

# A new decomposition approach to modeling financial returns: Conditioning sign on magnitude\*

Arsène Brou and Richard Luger<sup>†</sup>

Université Laval, Canada

*Author accepted manuscript.*

*Accepted for publication in the Journal of Banking and Finance.*

<https://doi.org/10.1016/j.jbankfin.2026.107716>

**Abstract:** Changes in volatility contain valuable information about the likelihood of positive versus negative returns. We propose a new approach to modeling financial returns that exploits this insight by decomposing returns into sign and magnitude (absolute value) components, with magnitude closely related to volatility. The joint distribution used to compute expected returns combines a model for the marginal distribution of magnitude with a model for the distribution of the sign, conditional on the contemporaneous magnitude. Unlike traditional linear predictive regressions, this decomposition framework captures nonlinear predictability in return dynamics. An out-of-sample forecasting evaluation using monthly U.S. stock market excess returns demonstrates substantial statistical and economic gains relative to linear regression and complete subset regression, while delivering performance that is competitive with copula-based return-decomposition methods and other nonlinear benchmarks.

**JEL classification:** C13, C22, C51, G11

**Keywords:** Stock return predictability; Return decomposition; Volatility; Sign predictability; Market timing.

---

\*This work draws on research supported by the Social Sciences and Humanities Research Council of Canada.

<sup>†</sup>Correspondence to: Department of Finance, Insurance and Real Estate, Université Laval, Quebec City, QC G1V 0A6, Canada.

*E-mail addresses:* kouakou-arsene.brou.1@ulaval.ca (A. Brou), richard.luger@fsa.ulaval.ca (R. Luger).

# 1. Introduction

Traditional methods for investigating return predictability typically rely on ordinary least squares (OLS) regressions in which the conditional expectation of returns is modeled as a linear function of lagged economic variables. Yet a large body of evidence shows that such linear predictive regressions deliver poor out-of-sample performance and often fail to outperform the historical average benchmark that treats returns as essentially unpredictable (see, e.g., Borup et al., 2024; Bossaerts and Hillion, 1999; Breen et al., 1989; Butler et al., 2005; Rapach and Zhou, 2013; Welch and Goyal, 2008). This has motivated alternative approaches that aim to better exploit the information content of existing predictors rather than introducing new ones.

One prominent approach that addresses model uncertainty in return prediction is the complete subset regression (CSR) method of Elliott et al. (2013). CSR averages forecasts across all linear models formed from fixed-size subsets of the available predictors. This is particularly well suited to settings where many predictors are only weakly informative and where conventional OLS forecasts are unstable. By systematically averaging across many small models, CSR mitigates model selection risk and implicitly regularizes estimation without imposing strong parametric structure. Elliott et al. (2013) show that CSR can outperform several common benchmarks, and Bianchi et al. (2025) document that CSR remains competitive with a wide range of modern statistical learning methods in out-of-sample equity premium forecasting.

In contrast to CSR, which remains within the linear OLS framework, Anatolyev and Gospodinov (2010, hereafter AG) propose a decomposition-based approach that targets non-linear return dynamics by expressing returns as the product of a sign and a magnitude (absolute value) component. The magnitude is specified via a multiplicative error model, the sign via a binary choice model, and the two components are then coupled with a bivariate

copula. The resulting joint distribution is used to compute expected returns. This framework accommodates nonlinear structure that may be missed by linear regressions, and it has motivated a growing literature; see, among others, Liu (2015), Liu (2017), Liu and Luger (2015), Frazier and Liu (2016), Anatolyev et al. (2017), and Anatolyev and Gospodinov (2019).

In this paper, we introduce a new decomposition strategy that we call the conditioning sign on magnitude (CSM) approach. Like AG, CSM decomposes returns into sign and magnitude components. However, rather than specifying their dependence through a copula, CSM models the magnitude using a multiplicative error model and the sign using a binary choice model that conditions explicitly on the contemporaneous magnitude. This structure avoids copula selection and estimation while still allowing the predictive content of volatility states to influence expected returns.

The economic motivation for conditioning sign on magnitude stems from the well-documented persistence of volatility. When the magnitude component is large, volatility tends to remain high, and this can shift the probability of positive versus negative returns. As shown by Christoffersen and Diebold (2006), volatility dynamics can therefore induce sign predictability. The CSM approach exploits this channel by letting the sign model incorporate information embedded in the contemporaneous magnitude, which in turn improves the implied conditional mean forecast.

