts outperform the unconditional mean benchmark more often than linear forecasts for labor, money, and interest rates series. In contrast, linear forecasts outperform the unconditional mean benchmark more often than CAVS-based forecasts for consumption and prices series. The performance of all models is similar in the remaining series. For $h > 1$ months ahead, CAVS-based forecasts outperform the unconditional mean more often than linear models for nearly all groups across all horizons. In particular, at $h = 6$ months ahead, CAVS-based forecasts outperform the unconditional mean for up to 40% of the Output series, in contrast to linear models, which outperform the unconditional mean for up to 25% of the series. Overall, CAVS-based forecasts outperform a unconditional mean benchmark for more series and across longer forecast horizons than linear models.

Model selection and forecast averaging. I consider the performance of forecast combinations, as described in Section 4.1.

TABLE 7 ABOUT HERE

Table 7 reports the MSE of each forecast combination strategy relative to that of PCR, for each selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability relative to PCR are carried out, and stars denote significance levels. Forecasts constructed by model averaging generally outperform those constructed by model selection for both linear and CAVS forecasts. Additionally, linear model averaging significantly outperforms PCR for nearly all series at all horizons considered. Moreover, forecast combinations that include CAVS forecasts outperform those based exclusively on linear models for the majority of series considered, and are particularly well-suited for intermediate forecast horizons. Finally, simple forecast combinations of linear and nonlinear models are a competitive strategy. Overall, the findings reported in Table 7 are in line with those reported in Table 1, suggesting that series characteristics, rather than specific modelling choices, are the determinants of whether CAVS will perform favorably compared to linear forecasts.

5 Concluding Remarks

This paper introduces CAVS forecasts, a nonlinear framework that combines forecasts of the sign and absolute value of a time series into conditional mean forecasts. In contrast to linear models, in which variables that affect the mean of the target variable must affect both its sign and absolute value, the proposed framework allows different predictors to affect either the sign, the absolute value, or both.

I provide a number of theoretical results for CAVS forecasts. First, I show that the conditional mean can be written as the product of MSE optimal forecasts of signs and absolute values and a covariance term that can be explicitly modeled. If the underlying DGP is additive with symmetric shocks, the covariance term is small relative to the variance of the shocks and therefore may be ignored for forecasting purposes. Second, I show that the performance of CAVS forecasts hinges on the ability to accurately forecast each of the components of the target series. This result highlights that CAVS-based forecasts are particularly suited for series that exhibit persistent volatility dynamics, and hence absolute value predictability. Third, I study a nonlinear DGP in which variables may affect signs and absolute values differently, and show that the MSE of the best linear predictor increases quadratically with the degree of nonlinearity.

The proposed methodology is applied to forecast each of the components of the FRED-MD dataset. I find that CAVS forecasts substantially outperform linear forecasts in series that exhibit strong conditional volatility dynamics, such as Output and Interest Rate series, and the performance gains remain sizable across forecast horizons. Moreover, I find that CAVS-based directional forecasts outperform linear forecasts for the majority of the selected series considered, across all horizons. Additionally, I document that CAVS forecasts outperform the recursively estimated unconditional mean benchmark for more series and across longer horizons than linear forecasts. Finally, I find that forecast combinations that include CAVS forecasts outperform those based exclusively on linear models for the majority of series and across all horizons considered. Overall, the empirical application highlights that exploiting directional and volatility predictability improves forecast accuracy in macroeconomic series.

References

Anatolyev, S. and Gerko, A. (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics*, **23**(4), 455–461.

Anatolyev, S. and Gospodinov, N. (2010). Modelling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, **28**(2), 232–245.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. Handbook of Economic Forecasting.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, **146**, 304–317.

Barone-Adesi, G., Engle, R. F., and Mancini, L. (2008). A garch option pricing model with filtered historical simulation. *Review of Financial Studies*, **21**(3), 1223–1258.

Boot, T. and Nibbering, D. (2019). Forecasting using random subspace methods. *Journal of Econometrics*, **209**(2), 391 – 406.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5–32.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theorey and Methods*. Springer.

Brownlees, C. T. and Souza, A. B. (2020). Backtesting global growth-at-risk. *Forthcoming Journal of Monetary Economics*.

Brownlees, C. T., Engle, R. F., and Kelly, B. (2011). A practical guide to volatility forecasting through calm and storm. *The Journal of Risk*, **14**(2), 3–22.

Cheng, X. and Hansen, B. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, **186**(2), 280–293.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, **39**(4), 841–862.

Clements, M. P., Franses, P. H., and Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, **20**, 169–183.

Devroye, L., Laszlo, G., and Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. Springer.

Diebold, F. X. and Christoffersen, P. F. (2006). Financial asset returns, direction-of-change forecasting and volatility dynamics. *Management Science*, **52**, 1273–1287.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economics Statistics*, **13**, 253–263.

Diebold, F. X. and Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*, **35**, 1679–1691.

Diebold, F. X., Christoffersen, P. F., Mariano, R. S., Tay, A. S., and Tse, Y. K. (2007). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *Journal of Financial Forecasting*, **1**(2), 1–22.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, **50**(4), 987–1007.

Freund, Y. and Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *European Conference on computation learning theory*, pages 23–37.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**(1), 1–22.

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

Kim, H. H. and Swanson, N. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, **178**(P2), 352–367.

Leung, M. T., Daouk, H., and Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecsating*, **16**(2), 173–190.

Marcellino, M. (2002). Instability and non-linearity in the emu. *Centre for Economic Policy Research*, **Discussion paper No. 3312**.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, **34**(4), 574–589.

Ng, S. (2013). Variable selection in predictive regressions. Handbook of Economic Forecasting.

Nyberg, H. (2011). Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting*, (27), 561–578.

Pesaran, H. M. and Timmerman, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economics Statistics*, **10**, 461–465.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, **51**(4), 1063–1119.

Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, **1**(1), 2–25.

Sinclair, T. M., Stekler, H. O., and Kitzinger, L. (2010). Directional forecasts of gdp and inflation: a joint evaluation with an application to federal reserve predictions. *Applied Economics*, **42**(18), 2289–2297.

Stock, J. and Watson, M. (1999). *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series*, pages 1–44. Oxford University Press, Oxford.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, **20**(2), 147–162.

Stock, J. H. and Watson, M. W. (2007). Why has u.s inflation become harder to forecast? *Journal of Money, Credt and Banking*, **39**(1), 4–33.

Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, **30**(4), 481–493.

Terasvirta, T. (2006). Forecasting economic variables with nonlinear models. Handbook of Economic Forecasting.

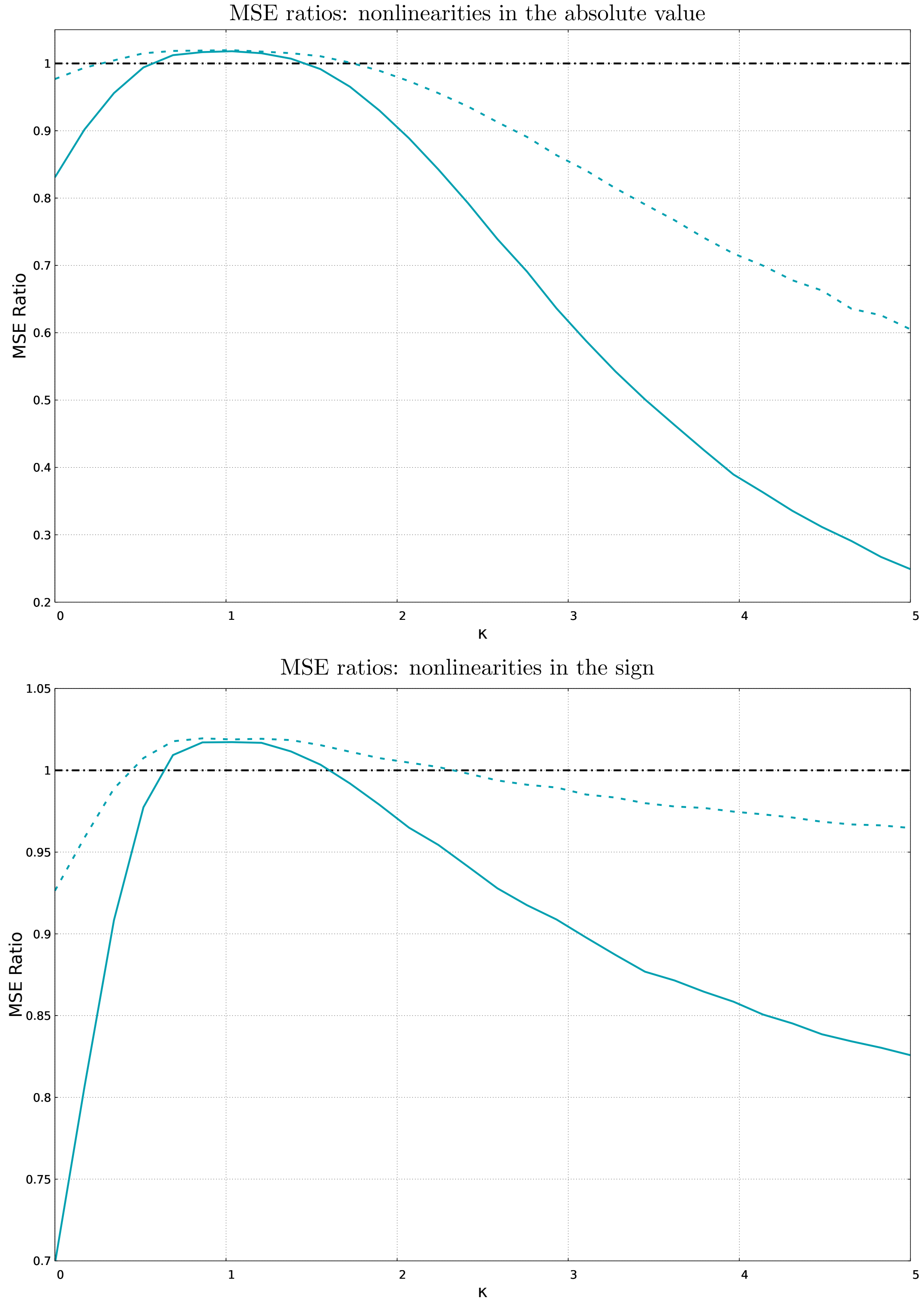
Timmerman, A. (2006). Forecast combinations. Handbook of Economic Forecasting.

White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, **3**(5), 169–183.

White, H. (2006). Approximate nonlinear forecasting. Handbook of Economic Forecasting.

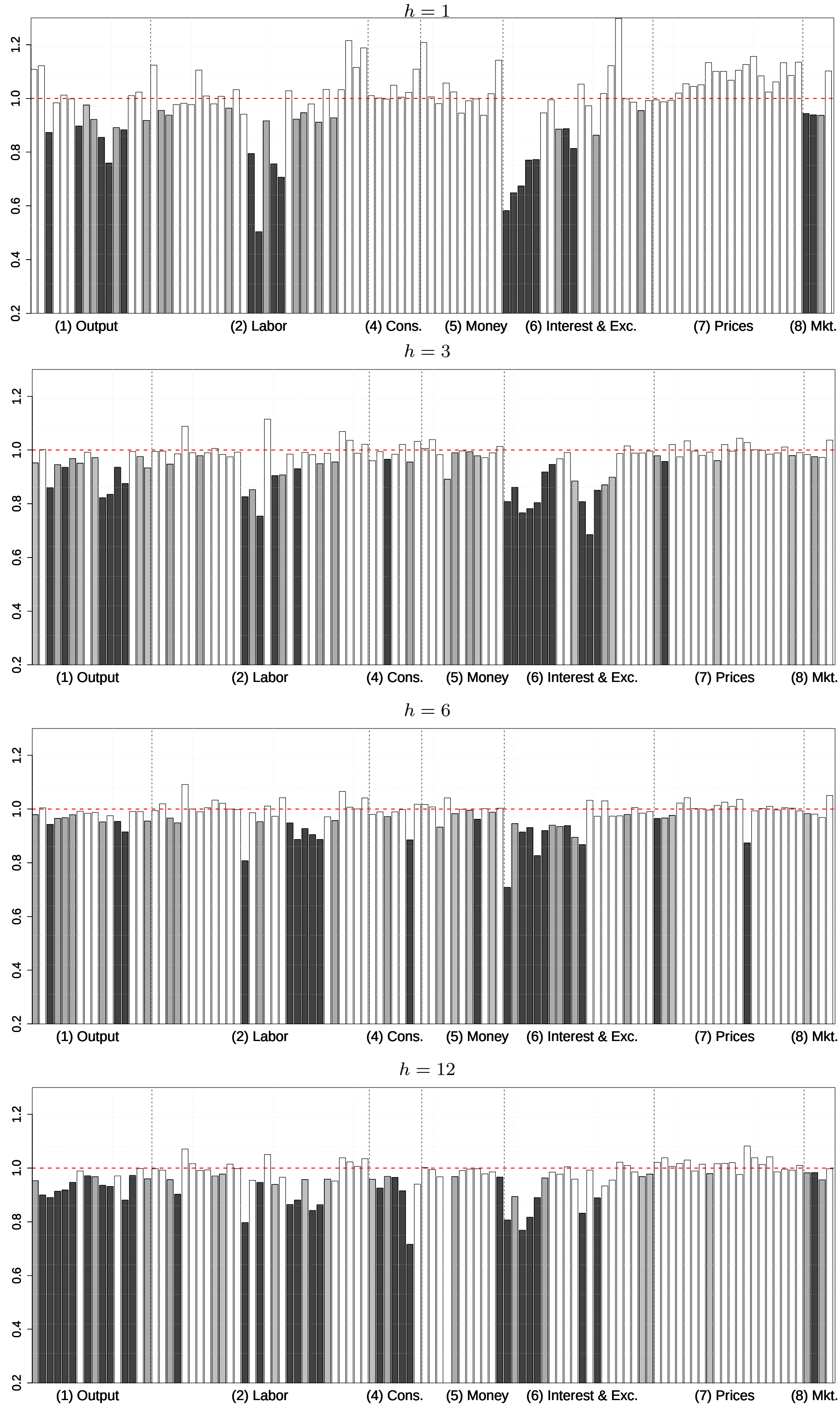
White, H. and Swanson, N. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, **13**, 439–461.

Figure 1: Simulation Study



This figure illustrates the findings of the simulation study. It depicts the MSE ratios of each strategy considered over the linear benchmark. The x-axis represents distance to linearity, which increases with $|\kappa - 1|$. Solid lines represent sign forecasts based on correctly specified models, whereas dashed lines represent sign forecasts based on a misspecified probit regression. Values below 1 indicate that the MSE of a given strategy is smaller than that of the linear forecast.

Figure 2: MSE ratios for all FRED-MD components



This figure reports the ratio of the MSE of the equally weighted CAVS forecast relative to that of the equally weighted linear forecast. Values below 1 indicate that CAVS forecasts outperform PCR forecasts. Colors indicate that CAVS combinations outperform linear combinations at the 10% level (gray), 5% level(dark-gray), or 1% level (black), based on a one-sided DM test.

Table 1: Linear and CAVS Forecasts

		Linear		*CAVS*		
h	Series	Ridge	LASSO	PCR	Ridge	LASSO
1	Fed Funds	0.725***	0.662***	0.692**	0.582***	**0.573*****
	Ind. Prod.	0.930***	0.982	0.891***	**0.874*****	0.894***
	Nonfarm Empl.	0.985	1.023	**0.922***	0.941	0.926*
	S&P 500	0.954***	0.958**	1.011	**0.944*****	0.961**
	Unemp.	0.980	0.987	**0.946**	0.982	1.014
	M2 (Real)	0.972	0.968	1.000	0.981	**0.960**
	PPI: FG	0.967*	**0.941****	1.134	0.995	0.995
	10-Year T. Rate	**0.948****	0.969	1.039	0.995	1.006
	CPI	0.902***	**0.874*****	1.224	1.051	1.058
	RPI ex. Rec.	1.019	1.038	1.069	1.122	1.138
3	Fed Funds	0.940*	0.971	**0.763*****	0.808***	0.818***
	Ind. Prod.	0.912***	0.960	0.878***	**0.860*****	0.874***
	Nonfarm Empl.	0.897***	0.947	0.854***	**0.826*****	0.856***
	S&P 500	0.988**	0.986**	1.000	0.983	**0.980***
	Unemp.	1.022	1.051	1.047	1.088	1.100
	M2 (Real)	**0.963***	0.966	0.987	0.983	0.992
	PPI: FG	0.997	1.000	**0.975****	0.979**	0.978**
	10-Year T. Rate	0.970*	0.972	0.966*	**0.947*****	0.951**
	CPI	0.972**	**0.971****	0.987	0.980	0.984
	RPI ex. Rec.	1.006	1.020	**0.990**	1.002	1.017
6	Fed Funds	0.739***	0.815***	**0.690*****	0.708***	0.727***
	Ind. Prod.	0.980	0.994	0.950**	**0.943*****	0.964**
	Nonfarm Empl.	0.886***	0.915**	0.853***	0.807***	**0.788*****
	S&P 500	0.994	0.992	0.991	0.983*	**0.982***
	Unemp.	1.006	1.011	1.107	1.091	1.105
	M2 (Real)	**0.916*****	0.954*	0.953*	0.933*	0.932*
	PPI: FG	0.984**	0.984**	0.985	0.965***	**0.964****
	10-Year T. Rate	0.973	0.970	0.973	**0.940****	0.955
	CPI	0.998	**0.996**	1.013	1.001	1.006
	RPI ex. Rec.	1.002	1.003	1.002	1.004	1.015
12	Fed Funds	0.839***	0.902**	0.826***	**0.807*****	0.814***
	Ind. Prod.	0.921***	0.947***	0.923***	**0.890*****	0.892***
	Nonfarm Empl.	0.979	1.051	0.802***	**0.797*****	0.805***
	S&P 500	0.991**	0.992*	0.989	0.982**	**0.982****
	Unemp.	1.023	1.019	1.081	1.071	1.094
	M2 (Real)	**0.924*****	0.952**	1.014	0.968	0.992
	PPI: FG	1.015	1.022	1.025	1.021	1.026
	10-Year T. Rate	**0.983***	0.990	1.015	0.985	0.991
	CPI	1.001	1.008	1.027	1.015	1.015
	RPI ex. Rec.	0.969**	0.973*	0.909***	0.900***	**0.897*****

This table reports the MSE of each forecasting strategy (columns) relative to that of PCR, for each of the selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out, and stars denote significance levels. ($^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$).