To benchmark the CSM approach against other economically interpretable nonlinear specifications, we also consider two widely used alternatives. First, we estimate a GARCH-in-mean model (Engle et al., 1987), in which conditional volatility enters the return equation directly. Second, we estimate a Markov-switching predictive regression with regime-dependent coefficients and variances, following the regime-switching literature (e.g., Ang and Bekaert, 2002, 2004; Hamilton, 1989; Henkel et al., 2011).

Our empirical results deliver three main messages. First, forecasting performance is

typically strongest at moderate predictor dimensions. Decomposition-based models tend to produce their most reliable out-of-sample gains when the information set is neither too small nor too large: improvements are generally more stable under absolute loss for small predictor sets, while under squared loss the gains concentrate at intermediate dimensions and can dissipate as the predictor set becomes large.

Second, formal model comparisons broadly reinforce this pattern. Across both loss functions, decomposition-based specifications are typically difficult to distinguish from the best-performing models, whereas the regime-switching benchmark is more often dominated. Overall, the evidence suggests that return decomposition captures salient nonlinearities that are missed by linear predictive regressions and by forecast averaging within the linear class.

Third, while copula-based and CSM specifications often exhibit similar statistical forecast accuracy, differences become more visible once performance is assessed through market-timing outcomes and investor utility. The CSM approach frequently delivers large economic gains relative to standard benchmarks and remains highly competitive across predictor dimensions. Copula-based strategies can occasionally match or slightly exceed CSM in some configurations, but these incremental utility gains are typically modest, suggesting that conditioning sign on magnitude provides a practical alternative to copula-based dependence modeling.

Our empirical analysis should be viewed as a small-scale application: we work with a deliberately moderate set of candidate predictors and focus on fixed-size subsets within this set. Recent studies such as Bianchi et al. (2025) and Goyal et al. (2024) consider substantially larger predictor sets. Moreover, Bianchi et al. (2025) emphasize that the optimal model size can be time-varying, whereas the CSR benchmark implemented here assumes a constant subset size throughout. Accordingly, our results demonstrate the incremental value of decomposition-based forecasting given our moderate predictor set and fixed subset size, rather than providing level comparisons to studies using larger predictor sets or time-varying

model selection.

The remainder of the paper proceeds as follows. Section 2 presents the statistical framework, including the benchmark specifications and our return decomposition approach. Section 3 presents the empirical evaluation, comparing out-of-sample forecasting performance and economic value against the historical average, linear predictive regression, CSR, and nonlinear benchmarks. Section 4 concludes. Technical details related to the copulas are provided in the Appendix and further empirical results are reported in the Supplementary material.

## 2. Statistical framework

We begin by reviewing the OLS predictive regression and CSR as linear benchmarks, then summarize the GARCH-in-mean and Markov-switching nonlinear benchmarks. Next, we introduce the sign-magnitude decomposition and a motivating example illustrating why it can capture nonlinear predictive structure missed by linear projections. We then present the AG copula-based approach and the CSM approach proposed in this paper.

Consider an asset return  $R_t$  at time  $t$  and a vector of variables  $\mathbf{X}_{t-1} = (X_{1,t-1}, \dots, X_{K,t-1})'$  observed at time  $t - 1$  that could potentially predict  $R_t$ . We treat  $R_t$  as a continuous random variable that can take both positive and negative values. The traditional method for investigating stock return predictability uses OLS to fit the linear predictive regression model

$$R_t = \beta_0 + \beta_1 X_{1,t-1} + \dots + \beta_K X_{K,t-1} + \varepsilon_t, \tag{1}$$

where the error term is assumed to satisfy  $\mathbb{E}(\varepsilon_t | \mathbf{X}_{t-1}) = 0$ , and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$  collects the slope coefficients. Let  $\text{Proj}(R_t | \mathbf{X}_{t-1})$  denote the linear projection of  $R_t$  onto the space

spanned by the constant and the components of  $\mathbf{X}_{t-1}$ . If the model in (1) is correctly specified, then  $\mathbb{E}(R_t | \mathbf{X}_{t-1}) = \text{Proj}(R_t | \mathbf{X}_{t-1}) = \beta_0 + \boldsymbol{\beta}' \mathbf{X}_{t-1}$ , and hence, for  $\mathbf{X}_{t-1} = \mathbf{x}_{t-1}$ , we have  $\mathbb{E}(R_t | \mathbf{x}_{t-1}) = [\text{Proj}(R_t | \mathbf{X}_{t-1})]_{\mathbf{X}_{t-1}=\mathbf{x}_{t-1}} = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_{t-1}$ . If  $\beta_i \neq 0$ , then  $X_{i,t-1}$  is said to have predictive ability for the mean of  $R_t$ . This predictive ability can be evaluated using in-sample or out-of-sample performance. When  $\beta_1 = \dots = \beta_K = 0$ , the model reduces to the historical average model, in which the conditional expectation becomes constant:  $\mathbb{E}(R_t | \mathbf{X}_{t-1}) = \beta_0$ , and hence  $\mathbb{E}(R_t) = \beta_0$ .

While conceptually straightforward, the OLS framework can perform poorly in empirical applications with many potential predictors. In such settings, forecasts may suffer from instability and weak out-of-sample performance due to estimation noise and model selection risk. To mitigate these issues, Elliott et al. (2013) introduce the CSR method, which averages forecasts across all linear models based on fixed-size subsets of  $k \leq K$  predictors. Formally, the CSR forecast is given by

$$\text{Proj}_{\text{CSR}}(R_t | \mathbf{x}_{t-1}) = \frac{1}{n_{k,K}} \sum_{i=1}^{n_{k,K}} \left[ \text{Proj}(R_t | S_i \mathbf{X}_{t-1}) \right]_{\mathbf{X}_{t-1}=\mathbf{x}_{t-1}},$$

where  $n_{k,K} = \binom{K}{k}$ , and  $S_i$  is a  $K \times K$  diagonal selection matrix with ones on the diagonal entries corresponding to the included variables (and zeros otherwise). In practice, the population projections are estimated by OLS on each subset, and the sample CSR forecast averages these OLS forecasts. This averaging scheme implicitly regularizes estimation and accounts for model uncertainty. When  $k = 1$ , CSR reduces to the forecast combination method of Rapach et al. (2010); when  $k = K$ , it becomes the full linear predictive regression model in (1).

As in Elliott et al. (2013), we implement CSR for a fixed subset size over the out-of-sample period. This differs from Bianchi et al. (2025), who emphasize that the optimal model size may vary over time and consider adaptive choices of the subset size. We keep the subset size

fixed to maintain comparability across models and to isolate the incremental role of return decomposition.

We next describe the two benchmark nonlinear specifications used in the empirical comparison. The GARCH-in-mean (GARCH-M) model (Engle et al., 1987) allows expected returns to depend on conditional volatility. Specifically,

$$R_t = \beta_0 + \boldsymbol{\beta}' \mathbf{X}_{t-1} + \lambda \sigma_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2),$$

where  $\sigma_t^2$  follows a standard GARCH(1,1) recursion,  $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta_g \sigma_{t-1}^2$ , with  $\omega > 0$ ,  $\alpha \geq 0$ ,  $\beta_g \geq 0$ , and  $\alpha + \beta_g < 1$ . The parameter  $\lambda$  captures the sensitivity of expected returns to time-varying risk. This specification aligns with the view emphasized by Cederburg et al. (2023) that volatility is an important state variable for equity premium predictability and the economic value of forecasting. The model is estimated via quasi-maximum likelihood.

To capture shifts in return dynamics, we also estimate a two-state Markov-switching (MS) predictive regression in the spirit of Hamilton (1989), Ang and Bekaert (2002, 2004), and Henkel et al. (2011). Returns depend on an unobserved state variable  $c_t \in \{1, 2\}$  that evolves as a first-order Markov chain. We specify

$$R_t = \beta_{0,c_t} + \boldsymbol{\beta}'_{c_t} \mathbf{X}_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{c_t}^2),$$

where  $\beta_{0,c_t}$ ,  $\boldsymbol{\beta}_{c_t} = (\beta_{1,c_t}, \dots, \beta_{K,c_t})'$ , and  $\sigma_{c_t}^2$  are regime-specific parameters. The transition probabilities  $\Pr(c_t = j \mid c_{t-1} = i) = p_{ij}$ ,  $i, j \in \{1, 2\}$ , govern movements between regimes. This specification allows the intercept, slope coefficients, and error variance to vary across regimes. We estimate the model by maximum likelihood using the expectation-maximization (EM) algorithm.<sup>1</sup>

---

<sup>1</sup>In our empirical application, we implement the MS model using the R package `MSwM` (Sanchez-Espigares and Lopez-Moreno, 2021).