Table 2: Directional Forecasts

h	Series	*Linear* PCR(%DL)	*Linear* Ridge	*Linear* LASSO	*CAVS: Sign* PCR	*CAVS: Sign* Ridge	*CAVS: Sign* LASSO
1	Fed Funds	40.24	0.994	0.970	0.994	0.976	**0.947**
	Ind. Prod.	31.19	0.947	0.985	**0.924**	1.023	1.008
	Nonfarm Empl.	11.43	0.958	0.938	1.000	0.938	**0.875**
	S&P 500	36.67	0.968	0.942	0.994	**0.935**	0.942
	Unemp.	57.38	1.004	1.021	**0.938**	0.967	0.992
	M2 (Real)	26.67	0.893	**0.884**	1.045	0.911	0.902
	PPI: FG	32.86	1.051	1.014	1.159	1.014	**0.971**
	10-Year T. Rate	40.95	1.023	1.023	1.012	**0.988**	1.017
	CPI	34.52	**0.945**	0.952	1.200	0.966	1.014
	RPI ex. Rec.	25.71	0.991	0.972	0.963	**0.954**	0.954
3	Fed Funds	50.48	1.100	1.261	**0.853*****	0.915*	0.872***
	Ind. Prod.	33.49	1.000	1.007	**0.914**	1.007	1.014
	Nonfarm Empl.	13.16	1.036	1.036	1.073	1.055	1.036
	S&P 500	40.19	0.976	0.970	0.946	0.917	**0.887***
	Unemp.	54.31	1.018	1.035	1.048	1.057	1.026
	M2 (Real)	29.43	1.016	0.992	**0.927**	0.976	0.976
	PPI: FG	51.91	1.018	1.037	0.977	**0.954**	0.954
	10-Year T. Rate	52.87	1.109	1.059	0.977	**0.941**	0.946
	CPI	51.20	1.000	1.009	**0.855*****	0.869***	0.855***
	RPI ex. Rec.	24.64	1.068	1.049	1.049	1.078	1.087
6	Fed Funds	52.53	1.009	1.018	**0.917***	0.931	1.028
	Ind. Prod.	35.66	0.980	0.986	**0.966**	0.966	0.973
	Nonfarm Empl.	15.90	1.030	1.076	0.955	**0.939**	1.015
	S&P 500	38.55	1.000	0.975	0.981	0.975	**0.950**
	Unemp.	56.14	1.017	1.000	1.052	1.056	1.056
	M2 (Real)	28.19	1.051	1.085	1.068	1.000	1.017
	PPI: FG	52.29	0.982	1.000	0.940	**0.931**	0.949
	10-Year T. Rate	50.84	1.043	0.962	0.976	**0.948**	0.957
	CPI	51.57	1.000	0.977	**0.897****	0.897**	0.902**
	RPI ex. Rec.	24.82	1.000	0.981	**0.971**	0.990	1.000
12	Fed Funds	50.86	1.014	1.202	1.067	1.091	1.168
	Ind. Prod.	35.70	**0.966**	0.966	0.966	0.966	0.966
	Nonfarm Empl.	20.29	0.928	0.928	0.940	**0.916**	0.928
	S&P 500	37.65	0.974	1.006	0.974	**0.961**	0.987
	Unemp.	58.19	1.071	1.042	1.042	1.034	1.038
	M2 (Real)	32.27	**0.955**	1.053	0.977	0.992	0.992
	PPI: FG	48.41	0.970	0.995	**0.914***	0.944	0.934
	10-Year T. Rate	46.45	1.074	1.021	1.095	1.105	1.084
	CPI	46.70	1.000	1.031	0.932	**0.901***	0.916
	RPI ex. Rec.	26.41	**0.981**	1.009	0.981	0.981	1.009

This table reports, for each selected series and forecast horizon, the directional loss ratio of each sign forecasting strategy relative to the sign of the linear PCR forecast. Linear sign forecasts are the signs of the linear forecasts constructed, whereas the remaining forecasts are obtained by directly targeting the sign of the series considered. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. Likelihood ratio tests of superior predictive ability relative to the linear PCR benchmark, as described in Section 4.1, are carried out. Stars denote significance levels. ($^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$).

Table 3: Absolute Value Forecasts

h	Series	*Linear*		*CAVS: Absolute Value*			
		Ridge	LASSO	PCR	Ridge	LASSO	GARCH
1	Fed Funds	0.679***	0.590***	0.862	0.752***	0.699***	**0.568*****
	Ind. Prod.	1.056	1.076	0.778***	**0.714*****	0.741***	0.807***
	Nonfarm Empl.	0.980	1.064	0.981	0.994	1.000	**0.917***
	S&P 500	1.006	1.030	0.639***	0.636***	**0.633*****	0.637***
	Unemp.	1.086	1.111	0.862***	**0.849*****	0.860***	0.856***
	M2 (Real)	1.007	1.014	0.910*	0.902**	**0.890****	0.925*
	PPI: FG	0.955**	0.927**	0.805***	0.794***	0.803***	**0.700*****
	10-Year T. Rate	1.006	1.002	0.833***	0.809***	0.797***	**0.738*****
	CPI	0.910**	0.894**	0.869	0.838**	0.866*	**0.731*****
	RPI ex. Rec.	0.994	1.007	0.768**	0.760**	**0.751****	0.854
3	Fed Funds	0.946	0.935	0.941	0.890	0.955	**0.810****
	Ind. Prod.	1.005	0.952*	0.775***	**0.760*****	0.777***	0.762***
	Nonfarm Empl.	0.921*	1.019	1.017	0.975	0.981	**0.908***
	S&P 500	1.017	1.009	0.586***	0.582***	0.589***	**0.578*****
	Unemp.	1.055	1.072	**0.704*****	0.715***	0.725***	0.708***
	M2 (Real)	0.984	0.981	0.829***	**0.806*****	0.812***	0.857***
	PPI: FG	1.001	0.986**	0.537***	0.512***	0.515***	**0.465*****
	10-Year T. Rate	1.092	1.091	0.666***	0.627***	0.626***	**0.564*****
	CPI	1.023	1.017	0.580***	0.563***	0.578***	**0.555*****
	RPI ex. Rec.	0.998	1.005	0.911***	0.905***	**0.904*****	1.010
6	Fed Funds	0.975	0.938	0.913	**0.909**	0.995	0.951
	Ind. Prod.	0.966**	0.954**	0.780***	**0.779*****	0.785***	0.810***
	Nonfarm Empl.	0.891***	0.840***	0.895*	0.866**	0.866**	**0.835*****
	S&P 500	1.019	1.031	0.588***	0.579***	0.581***	**0.571*****
	Unemp.	1.079	1.087	0.666***	0.652***	**0.648*****	0.652***
	M2 (Real)	0.952*	0.977	0.813***	0.783***	**0.781*****	0.805***
	PPI: FG	1.024	1.029	0.540***	0.536***	0.536***	**0.491*****
	10-Year T. Rate	0.950**	0.967*	0.622***	0.613***	0.649***	**0.577*****
	CPI	1.016	1.011	0.583***	0.576***	**0.576*****	0.582***
	RPI ex. Rec.	0.995*	0.986***	0.915***	**0.914*****	0.921***	0.954
12	Fed Funds	1.018	1.002	1.035	1.088	1.301	1.199
	Ind. Prod.	1.047	1.026	0.948	**0.894****	0.906**	0.952
	Nonfarm Empl.	0.788***	**0.777*****	0.866**	0.861**	0.860**	0.842***
	S&P 500	1.018	1.013	0.609***	0.597***	0.602***	**0.596*****
	Unemp.	1.079	1.051	0.672***	0.651***	0.649***	**0.642*****
	M2 (Real)	0.940***	0.974	0.885***	0.848***	0.844***	**0.820*****
	PPI: FG	1.040	1.042	0.608***	0.602***	0.609***	**0.581*****
	10-Year T. Rate	1.052	1.039	0.596***	0.571***	0.592***	**0.563*****
	CPI	1.003	0.997	0.641***	**0.622*****	0.622***	0.630***
	RPI ex. Rec.	1.033	1.048	0.964	0.956	**0.951**	0.989

This table reports the MSE of absolute value forecasts, for each forecasting strategy relative to that of PCR, for each of the selected series and forecast horizons. Absolute values of linear models are the absolute values of the linear forecasts constructed. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out and stars denote significance levels. ($^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$).