The return decomposition approach offers an alternative nonlinear perspective to (1) by first separating  $R_t$  into distinct components. This approach is based on the mathematical identity

$$R_t = \text{sign}(R_t)|R_t|, \tag{2}$$

where  $\text{sign}(R_t)$  and  $|R_t|$  are the sign and magnitude (or absolute value) components of  $R_t$ , respectively. Note that there is no loss of information in this decomposition. Following AG, we apply this decomposition directly to raw returns rather than mean-adjusted returns; Christoffersen and Diebold (2006) show that removing unconditional means before decomposing can disguise sign predictability, providing the rationale for this choice.

The conditional mean of  $R_t$ , given that  $\mathbf{X}_{t-1} = \mathbf{x}_{t-1}$ , can then be expressed as

$$\mathbb{E}(R_t \mid \mathbf{x}_{t-1}) = \mathbb{E}(\text{sign}(R_t) |R_t| \mid \mathbf{x}_{t-1}).$$

The objective of decomposing  $R_t$  as in (2) is to obtain  $\mathbb{E}(R_t \mid \mathbf{x}_{t-1})$  by modeling the joint conditional distribution of  $\text{sign}(R_t)$  and  $|R_t|$ , given the predictors. Next, we present a simple example that illustrates why such a decomposition-based approach can be advantageous compared to OLS and its extensions, including CSR, which relies on linear projections as its foundational building blocks.

## 2.1. Motivating example

This toy example illustrates how linear predictive regression models can fail to capture fundamental nonlinear dynamics, such as those arising from the decomposition in (2). While this identity holds for any real-valued random variable, linear models may fail to exploit the structure it reveals, thereby missing essential aspects of the data-generating process.

Assume that the asset return is related to two predictor variables via the nonlinear

relationship

$$R_t = \text{sign}(X_{1,t-1}) |X_{2,t-1}|,$$

where  $X_{1,t-1}$  and  $X_{2,t-1}$  are independent random variables, each following an  $N(0, 1)$  distribution. This independence assumption is purely for expositional simplicity in this toy example; the CSM framework developed below does not require the predictors to be independent.

Economically, this stylized structure can be viewed as separating direction and magnitude in return dynamics. The variable  $X_{1,t-1}$  summarizes information that tilts the market toward “good” or “bad” states, and its sign determines whether the return is positive or negative. In contrast,  $X_{2,t-1}$  can be interpreted as a magnitude (or “volatility-state”) signal:  $|X_{2,t-1}|$  determines how large the return move is. The realized return is then the product of a directional signal and a magnitude signal.

The return is fully predictable because the population conditional expectation satisfies

$$\mathbb{E}(R_t | X_{1,t-1}, X_{2,t-1}) = \text{sign}(X_{1,t-1}) |X_{2,t-1}|,$$

which equals  $R_t$  with probability one, since  $R_t$  is a deterministic function of its predictors. This confirms that both  $X_{1,t-1}$  and  $X_{2,t-1}$  play an essential role in predicting  $R_t$ .

If, however, one assumes the linear relationship

$$R_t = \beta_1 X_{1,t-1} + \beta_2 X_{2,t-1} + \varepsilon_t$$

with  $\mathbb{E}(\varepsilon_t | X_{1,t-1}, X_{2,t-1}) = 0$  and fits the model using OLS, then only  $X_{1,t-1}$  will have a nonzero coefficient in the population regression. Indeed, the population OLS coefficient  $\beta_1$

is given by  $\beta_1 = \mathbb{E}(R_t X_{1,t-1})$ . Substituting the definition of  $R_t$ , we obtain

$$\beta_1 = \mathbb{E}[\text{sign}(X_{1,t-1}) |X_{2,t-1}| X_{1,t-1}].$$

Observe that  $\text{sign}(X_{1,t-1})X_{1,t-1} = |X_{1,t-1}|$ , so the expression simplifies to

$$\beta_1 = \mathbb{E}[|X_{1,t-1}| |X_{2,t-1}|].$$

Because  $X_{1,t-1}$  and  $X_{2,t-1}$  are independent standard normal variables, the expectation factorizes:

$$\beta_1 = \mathbb{E}(|X_{1,t-1}|)\mathbb{E}(|X_{2,t-1}|) = \left(\sqrt{\frac{2}{\pi}}\right)^2 = \frac{2}{\pi}.$$

On the other hand, the contribution of  $X_{2,t-1}$  is overlooked, since the population OLS coefficient  $\beta_2$  is

$$\beta_2 = \mathbb{E}(R_t X_{2,t-1}) = \mathbb{E}(\text{sign}(X_{1,t-1}))\mathbb{E}(|X_{2,t-1}|X_{2,t-1}) = 0,$$

due to the independence of  $X_{1,t-1}$  and  $X_{2,t-1}$  and the symmetry of  $X_{1,t-1}$ , which implies  $\mathbb{E}(\text{sign}(X_{1,t-1})) = 0$ .