Table 4: Additional CAVS Specifications

	Absolute Value:	*PCR*			*Ridge*			*LASSO*			*GARCH*		
h	Series	PCR	Ridge	LASSO	PCR	Ridge	LASSO	PCR	Ridge	LASSO	PCR	Ridge	LASSO
1	Fed Funds	0.692**	0.653***	0.687**	0.587***	0.582***	0.584***	0.569***	0.585***	0.573***	0.578***	0.579***	**0.561*****
	Ind. Prod.	0.891***	0.876***	0.898***	0.877***	0.874***	0.903***	0.867***	**0.865*****	0.894***	0.900***	0.899***	0.926**
	Nonfarm Empl.	0.922*	0.975	0.963	**0.898****	0.941	0.940	0.901**	0.931*	0.926*	0.923*	0.972	0.963
	S&P 500	1.011	0.946***	0.962**	1.008	0.944***	0.960**	1.004	0.944***	0.961**	1.014	**0.941*****	0.956**
	Unemp.	**0.946**	0.970	1.004	0.956	0.982	1.011	0.961	0.986	1.014	0.986	1.001	1.027
	M2 (Real)	1.000	0.981	0.970	0.997	0.981	0.971	0.988	0.969	**0.960**	1.018	1.014	1.002
	PPI: FG	1.134	1.001	1.002	1.142	0.995	0.996	1.140	0.995	0.995	1.112	**0.969**	0.972
	10-Year T. Rate	1.039	0.995	1.006	1.034	0.995	1.009	1.026	0.993	1.006	1.020	**0.980**	0.996
	CPI	1.224	1.108	1.133	1.193	1.051	1.068	1.191	1.038	1.058	1.149	1.044	1.055
	RPI ex. Rec.	1.069	1.115	1.128	1.074	1.122	1.136	1.077	1.126	1.138	1.128	1.205	1.217
3	Fed Funds	**0.763*****	0.786***	0.793***	0.791***	0.808***	0.814***	0.794***	0.814***	0.818***	0.774***	0.783***	0.793***
	Ind. Prod.	0.878***	0.867***	0.884***	0.870***	0.860***	0.878***	0.868***	0.857***	0.874***	0.850***	**0.839*****	0.859***
	Nonfarm Empl.	0.854***	0.864***	0.890**	0.810***	0.826***	0.856***	**0.810*****	0.826***	0.856***	0.822***	0.847***	0.880**
	S&P 500	1.000	0.984	0.981*	0.998	0.983	0.980*	0.997	0.983	**0.980***	0.998	0.983	0.980*
	Unemp.	1.047	1.075	1.089	1.064	1.088	1.099	1.065	1.090	1.100	1.066	1.089	1.101
	M2 (Real)	0.987	0.998	1.006	0.973	0.983	0.990	**0.973**	0.985	0.992	0.985	0.994	1.001
	PPI: FG	**0.975****	0.978**	0.979**	0.977**	0.979**	0.979**	0.977*	0.978**	0.978**	0.983	0.985	0.986
	10-Year T. Rate	0.966*	0.945***	0.951**	0.970*	0.947***	0.954**	0.962**	**0.943*****	0.951**	0.971*	0.951**	0.959**
	CPI	0.987	0.983	0.982	**0.978**	0.980	0.980	0.988	0.986	0.984	0.996	0.993	0.992
	RPI ex. Rec.	0.990	1.000	1.006	**0.989**	1.002	1.009	0.990	1.006	1.017	1.044	1.056	1.068
6	Fed Funds	**0.690*****	0.707***	0.730***	0.694***	0.708***	0.726***	0.700***	0.711***	0.727***	0.704***	0.721***	0.736***
	Ind. Prod.	0.950**	0.948***	0.977	0.944***	0.943***	0.971*	0.941***	0.938***	0.964**	0.940***	**0.934*****	0.963**
	Nonfarm Empl.	0.853***	0.824***	0.805***	0.835***	0.807***	0.790***	0.833***	0.806***	**0.788*****	0.827***	0.802***	0.796***
	S&P 500	0.991	0.984	0.981*	0.990	0.983*	**0.981***	0.991	0.984*	0.982*	0.988	0.984*	0.982*
	Unemp.	1.107	1.098	1.114	1.096	1.091	1.104	1.095	1.093	1.105	1.089	1.087	1.098
	M2 (Real)	0.953*	0.955	0.967	0.930*	0.933*	0.944	**0.917****	0.920*	0.932*	0.931*	0.930*	0.941*
	PPI: FG	0.985	0.966***	0.967**	0.982	0.965***	0.966**	0.984	**0.964*****	0.964**	0.986	0.966***	0.969**
	10-Year T. Rate	0.973	0.940**	0.965	0.971	0.940**	0.959	0.963	**0.937****	0.955	0.975	0.942*	0.957
	CPI	1.013	1.000	1.006	1.012	1.001	1.007	1.010	1.001	1.006	1.025	1.012	1.023
	RPI ex. Rec.	1.002	1.001	1.004	1.004	1.004	1.007	1.011	1.012	1.015	1.037	1.039	1.041
12	Fed Funds	0.826***	0.815***	0.805***	0.813***	0.807***	**0.798*****	0.844***	0.818***	0.814***	0.823***	0.813***	0.815***
	Ind. Prod.	0.923***	0.894***	0.901***	0.920***	0.890***	0.898***	0.908***	**0.885*****	0.892***	0.926***	0.893***	0.910***
	Nonfarm Empl.	0.802***	**0.792*****	0.798***	0.805***	0.797***	0.804***	0.805***	0.797***	0.805***	0.845***	0.842***	0.856***
	S&P 500	0.989	0.986*	0.986*	0.985*	0.982**	0.983**	0.984**	0.981**	0.982**	0.980**	**0.977*****	0.978***
	Unemp.	1.081	1.070	1.092	1.082	1.071	1.091	1.084	1.073	1.094	1.064	1.054	1.070
	M2 (Real)	1.014	0.982	1.011	0.989	0.968	0.995	0.986	0.964	0.992	0.968	**0.946***	0.974
	PPI: FG	1.025	1.027	1.030	1.019	1.021	1.025	1.019	1.022	1.026	1.013	1.020	1.025
	10-Year T. Rate	1.015	0.988	0.991	1.018	**0.985**	0.988	1.016	0.987	0.991	1.010	0.986	0.990
	CPI	1.027	1.019	1.019	1.021	1.015	1.014	1.023	1.016	1.015	1.026	1.020	1.020
	RPI ex. Rec.	0.909***	0.896***	**0.891*****	0.911***	0.900***	0.895***	0.914***	0.903***	0.897***	0.940*	0.929**	0.923**

This table reports the MSE of CAVS-based forecasts relative to that of PCR, for each of the selected series and forecast horizons. The first row denotes the Absolute Value forecasting strategy, and the second row the sign forecasting strategy considered. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out, and stars denote significance levels. ($^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$).

Table 5: Machine Learning Forecasts

Series	PCR	Ridge	LASSO	RF	Adab.	k-NN	NN
Fed Funds	0.587	0.582	0.584	**0.473**	0.540	0.577	0.829
Ind. Prod.	0.877	**0.874**	0.903	0.880	0.892	0.937	1.115
Nonfarm Empl.	**0.898**	0.941	0.940	0.903	1.154	1.197	1.581
S&P 500	1.008***	**0.944*****	0.960***	0.986***	1.043***	0.988*	1.076
Unemp.	**0.956**	0.982	1.011	0.977	1.017	1.043	1.260
M2 (Real)	0.997	0.981	**0.971**	1.095	1.075	1.119	1.113
PPI: FG	1.142	**0.995**	0.996	1.215	1.256	1.233	1.173
10-Year T. Rate	1.034	**0.995**	1.009	1.018	1.025	1.036	1.251
CPI	1.193	1.051	1.068	1.107	1.123	1.165	1.212
RPI ex. Rec.	1.074	1.122	1.136	1.015	1.028	1.055	1.219

This table reports the results for select series. For each series, the table reports the ratios of the MSE of the CAVS specification considered relative to PCR forecasts. The absolute values employed in this CAVS forecast are obtained by Ridge, and the sign forecasts are based on the methods in the columns. Best performing methods are highlighted in boldface and stars represent significance of a superior predictive ability test relative to the PCR forecast at 5% significance level. If no method is highlighted, the PCR benchmark is the best performing method.

Table 6: Predictability

h	Series	Linear			CAVS		
		PCR	Ridge	LASSO	PCR	Ridge	LASSO
1	Output	37.50	**68.75**	50.00	**68.75**	**68.75**	**68.75**
	Consumption	**85.71**	**85.71**	71.43	57.14	71.43	71.43
	Labor	82.76	89.66	82.76	**93.10**	**93.10**	86.21
	Money	50.00	57.14	57.14	35.71	**64.29**	57.14
	Stocks	25.00	**100.00**	75.00	25.00	**100.00**	**100.00**
	Interest and Exc.	45.00	55.00	45.00	60.00	**95.00**	85.00
	Prices	85.00	**95.00**	**95.00**	45.00	60.00	65.00
	FRED-MD	63.64	77.27	69.09	62.73	**79.09**	75.45
3	Output	12.50	43.75	31.25	56.25	**62.50**	50.00
	Consumption	28.57	28.57	28.57	**42.86**	**42.86**	**42.86**
	Labor	58.62	65.52	62.07	68.97	**72.41**	65.52
	Money	0.00	**7.14**	**7.14**	**7.14**	**7.14**	**7.14**
	Stocks	0.00	0.00	0.00	0.00	0.00	0.00
	Interest and Exc.	30.00	35.00	35.00	**40.00**	35.00	35.00
	Prices	0.00	0.00	0.00	0.00	0.00	0.00
	FRED-MD	24.55	32.73	30.00	37.27	**38.18**	34.55
6	Output	18.75	25.00	18.75	25.00	**43.75**	37.50
	Consumption	**28.57**	**28.57**	**28.57**	**28.57**	**28.57**	**28.57**
	Labor	68.97	**75.86**	65.52	68.97	72.41	72.41
	Money	0.00	0.00	0.00	0.00	0.00	0.00
	Stocks	0.00	0.00	0.00	0.00	0.00	0.00
	Interest and Exc.	25.00	30.00	30.00	**35.00**	**35.00**	25.00
	Prices	0.00	0.00	0.00	0.00	0.00	0.00
	FRED-MD	27.27	30.91	27.27	30.00	**33.64**	30.91
12	Output	0.00	6.25	6.25	6.25	**37.50**	31.25
	Consumption	0.00	14.29	0.00	**28.57**	**28.57**	**28.57**
	Labor	31.03	37.93	27.59	41.38	**48.28**	41.38
	Money	**7.14**	**7.14**	**7.14**	**7.14**	**7.14**	**7.14**
	Stocks	0.00	0.00	0.00	0.00	0.00	0.00
	Interest and Exc.	25.00	25.00	20.00	**30.00**	25.00	25.00
	Prices	**5.00**	**5.00**	0.00	**5.00**	**5.00**	0.00
	FRED-MD	14.55	18.18	12.73	20.91	**26.36**	22.73

This table reports the results for each group in the FRED-MD, as well as for the dataset as a whole. For each group and forecast horizon, this table reports the percentage of series within a group for which each method outperforms an unconditional mean benchmark according to a DM test at the 5% significance level. Best performing methods for each group are highlighted in boldface.

Table 7: Forecast Combinations

		Model Selection		*Forecast Averaging*		
h	Series	Linear	CAVS	Linear	CAVS	*Hybrid*
1	Fed Funds	1.000	0.580***	0.719***	0.574***	**0.538*****
	Ind. Prod.	0.970**	0.880***	0.951***	**0.867*****	0.891***
	Nonfarm Empl.	1.000	0.946	0.953**	**0.892****	0.901***
	S&P 500	0.969*	**0.948*****	0.964***	0.960***	0.951***
	Unemp.	0.987	0.996	0.969**	0.973	**0.946****
	M2 (Real)	0.978	1.001	0.967**	**0.956**	0.973
	PPI: FG	**0.941****	0.981	0.961**	1.010	0.957**
	10-Year T. Rate	0.994	1.001	0.954***	1.000	**0.940****
	CPI	**0.874*****	1.066	0.907***	1.084	0.946
	RPI ex. Rec.	1.000	1.080	1.015	1.103	1.023
3	Fed Funds	1.000	0.792***	0.930**	**0.785*****	0.842***
	Ind. Prod.	0.951***	0.873***	0.939***	**0.863*****	0.897***
	Nonfarm Empl.	1.003	**0.822*****	0.910***	0.828***	0.843***
	S&P 500	0.995	0.987	0.990**	**0.986**	0.986*
	Unemp.	1.002	1.061	1.018	1.076	1.023
	M2 (Real)	0.973	0.982	0.969*	0.981	**0.964***
	PPI: FG	1.002	0.983	0.998	**0.977****	0.986**
	10-Year T. Rate	0.991	0.967**	0.973**	**0.950****	0.962***
	CPI	1.000	0.986	0.975**	0.979	**0.973****
	RPI ex. Rec.	1.003	**0.981**	1.006	0.998	0.989
6	Fed Funds	0.807***	0.717***	0.796***	**0.705*****	0.730***
	Ind. Prod.	0.980	0.949**	0.985	**0.945*****	0.957***
	Nonfarm Empl.	0.902***	0.817***	0.904***	**0.796*****	0.818***
	S&P 500	1.000	0.986	0.994	**0.983***	0.987*
	Unemp.	1.003	1.089	**0.996**	1.096	1.021
	M2 (Real)	**0.918*****	0.935*	0.949***	0.936*	0.931**
	PPI: FG	0.999	0.970**	0.988**	**0.970****	0.975***
	10-Year T. Rate	1.002	0.952*	0.965**	0.949*	**0.946****
	CPI	0.998	1.008	0.997	1.002	**0.993**
	RPI ex. Rec.	1.004	1.015	1.002	1.006	1.005
12	Fed Funds	1.000	0.820***	0.869***	**0.807*****	0.829***
	Ind. Prod.	0.953***	0.909***	0.943***	**0.896*****	0.920***
	Nonfarm Empl.	0.997	0.791***	0.984	**0.789*****	0.857***
	S&P 500	0.996*	**0.978*****	0.994**	0.984**	0.984***
	Unemp.	1.019	1.076	1.006	1.077	1.021
	M2 (Real)	**0.929*****	0.948*	0.948***	0.984	0.938***
	PPI: FG	1.000	1.023	1.010	1.020	1.010
	10-Year T. Rate	0.994	0.992	0.987*	0.993	**0.984**
	CPI	1.003	1.014	1.002	1.018	**0.999**
	RPI ex. Rec.	0.980	0.911**	0.977**	**0.899*****	0.935***

This table reports the MSE of each forecast combination strategy (columns) relative to that of PCR, for each of the selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out, and stars denote significance level. ($^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$).

A Proofs

Proof of propositions 1 and 2. We need to show that

$$\mathbb{E}\Big[\Big(Y_t - \mu_{\text{CAVS}}(X_t)\Big)^2\Big] \leq \sigma_u^2 + a_1\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big]$$

where a_1 is a constant that depends on $\mathbb{E}[Y_t^2]$. Let $M = (\mu_A(X_t)\ ,\ m_S^*(X_t))'$ and $\varepsilon = \Big(\mu_S(X_t) - m_S^*(X_t), \mu_A(X_t) - m_A^*(X_t)\Big)'$. First, note that

$$||M'\varepsilon||_2^2 = \Big(\mu_A(X_t)(\mu_S(X_t) - m_S^*(X_t)) + m_S^*(X_t)(\mu_A(X_t) - m_A^*(X_t)\Big)^2 ,$$

and

$$||M||_2^2 = \mu_A(X_t)^2 + m_S^*(X_t)^2 .$$

Note that $|m_S^*(X_t)| \leq 1$. Additionally,

$$||\varepsilon||_2^2 = \Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2 .$$

Applying Cauchy-Schwarz's inequality, we have

$$\begin{aligned}
||M'\varepsilon||^2 &\leq ||M||_2^2||\varepsilon||_2^2 \\
&\leq (1 + \mu_A(X_t)^2)||\varepsilon||_2^2 \\
&= (1 + \mu_A(X_t)^2)\Big(\big(\mu_S(X_t) - m_S^*(X_t)\big)^2 + \big(\mu_A(X_t) - m_A^*(X_t)\big)^2\Big)
\end{aligned}$$

Note that $\mu_A(X_t) = \mathbb{E}\Big[|Y_t|\Big|X_t\Big]$. Next, because $\mathbb{V}\text{ar}(|Y_t|\ |X_t) = \mathbb{E}[Y_t^2|X_t] - \mu_A(X_t)^2 \geq 0$, It follows that $\mu_A(X_t)^2 \leq \mathbb{E}[Y_t^2|X_t]$. Hence

$$\begin{aligned}
\mathbb{E}\Big[\Big(Y_t - \mu_{\text{CAVS}}(X_t)\Big)^2\Big] &= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu(X_t) - m_A^*(X_t)m_S^*(X_t) - c(X_t)\Big)^2\Big] \\
&= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu_A(X_t)\mu_S(X_t) - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \\
&= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu_A(X_t)\big(\mu_S(X_t) - m_S^*(X_t)\big) + m_S^*(X_t)\big(\mu_A(X_t) - m_A^*(X_t)\big)\Big)^2\Big] \\
&\leq \sigma_u^2 + \mathbb{E}\Big[||M'\varepsilon||^2\Big] \\
&\leq \sigma_u^2 + \mathbb{E}\Big[\big(1 + \mathbb{E}[Y_t^2|X_t]\big)\Big(\big(\mu_S(X_t) - m_S^*(X_t)\big)^2 + \big(\mu_A(X_t) - m_A^*(X_t)\big)^2\Big)\Big] \\
&\leq \sigma_u^2 + a_1\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big]
\end{aligned}$$

where the last line follows from the assumption that $\mathbb{E}[Y_t^2] < \infty$ and $\mathbb{E}[Y_t|X_t] < \infty$, which together imply that $\mathbb{E}[Y_t^2|X_t] \leq a_0 < \infty$ for all X_t with positive measure.