To quantify the consequences of this misspecification, we compare the population mean squared error (MSE) of the OLS predictor to that of the true conditional expectation. Since  $R_t = \text{sign}(X_{1,t-1}) |X_{2,t-1}|$  by construction, the oracle predictor  $\mathbb{E}(R_t | X_{1,t-1}, X_{2,t-1})$  perfectly recovers  $R_t$ , resulting in  $\text{MSE}_{\text{oracle}} = 0$ . In contrast, the population OLS predictor corresponds to the best linear approximation in mean square, given by

$$\text{Proj}(R_t | X_{1,t-1}, X_{2,t-1}) = \frac{2}{\pi} X_{1,t-1},$$

which fails to capture the nonlinear interaction between the sign and magnitude components.

The resulting population MSE is strictly positive:

$$\text{MSE}_{\text{OLS}} = \mathbb{E} \left[ \left( R_t - \frac{2}{\pi} X_{1,t-1} \right)^2 \right] = 1 - \frac{4}{\pi^2} \approx 0.595.$$

This demonstrates that although OLS correctly identifies  $X_{1,t-1}$  as having predictive power, it fails to recover the true nonlinear structure and overlooks the essential role of  $X_{2,t-1}$ , resulting in a substantial loss of predictive accuracy.

This example highlights a key limitation of linear predictive models: their inability to capture interactions between sign and magnitude that are central to return dynamics. While the CSR method improves upon OLS by averaging over many low-dimensional linear models to reduce estimation risk, it remains confined to the linear span of predictors and cannot recover the nonlinear structure illustrated here. More generally, when return dynamics involve distinct directional and scale components, purely linear predictive models may miss economically meaningful structure. This limitation motivates the use of decomposition-based models, notably the proposed CSM approach; see Anatolyev et al. (2017) for another illustrative example.

To preview the empirical relevance of this decomposition, we turn to the data used in our application. Table 1 reports simple correlations between lagged predictors and the realized return  $r_t$ , as well as with its sign and magnitude components, denoted  $s_t$  and  $m_t$ , respectively. Although these unconditional correlations are generally modest, several predictors exhibit stronger associations with one of the two components than with  $r_t$  itself. For example, the short-term interest rate  $tbl_{t-1}$  is more strongly correlated with the sign component ( $-0.143$ ) than with returns ( $-0.097$ ), whereas credit-related variables such as  $dfy_{t-1}$  and  $dfr_{t-1}$  display the strongest correlations with the magnitude component ( $0.166$  and  $-0.146$ , respectively). These patterns suggest that modeling directional and scale information jointly, rather than forcing both channels through a single linear conditional-mean

relation, may capture economically meaningful structure.

## 2.2. Copula-based decomposition approach

In this subsection, we present the AG copula-based decomposition approach. This method builds upon (2) by first rewriting it as

$$R_t = M_t(2S_t - 1), \quad (3)$$

where  $M_t = |R_t|$  is a non-negative random variable and  $S_t = \frac{1}{2}(\text{sign}(R_t) + 1) = \mathbb{1}\{R_t > 0\}$  is a Bernoulli random variable. Here  $\mathbb{1}\{\cdot\}$  is the indicator function. In the decomposition approach,  $M_t$  and  $S_t$  are treated as the dependent variables. AG specify a multiplicative error model (MEM) for the “volatility” component  $M_t$  (since under a scale model  $R_t = \sigma_t Z_t$ , we have  $|R_t| = \sigma_t |Z_t|$ , so the absolute return is proportional to volatility; Granger and Sin 2000), a binary choice model for the direction-of-change component  $S_t$ , and then construct their joint distribution by using a bivariate copula that links the two marginal distributions.

### 2.2.1. Marginal distributions

Following the framework of Engle (2002), AG specify the dynamics of  $M_t \geq 0$  to be an MEM model with a Weibull distribution for the error term. In the present context with  $\mathbf{X}_{t-1}$  comprising the driving variables, the model for the magnitude component in (3) becomes

$$\begin{aligned} M_t &= \psi_t \eta_t, \\ \log(\psi_t) &= w_v + \boldsymbol{\delta}'_v \mathbf{x}_{t-1}, \end{aligned} \quad (4)$$

where  $\psi_t = \mathbb{E}(M_t | \mathbf{x}_{t-1})$  and the error term  $\eta_t \geq 0$  follows a Weibull distribution, scaled to ensure  $\mathbb{E}(\eta_t | \mathbf{x}_{t-1}) = 1$ . AG consider a richer specification that includes dynamic autoregressive terms in addition to the predictor variables. However, for our comparison with

the traditional OLS predictive regression, we focus on specifications that use only the same predictor variables as in (1). Under (4), the cumulative distribution function (CDF) of  $M_t$  becomes

$$F_{M_t|\mathbf{x}_{t-1}}(m|\mathbf{x}_{t-1}) = 1 - \exp\left(-\left(\frac{m}{\psi_t}\Gamma(1+\kappa^{-1})\right)^\kappa\right), \quad (5)$$

with corresponding probability density function

$$f_{M_t|\mathbf{x}_{t-1}}(m|\mathbf{x}_{t-1}) = \kappa\psi_t^{-\kappa}\Gamma(1+\kappa^{-1})^\kappa m^{\kappa-1} \exp\left(-\left(\frac{m}{\psi_t}\Gamma(1+\kappa^{-1})\right)^\kappa\right), \quad (6)$$

where  $\Gamma(\cdot)$  is the gamma function and  $\kappa > 0$  is a shape parameter. We stack the parameters on which (5) and (6) depend into the vector  $\boldsymbol{\theta}_v = (w_v, \boldsymbol{\delta}'_v, \kappa)'$ .

The conditional distribution of  $S_t$  given  $\mathbf{x}_{t-1}$  is of course Bernoulli  $\mathcal{B}(p_t)$  with probability mass function (PMF)  $f_{S_t|\mathbf{x}_{t-1}}(v|\mathbf{x}_{t-1}) = p_t^v(1-p_t)^{1-v}$ ,  $v \in \{0, 1\}$  and cumulative mass function  $F_{S_t|\mathbf{x}_{t-1}} = 1 - p_t(1-v)$ , where  $p_t = \Pr(R_t > 0 | \mathbf{x}_{t-1})$ . Following Nyberg (2011) and Pönkä (2017), we parameterize  $p_t$  as a predictive probit model of the form

$$\begin{aligned} p_t &= \Phi(\theta_t), \\ \theta_t &= w_d + \boldsymbol{\delta}'_d \mathbf{x}_{t-1}, \end{aligned} \quad (7)$$

where  $\Phi(\cdot)$  is the CDF of the standard normal distribution, and we let  $\boldsymbol{\theta}_d = (w_d, \boldsymbol{\delta}'_d)'$  regroup the associated parameters.<sup>2</sup> AG consider a more general binary choice model that allows for own dynamics. As already mentioned, our focus here is on specifications that use only the predictor variables appearing in (1), as we aim to investigate the channels through which their predictive ability operates.

---

<sup>2</sup>We prefer probit over logit as it is conformable with the Gaussian copula (Anatolyev and Gospodinov, 2019).

### 2.2.2. Joint distribution

In order to construct the bivariate distribution of  $Y_t = (M_t, S_t)'$ , the AG approach appeals to the theory of copulas. Specifically, a conditional meta-distribution is created as

$$F_{Y_t|\mathbf{X}_{t-1}}(u, v | \mathbf{x}_{t-1}) = C(F_{M_t|\mathbf{X}_{t-1}}(u | \mathbf{x}_{t-1}), F_{S_t|\mathbf{X}_{t-1}}(v | \mathbf{x}_{t-1})), \quad (8)$$

where  $C(w_1, w_2)$  is a copula distribution function defined on  $[0, 1]^2$ , with dependency parameter  $\theta_c$ . Upon differentiation of (8), AG obtain the joint conditional density/mass function of  $M_t$  and  $S_t$  as

$$f_{Y_t|\mathbf{X}_{t-1}}(u, v | \mathbf{x}_{t-1}) = \varrho_t \left( F_{M_t|\mathbf{X}_{t-1}}(u | \mathbf{x}_{t-1}) \right)^v \left( 1 - \varrho_t \left( F_{M_t|\mathbf{X}_{t-1}}(u | \mathbf{x}_{t-1}) \right) \right)^{1-v} f_{M_t|\mathbf{X}_{t-1}}(u | \mathbf{x}_{t-1}), \quad (9)$$

where  $\varrho_t(z) = 1 - \partial C(z, 1 - p_t) / \partial w_1$  represents a “deformed” probability mass of  $S_t$  whose probability of success is given by  $\varrho_t(F_{M_t|\mathbf{X}_{t-1}}(u | \mathbf{x}_{t-1}))$ .