Proof of Proposition 2. Note that

$$\begin{aligned}\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] &= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t) + \mu_{\text{CAVS}}(X_t) - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \\ &= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t) + c(X_t)\Big)^2\Big]\end{aligned}$$

Next using the fact that $(a-b)^2 \le 2(a^2+b^2)$, we write

$$\mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t) + c(X_t)\Big)^2\Big] \le 2\mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t)\Big)^2 + c(X_t)^2\Big]$$

Applying Proposition 1, we have

$$\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \le \sigma_u^2 + 2\mathbb{E}\Big[c(X_t)^2\Big] + a_2\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big]$$

□

Proof of proposition 3. (To be completed)
Let

$$Y_t = \mu(X_t) + \sigma_u u_t \;\; u_t \sim \mathcal{D}(0,1) \; ,$$

and denote by $\mu_A(X_t) = \mathbb{E}\Big[\big|\mu(X_t) + \sigma_u u_t\big|\Big|X_t\Big]$ and by $\mu_S(X_t) = \mathbb{E}\Big[\text{sign}(Y_t)|X_t\Big]$. We will show that

$$\Big(\mu(X_t) - \mu_A(X_t)\mu_S(X_t)\Big)^2 \le \gamma\sigma_u^2 \; ,$$

which implies that $\mathbb{E}\Big[\Big(\mu(X_t) - \mu_A(X_t)\mu_S(X_t)\Big)^2\Big] \le \gamma\sigma_u^2$. Let $z = \frac{\mu(X_t)}{\sigma_u}$, and note that

$$\begin{aligned}&\mu_A(X_t) = \sigma_u\mathbb{E}_u\Big[\big|z + u_t\big|\Big], \text{ and} \\ &\mu_S(X_t) = 1 - 2F_{\mathcal{D}}\Big(-\frac{\mu(X_t)}{\sigma_u}\Big) = 1 - 2F_{\mathcal{D}}(-z)\end{aligned}$$

where $\mathbb{E}_u$ denotes expectation taken with respect to $\mathcal{D}$, the distribution of u_t. In addition, note that z is measurable with respect to X_t. Our goal is to bound

$$\begin{aligned}d(z) &= \Big(\mu(X_t) - \mu_A(X_t)\mu_S(X_t)\Big)^2 \\ &= \Big(\sigma_u z - \sigma_u\mathbb{E}_u\Big[|z+u_t|\Big]\big(1 - 2F_{\mathcal{D}}^{-1}(-z)\big)\Big)^2 \\ &= \sigma_u^2\Big(z - \mathbb{E}_u\Big[|z+u_t|\Big]\big(1 - 2F_{\mathcal{D}}^{-1}(-z)\big)\Big)^2 \; .\end{aligned}$$

Applying Lemma 1 we obtain

$$d(z) \le \sigma^2 M\frac{(z^*)^6}{36}$$

We can numerically compute M and z^* for different distributions. In particular, we numerically evaluate $g'''(z^*)$ for the $t-$ distribution to obtain that $|g'''(z^*)| \leq 3$, which implies $\gamma \leq 1/2$. A more general proof that does not require numerical evaluation (and hence does not assume the t distribution) is being completed.

For the Gaussian case, note that

$$\begin{aligned}\mathbb{E}\Big[(Y_t - \mu_A(X_t)\mu_S(X_t))^2\Big] &= \sigma_u^2 + \mathbb{E}\Big[(\mu(X_t) - \mu_A(X_t)\mu_S(X_t))^2\Big] \\ &= \sigma_u^2 + \mathbb{E}\Big[\mathbb{C}\mathrm{ov}\Big(|Y_t|, \mathrm{sign}(Y_t)\Big|X_t\Big)^2\Big]\end{aligned}$$

Hence, clearly

$$\frac{\mathbb{E}\Big[\Big(Y_t - \mu_A(X_t)\mu_S(X_t)\Big)^2\Big]}{\mathbb{E}\Big[\Big(Y_t - \mu(X_t)\Big)^2\Big]} = 1 + \frac{1}{\sigma_u^2}\mathbb{E}\Big[\mathbb{C}\mathrm{ov}\Big(|Y_t|, \mathrm{sign}(Y_t)\Big|X_t\Big)^2\Big]$$

Next, we compute the conditional covariance between the signs and absolute values of Y_t. We must evaluate the expectations. First, note that $Y_t\big|X_t \sim N(\mu(X_t), \sigma_u^2)$. It follows that $|Y_t|\big|X_t$ is the absolute value of a Gaussian random variable, and hence it is distributed as a Folded Normal. The expected value of $|Y_t|\big|X_t$ can be directly computed as:

$$\mathbb{E}\Big[|Y_t|\Big|X_t\Big] = 2\sigma_u\phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) + \mu(X_t)\Big(2\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) - 1\Big)$$

In addition, noting that

$$P\Big(\mu(X_t) + u_{t+1} \geq 0\big|X_t\Big) = \Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)$$

the expected value of $\mathrm{sign}(Y_t)|X_t$ is given by

$$\mathbb{E}_t\Big[\mathrm{sign}(Y_{t+1})\big|X_t\Big] = 2\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) - 1$$

Applying the definition of covariance, we obtain

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\Big(|Y_t|, \mathrm{sign}(Y_t)\Big|X_t\Big) = \mu(X_t) - 2\sigma_u\phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big(2\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) - 1\Big) - \mu(X_t)\Big(2\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) - 1\Big)^2 \\ &= \mu(X_t) - 2\sigma_u\phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big(2\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) - 1\Big) - \mu(X_t)\Big(1 - 4\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big(1 - \Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big) \\ &= 4\mu(X_t)\Big(\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big(1 - \Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big) - 2\sigma_u\phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big)\Big(2\Phi\Big(\frac{\mu(X_t)}{\sigma_u}\Big) - 1\Big)\end{aligned}$$

Setting $z = \frac{\mu(X_t)}{\sigma_u}$, we can write

$$\begin{aligned}\mathbb{C}\mathrm{ov}\Big(|Y_t|, \mathrm{sign}(Y_t)\Big|X_t\Big) &= 4z\sigma_u\big(\Phi(z)(1 - \Phi(z))\big) - 2\sigma_u\phi(z)\big(2\Phi(z) - 1\big) \\ &= 2\sigma_u\big(2z\Phi(z)(1 - \Phi(z)) - 2\phi(z)\Phi(z) + \phi(z)\big)\end{aligned}$$

Clearly, this function is 0 at $z = 0$. Additionally, it is easy to see that as $|z| \to \infty$, this function goes to 0. Scaling the covariance by the irreducible MSE, we obtain:

$$\begin{aligned}
\frac{1}{\sigma_u^2}\mathbb{E}\Big[(Y_t - \mu_A(X_t)\mu_S(X_t))^2\Big] &= 1 + \frac{1}{\sigma_u^2}\mathbb{E}\Big[\mathbb{C}\mathrm{ov}\Big(|Y_t|, \mathrm{sign}(Y_t)\Big|X_t\Big)^2\Big] \\
&= 1 + \mathbb{E}\Big[\frac{4\sigma_u^2\big(2z\Phi(z)(1-\Phi(z)) - 2\phi(z)\Phi(z) + \phi(z)\big)^2}{\sigma_u^2}\Big] \\
&= 1 + 4\mathbb{E}\Big[\big(2z\Phi(z)(1-\Phi(z)) - 2\phi(z)\Phi(z) + \phi(z)\big)^2\Big] \\
&\leq 1 + 4(2\tilde{z}\Phi(z^*)(1-\Phi(z^*)) - 2\phi(z^*)\Phi(z^*) + \phi(z^*)\big)^2 \\
&\approx 1.04187
\end{aligned}$$

where $z^* = \arg\max_{z\in\mathbb{R}} \big(2z\Phi(z)(1-\Phi(z)) - 2\phi(z)\Phi(z) + \phi(z)\big)^2$, evaluated numerically.

□

Proof of proposition 4. Let $X \sim \mathcal{D}(0, \Sigma)$, with Σ a diagonal matrix with elemtens given by σ_1^2 and σ_2^2. Write

$$Y_t = |x_1\kappa\beta_1 + x_2\beta_2|\mathrm{sign}(x_1\beta_1 + x_2\beta_2) + u_t \ , u_t \sim \mathcal{D}(0,1)$$

The OLS estimator for β_1^* is given by

$$\begin{aligned}
\beta_1^* &= \frac{1}{\sigma_1^2}\mathbb{E}\Big[x_1|x_1\kappa\beta_1 + x_2\beta_2|\mathrm{sign}(x_1\beta_1 + x_2\beta_2)\Big] \\
&= \frac{1}{\sigma_1^2}\mathbb{E}\Big[|x_1||x_1\kappa\beta_1 + x_2\beta_2|\mathrm{sign}(x_1)\mathrm{sign}(x_1\beta_1 + x_2\beta_2)\Big] \\
&= \frac{1}{\sigma_1^2}\mathbb{E}\Big[|x_1^2\kappa\beta_1 + x_1x_2\beta_2|\mathrm{sign}(x_1^2\beta_1 + x_1x_2\beta_2)\Big] \\
&= \frac{1}{\sigma_1^2}\mathbb{E}\Big[(x_1^2\kappa\beta_1 + x_1x_2\beta_2)\mathrm{sign}(x_1^2\beta_1^2\kappa + x_1x_2\beta_1\beta_2(1+k) + x_2^2\beta_2)\Big]
\end{aligned}$$