The bivariate copulas used in the empirical application are the Gaussian, Frank, Clayton, and Farlie-Gumbel-Morgenstern (FGM) copulas, following AG and Anatolyev and Gospodinov (2019). Since these copulas are well known, the Appendix provides only their definitions and the corresponding expressions for the deformed probability of success,  $\varrho_t(z)$ ; see Trivedi and Zimmer (2007) for more detailed discussions.

### 2.2.3. Conditional mean prediction and parameter estimation

The decomposition in (3) implies that the conditional mean of  $R_t$  can be written as

$$\mathbb{E}(R_t | \mathbf{x}_{t-1}) = 2\xi_t - \psi_t,$$

where  $\xi_t = \mathbb{E}(M_t S_t | \mathbf{x}_{t-1})$  is the expected cross-product of  $M_t$  and  $S_t$ , and  $\psi_t = \mathbb{E}(M_t | \mathbf{x}_{t-1})$  is given in (4). The cross-product term is given by AG as

$$\xi_t = \int_0^1 F_{M_t | \mathbf{x}_{t-1}}^{-1}(z | \mathbf{x}_{t-1}) \varrho_t(z) dz, \quad (10)$$

where  $F_{M_t | \mathbf{x}_{t-1}}^{-1}(z | \mathbf{x}_{t-1}) = \psi_t \Gamma(1 + \kappa^{-1})^{-1} (-\log(1 - z))^{1/\kappa}$  is the conditional quantile function of  $M_t$ . For practical applications, we evaluate the above integral by Monte Carlo integration (Robert and Casella, 2010, Ch. 3).

The complete vector of model parameters is  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_v, \boldsymbol{\theta}'_d, \theta_c)'$ , comprising: (i)  $\boldsymbol{\theta}_v = (w_v, \boldsymbol{\delta}'_v, \kappa)'$ , the parameters of the marginal distribution of  $M_t$  in (4); (ii)  $\boldsymbol{\theta}_d = (w_d, \boldsymbol{\delta}'_d)'$ , the parameters of the marginal distribution of  $S_t$  in (7); and (iii)  $\theta_c$ , the parameter of the copula distribution in (8). Given the sample of returns  $r_1, \dots, r_T$ , which upon decomposition yields the realizations  $(m_1, s_1), \dots, (m_T, s_T)$ , along with the realized values of the predictor variables, the full log-likelihood function can be computed as

$$L(\boldsymbol{\theta}) = L_v(\boldsymbol{\theta}_v) + L_c(\boldsymbol{\theta}_v, \boldsymbol{\theta}_d, \theta_c) \quad (11)$$

with

$$L_v(\boldsymbol{\theta}_v) = \sum_{t=1}^T \log f_{M_t | \mathbf{x}_{t-1}}(m_t | \mathbf{x}_{t-1})$$

and

$$L_c(\boldsymbol{\theta}_v, \boldsymbol{\theta}_d, \theta_c) = \sum_{t=1}^T s_t \log \varrho_t(u_t) + (1 - s_t) \log (1 - \varrho_t(u_t)),$$

where  $u_t = F_{M_t | \mathbf{x}_{t-1}}(m_t | \mathbf{x}_{t-1})$ . The maximum likelihood estimates (MLE) are then obtained as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}_v, \boldsymbol{\theta}_d, \theta_c} L(\boldsymbol{\theta}),$$

using standard numerical optimization methods (Judd, 1998, Ch. 4).