Alternatively, for β_2^*, we have

$$\begin{aligned}
\beta_2^* &= \frac{1}{\sigma_2^2}\mathbb{E}\Big[x_2|x_1\kappa\beta_1 + x_2\beta_2|\mathrm{sign}(x_1\beta_1 + x_2\beta_2)\Big] \\
&= \frac{1}{\sigma_2^2}\mathbb{E}\Big[|x_2||x_1\kappa\beta_1 + x_2\beta_2|\mathrm{sign}(x_2)\mathrm{sign}(x_1\beta_1 + x_2\beta_2)\Big] \\
&= \frac{1}{\sigma_2^2}\mathbb{E}\Big[|x_1x_2\kappa\beta_1 + x_2^2\beta_2|\mathrm{sign}(x_2x_1\beta_1 + x_2^2\beta_2)\Big] \\
&= \frac{1}{\sigma_2^2}\mathbb{E}\Big[(x_1x_2\kappa\beta_1 + x_2^2\beta_2)\mathrm{sign}(x_1x_2\beta_1 + x_2^2\beta_2)\mathrm{sign}(x_1x_2\kappa\beta_1 + x_2^2\beta_2)\Big] \\
&= \frac{1}{\sigma_2^2}\mathbb{E}\Big[(x_1x_2\kappa\beta_1 + x_2^2\beta_2)\mathrm{sign}(x_1^2\beta_1^2\kappa + x_1x_2\beta_1\beta_2(1+k) + x_2^2\beta_2)\Big]
\end{aligned}$$

So

$$
\begin{aligned}
\beta_1^* &= \frac{1}{\sigma_1^2}\mathbb{E}\Big[(x_1^2\kappa\beta_1 + x_1x_2\beta_2)\text{sign}(x_1^2\beta_1^2\kappa + x_1x_2\beta_1\beta_2(1+k) + x_2^2\beta_2)\Big], \text{ and}\\
\beta_2^* &= \frac{1}{\sigma_2^2}\mathbb{E}\Big[(x_1x_2\kappa\beta_1 + x_2^2\beta_2)\text{sign}(x_1^2\beta_1^2\kappa + x_1x_2\beta_1\beta_2(1+k) + x_2^2\beta_2)\Big]
\end{aligned}
$$

Further, assuming $\beta_i = \beta$ for all i:

$$
\begin{aligned}
\beta_1^* &= \frac{1}{\sigma_1^2}\beta\mathbb{E}\Big[(x_1^2\kappa + x_1x_2)\text{sign}(x_1^2\kappa + x_1x_2(1+\kappa) + x_2^2)\Big], \text{ and}\\
\beta_2^* &= \frac{1}{\sigma_2^2}\beta\mathbb{E}\Big[(x_1x_2\kappa + x_2^2)\text{sign}(x_1^2\kappa + x_1x_2(1+\kappa) + x_2^2)\Big]
\end{aligned}
$$

Let $S(\kappa) = \text{sign}(x_1^2\kappa + x_1x_2(1+\kappa) + x_2^2)$ and denote by $a = \mathbb{E}\Big[x_1^2 S(\kappa)\Big]$, $b = \mathbb{E}\Big[x_2^2 S(\kappa)\Big]$ and $c = \mathbb{E}\Big[x_1x_2 S(\kappa)\Big]$.

$$
\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2 = \mathbb{E}\Big[\Big(\beta(x_1\kappa + x_2)S(\kappa) - \beta\frac{x_1}{\sigma_1^2}\mathbb{E}\Big[(x_1^2\kappa + x_1x_2)S(\kappa)\Big] - \beta\frac{x_2}{\sigma_2^2}\mathbb{E}\Big[(x_1x_2\kappa + x_2^2)S(\kappa)\Big]\Big)^2\Big]
$$

and we can write the scaled excess MSE as:

$$
\begin{aligned}
\beta^{-2}\Big(\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2\Big) &= \mathbb{E}\Big[\Big((x_1\kappa + x_2)S(\kappa) - \frac{x_1}{\sigma_1^2}\mathbb{E}\Big[(x_1^2\kappa + x_1x_2)S(\kappa)\Big] - \frac{x_2}{\sigma_2^2}\mathbb{E}\Big[(x_1x_2\kappa + x_2^2)S(\kappa)\Big]\Big)^2\Big]\\
&= \mathbb{E}\Big[\Big(S(\kappa)x_1\kappa + S(\kappa)x_2 - \frac{x_1}{\sigma_1^2}\kappa a - \frac{c}{\sigma_1^2}x_1 - \frac{c}{\sigma_2^2}\kappa x_2 - \frac{x_2}{\sigma_2^2}b\Big)^2\Big]\\
&= \mathbb{E}\Big[\Big(\underbrace{S(\kappa)(x_1\kappa + x_2)}_{(A)} - \underbrace{\kappa\big(\frac{x_1}{\sigma_1^2}a + \frac{c}{\sigma_2^2}x_2\big)}_{(B)} - \underbrace{\big(\frac{c}{\sigma_1^2}x_1 + \frac{x_2}{\sigma_2^2}b\big)}_{(C)}\Big)^2\Big]
\end{aligned}
$$

Term by term, first A^2, using that $\mathbb{E}\Big[x_1x_2\Big] = 0$:

$$
\begin{aligned}
\mathbb{E}\Big[A^2\Big] &= \mathbb{E}\Big[S(\kappa)^2(x_1\kappa + x_2)^2\Big]\\
&= \mathbb{E}\Big[(x_1\kappa + x_2)^2\Big]\\
&= \kappa^2\sigma_1^2 + \sigma_2^2
\end{aligned}
$$

Then, B^2:

$$
\begin{aligned}
\mathbb{E}\Big[B^2\Big] &= \kappa^2\mathbb{E}\Big[\Big(\frac{x_1}{\sigma_1^2}a + \frac{c}{\sigma_2^2}x_2\Big)^2\Big]\\
&= \kappa^2\Big(\frac{a^2}{\sigma_1^2} + \frac{c^2}{\sigma_2^2}\Big)
\end{aligned}
$$

For C^2:

$$\begin{aligned}
\mathbb{E}\Big[C^2\Big] &= \mathbb{E}\Big[(\frac{c}{\sigma_1^2}x_1 + \frac{x_2}{\sigma_2^2}b)^2\Big] \\
&= \frac{c^2}{\sigma_1^2} + \frac{b^2}{\sigma_2^2}
\end{aligned}$$

Then for AB

$$\begin{aligned}
\mathbb{E}\Big[AB\Big] &= \kappa\mathbb{E}\Big[S(\kappa)(x_1\kappa + x_2)(\frac{x_1}{\sigma_1^2}a + \frac{c}{\sigma_2^2}x_2)\Big] \\
&= \kappa\mathbb{E}\Big[S(\kappa)(\sigma_1^{-2}x_1^2\kappa a + x_1x_2\kappa c\sigma_2^{-2} + x_2x_1\sigma_1^{-2}a + c\sigma_2^{-2}x_2^2)\Big] \\
&= \kappa\Big(\sigma_1^{-2}\kappa a\mathbb{E}\Big[S(\kappa)x_1^2\Big] + \kappa c\sigma_2^{-2}\mathbb{E}\Big[S(\kappa)x_1x_2\Big] + \sigma_1^{-2}a\mathbb{E}\Big[S(\kappa)x_2x_1\Big] + c\sigma_2^{-2}\mathbb{E}\Big[S(\kappa)x_2^2\Big]\Big) \\
&= \kappa\Big(\sigma_1^{-2}\kappa a^2 + \kappa c^2\sigma_2^{-2} + \sigma_1^{-2}ac + bc\sigma_2^{-2}\Big) \\
&= \kappa^2\Big(\frac{a^2}{\sigma_1^2} + \frac{c^2}{\sigma_2^2}\Big) + \kappa c\Big(\frac{a}{\sigma_1^2} + \frac{b}{\sigma_2^2}\Big)
\end{aligned}$$

For BC

$$\begin{aligned}
\mathbb{E}\Big[BC\Big] &= \kappa\mathbb{E}\Big[\big(\frac{a}{\sigma_1^2}x_1 + \frac{c}{\sigma_2^2}x_2\big)\big(\frac{c}{\sigma_1^2}x_1 + \frac{b}{\sigma_2^2}x_2\big)\Big] \\
&= \kappa\mathbb{E}\Big[\big(\frac{ac}{\sigma_1^4}x_1^2 + \frac{ab}{\sigma_1^2\sigma_2^2}x_1x_2 + \frac{c^2}{\sigma_2^2\sigma_1^2}x_2x_1 + \frac{bc}{\sigma_2^4}x_2^2\big)\Big] \\
&= \kappa c\big(\frac{a}{\sigma_1^2} + \frac{b}{\sigma_2^2}\big)
\end{aligned}$$

For AC

$$\begin{aligned}
\mathbb{E}\Big[AC\Big] &= \mathbb{E}\Big[S(\kappa)(x_1\kappa + x_2)\big(\frac{x_1}{\sigma_1^2}c + \frac{x_2}{\sigma_2^2}b\big)\Big] \\
&= \mathbb{E}\Big[S(\kappa)\big(\frac{x_1^2}{\sigma_1^2}\kappa c + x_1\frac{x_2}{\sigma_2^2}\kappa b + x_2\frac{x_1}{\sigma_1^2}c + \frac{x_2^2}{\sigma_2^2}b\big)\Big] \\
&= \frac{1}{\sigma_1^2}\kappa c\mathbb{E}\Big[S(\kappa)x_1^2\Big] + \frac{1}{\sigma_2^2}\kappa b\mathbb{E}\Big[S(\kappa)x_1x_2\Big] + \frac{1}{\sigma_1^2}c\mathbb{E}\Big[S(\kappa)x_2x_1\Big] + \frac{1}{\sigma_2^2}b\mathbb{E}\Big[S(\kappa)x_2^2\Big]\big) \\
&= \frac{1}{\sigma_1^2}\kappa ac + \frac{1}{\sigma_2^2}\kappa bc + \frac{1}{\sigma_1^2}c^2 + \frac{1}{\sigma_2^2}b^2 \\
&= \kappa c\Big(\frac{a}{\sigma_1^2} + \frac{b}{\sigma_2^2}\Big) + \frac{c^2}{\sigma_1^2} + \frac{b^2}{\sigma_2^2}
\end{aligned}$$

Combining everything and noting that $\mathbb{E}\Big[AB\Big] = \mathbb{E}\Big[B^2\Big] + \mathbb{E}\Big[BC\Big]$ and $\mathbb{E}\Big[AC\Big] = \mathbb{E}\Big[C^2\Big] +$

$\mathbb{E}\Big[BC\Big]$, we have

$$\beta^{-2}\Big(\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2\Big) = \kappa^2\Big(\frac{1}{\sigma_1^2}(\sigma_1^4 - a^2) - \frac{c^2}{\sigma_2^2}\Big) - 2\kappa c\Big(\frac{a}{\sigma_1^2} + \frac{b}{\sigma_2^2}\Big) + \frac{1}{\sigma_2^2}(\sigma_2^4 - b^2) - \frac{c^2}{\sigma_1^2}$$

Plugging the values for a, b and c, and denoting the excess risk by $R(\kappa)$, we have

$$\begin{aligned}
R(\kappa) &= \kappa^2\Big(\frac{1}{\sigma_1^2}\Big(\sigma_1^4 - \mathbb{E}[x_1^2 S(\kappa)]^2\Big) - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_2^2}\Big) - 2\kappa\mathbb{E}[x_1x_2S(\kappa)]\Big(\frac{\mathbb{E}[x_1^2S(\kappa)]}{\sigma_1^2} + \frac{\mathbb{E}[x_2^2S(\kappa)]}{\sigma_2^2}\Big) \\
&+ \frac{1}{\sigma_2^2}(\sigma_2^4 - \mathbb{E}[x_2^2S(\kappa)]^2) - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_1^2} \\
&= \kappa^2\Big(\sigma_1^2 - \frac{\mathbb{E}[x_1^2S(\kappa)]^2}{\sigma_1^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_2^2}\Big) - 2\kappa\mathbb{E}[x_1x_2S(\kappa)]\Big(\frac{\mathbb{E}[x_1^2S(\kappa)]}{\sigma_1^2} + \frac{\mathbb{E}[x_2^2S(\kappa)]}{\sigma_2^2}\Big) \\
&+ \sigma_2^2 - \frac{\mathbb{E}[x_2^2S(\kappa)]^2}{\sigma_2^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_1^2}
\end{aligned}$$

□

Proof of Corollary 2. (To be completed) Let

$$\begin{aligned}
R(\kappa) &= \kappa^2\Big(\sigma_1^2 - \frac{\mathbb{E}[x_1^2S(\kappa)]^2}{\sigma_1^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_2^2}\Big) - 2\kappa\mathbb{E}[x_1x_2S(\kappa)]\Big(\frac{\mathbb{E}[x_1^2S(\kappa)]}{\sigma_1^2} + \frac{\mathbb{E}[x_2^2S(\kappa)]}{\sigma_2^2}\Big) \\
&+ \sigma_2^2 - \frac{\mathbb{E}[x_2^2S(\kappa)]^2}{\sigma_2^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_1^2}
\end{aligned}$$

We need to show that $R(\kappa) \in \Omega(\kappa^2)$, that is, we need to show that there exists a constant $a_3 > 0$ and κ_0 such $R(\kappa) > a_3\kappa^2$ for every $\kappa > \kappa_0$. Consider

$$\begin{aligned}
\frac{R(\kappa)}{\kappa^2} &= \Big(\sigma_1^2 - \frac{\mathbb{E}[x_1^2S(\kappa)]^2}{\sigma_1^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_2^2}\Big) - \frac{2}{\kappa}\mathbb{E}[x_1x_2S(\kappa)]\Big(\frac{\mathbb{E}[x_1^2S(\kappa)]}{\sigma_1^2} + \frac{\mathbb{E}[x_2^2S(\kappa)]}{\sigma_2^2}\Big) \\
&+ \frac{1}{\kappa^2}\Big(\sigma_2^2 - \frac{\mathbb{E}[x_2^2S(\kappa)]^2}{\sigma_2^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_1^2}\Big)
\end{aligned}$$

and assume $\sigma_1 = \sigma_2$. Note that $|\mathbb{E}[x_1^2S(\kappa)]| \leq \sigma_1^2$, and the same holds for x_2. It follows that, for large enough κ,

$$\frac{R(\kappa)}{\kappa^2} = \Big(\sigma_1^2 - \frac{\mathbb{E}[x_1^2S(\kappa)]^2}{\sigma_1^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma_1^2}\Big) + o(\kappa^{-1}) + o(\kappa^{-2})$$

It remains to be shown that

$$\mathbb{E}[x_1x_2S(\kappa)]^2 < 1 - \mathbb{E}[x_1^2S(\kappa)]^2$$

(To be completed) □

Lemma 1. *Let $u \sim \mathcal{D}(0,1)$ denote a random variable with mean zero and unit variance and define $h(x) : \mathbb{R} \to \mathbb{R}^+$, $v(x) : \mathbb{R} \to [-1,1]$ and $g(x) : \mathbb{R} \to \mathbb{R}$ as*

$$\begin{aligned} h(x) &= \mathbb{E}_u[|x+u|] \\ v(x) &= 1 - 2F_{\mathcal{D}}(-x) \ , \ \textit{and,} \\ g(x) &= h(x)v(x) \ . \end{aligned}$$

If $\mathcal{D}$ is symmetric about 0 and $|g'''(x)| \leq \sqrt{M}$ for all x, then

$$d(x) = (x - g(x))^2 \leq M\frac{(z^*)^6}{36}$$

Proof. First, note that if $\mathcal{D}$ is symmetric about 0, then $F_{\mathcal{D}}(0) = 1/2$ and $v(0) = g(0) = 0$, which in turn implies $d(0) = 0$. Moreover, note that

$$\begin{aligned} h(x) &= \int_{-\infty}^{\infty} (x+u)\text{sign}(x+u) f_u d(u) \\ &= \int_{-x}^{\infty} (x+u) f_u d(u) - \int_{-\infty}^{-x} (x+u) f_u d(u) \\ &= \int_{-x}^{\infty} u f_u d(u) - \int_{-\infty}^{-x} u f_u d(u) + \int_{-x}^{\infty} x f_u d(u) - \int_{-\infty}^{-x} x f_u d(u) \\ &= \int_{-x}^{\infty} u f_u d(u) - \int_{-\infty}^{-x} u f_u d(u) + x\big(1 - 2F_{\mathcal{D}}(-x)\big) \\ &= \int_{-x}^{\infty} u f_u d(u) - \int_{-\infty}^{-x} u f_u d(u) + xv(x) \\ &= 2\int_{-x}^{\infty} u f_u d(u) + xv(x). \end{aligned}$$

Hence we can write:

$$\begin{aligned} d(x) &= (x - v(x)h(x))^2 \\ &= \Big(x - 2v(x)\int_{-x}^{\infty} u f_u d(u) - xv(x)^2\Big)\Big)^2 \\ &= \Big(x(1 - v(x)^2) - 2v(x)\int_{-x}^{\infty} u f_u d(u)\Big)\Big)^2 \end{aligned}$$

Noting that $\lim_{x\to\infty} v(x)^2 = 1$ and that $\lim_{x\to\infty} \int_{-x}^{\infty} = \mathbb{E}[u] = 0$, it follows that $\lim_{x\to\infty} d(x) = 0$ Hence we know that $d(0) = 0$, $d(x) \geq 0$ for any x and $\lim_{x\to\infty} d(x) = 0$. Therefore, there is some $x^* \in \arg\max d(x)$. Taking derivatives with respect to x yields

$$d'(x) = 2(x - g(x))(g'(x) - 1)$$

Clearly, $d'(x) = 0$ at $x = g(x)$ and $g'(x) = 1$. Let $x^* = \{x : g'(x) = 1\}$ and consider a Taylor expansion of $g(0)$ around $g(x^*)$:

$$\begin{aligned} g(0) &= g(x^*) + g'(x^*)(-x^*) + \sum_{n=2}^{\infty} \frac{g^{(n)}(x^*)}{n!}(-x^*)^n \\ x^* &= g(x^*) + R_2(0) \end{aligned}$$

where the second line follows from the fact that $g'(x^*) = 1$ and $g(0) = 0$, and denoting $R_2(0) = \sum_{n=2}^{\infty} \frac{g^{(n)}(x^*)}{n!}(-x^*)^n$. It follows from Taylor's theorem that if $|g'''(x^*)| \leq \sqrt{M}$, then $|R_2(0)| \leq \sqrt{M}\frac{|x^*|^3}{6}$, which implies the result.

□