The estimation can be simplified by employing the method of inference for margins (IFM) of Shih and Louis (1995) and Joe and Xu (1996). The IFM method in the present context consists of first estimating  $\boldsymbol{\theta}_v$  by computing  $\hat{\boldsymbol{\theta}}_v = \arg \max_{\boldsymbol{\theta}_v} L_v(\boldsymbol{\theta}_v)$ , and then obtaining estimates of the remaining parameters as

$$(\hat{\boldsymbol{\theta}}_d, \hat{\theta}_c) = \arg \max_{\boldsymbol{\theta}_d, \theta_c} L_c(\hat{\boldsymbol{\theta}}_v, \boldsymbol{\theta}_d, \theta_c),$$

in which the first-step estimates  $\hat{\boldsymbol{\theta}}_v$  are taken as given. This procedure is computationally simpler than estimating all the parameters in  $\boldsymbol{\theta}$  simultaneously. The drawback of the IFM approach is that there is some loss of statistical efficiency relative to the full MLE because the first step ignores the dependency of  $L_c(\boldsymbol{\theta}_v, \boldsymbol{\theta}_d, \theta_c)$  on  $\boldsymbol{\theta}_v$  when finding  $\hat{\boldsymbol{\theta}}_v$ .

## 2.3. CSM decomposition approach

### 2.3.1. Factorization and its motivation

The CSM approach constructs the joint density of  $Y_t = (M_t, S_t)'$  by conditioning the distribution of  $S_t$  on both  $\boldsymbol{x}_{t-1}$  and the contemporaneous value of the return magnitude. Specifically, the CSM approach proceeds via the factorization

$$f_{Y_t|\boldsymbol{X}_{t-1}}(m_t, s_t | \boldsymbol{x}_{t-1}) = f_{S_t|M_t, \boldsymbol{X}_{t-1}}(s_t | m_t, \boldsymbol{x}_{t-1}) f_{M_t|\boldsymbol{X}_{t-1}}(m_t | \boldsymbol{x}_{t-1}), \quad (12)$$

where  $f_{M_t|\boldsymbol{X}_{t-1}}(m_t | \boldsymbol{x}_{t-1})$  is given by (6), and the distribution  $f_{S_t|M_t, \boldsymbol{X}_{t-1}}(s_t | m_t, \boldsymbol{x}_{t-1})$  is conditional on  $M_t = m_t$ . This way of expressing a joint distribution of two random variables as the product of a conditional distribution and a marginal distribution is a standard result in probability theory; it also implicitly defines a copula linking the two components (Sklar, 1959). Since the magnitude component is modeled identically in both (9) and (12), any performance disparities between the copula-based and CSM approaches can be attributed

solely to differences in how the sign-magnitude dependence structure is modeled.

Why should we prefer conditioning the sign on the magnitude as in (12), instead of the magnitude on the sign? There are at least three reasons. First, the magnitude is more likely to be predictive of the sign. Indeed, there is overwhelming evidence of GARCH-type volatility clustering in financial markets, where periods of high volatility (large  $m_t$ ) are followed by more high-volatility periods. Christoffersen and Diebold (2006) demonstrate that these volatility dynamics can lead to predictability in the sign of returns, even in the absence of mean predictability. This means that as  $m_t$  changes, so too does the probability of observing a positive return.

Second, behavioral finance insights support this approach. Due to differences in the timing of information receipt, information processing, behavioral biases, and feedback trading, Treynor and Ferguson (1985), Brown and Jennings (1989), Hong and Stein (1999), Cespa and Vives (2012), and Edmans et al. (2015), among others, show that past stock prices can lead to predictable patterns in returns. For instance, large values of  $m_t$  can trigger specific investor behaviors: significant negative returns might lead to panic selling, increasing the likelihood of subsequent negative returns, while large positive returns might lead to profit-taking or increased buying pressure, influencing the sign of the next return.

Third, in econometrics, conditioning on a variable that explains more variation in the dependent variable usually leads to better model performance. Since the magnitude exerts a greater influence on the sign of the return than vice versa, modeling  $S_t$  given  $m_t$  is likely to improve the accuracy and reliability of our predictions. Related empirical evidence supporting the usefulness of return volatility in forecasting the direction of asset returns is reported in Christoffersen et al. (2007), Bekiros and Georgoutsos (2008), Chevapatrakul (2013), Algieri and Leccadito (2019), and Campisi et al. (2024).

### 2.3.2. CSM model

In the CSM approach, the specification for the magnitude component  $M_t$  remains as described in (4), with the corresponding density function  $f_{M_t|\mathbf{X}_{t-1}}(m_t | \mathbf{x}_{t-1})$  given in (6). Importantly, the conditional distribution of the sign component  $S_t$  becomes Bernoulli  $\mathcal{B}(p_t^*)$ , where  $p_t^* = \Pr(R_t > 0 | m_t, \mathbf{x}_{t-1})$  now conditions on  $m_t$  in addition to  $\mathbf{x}_{t-1}$ . We specify  $p_t^*$  as

$$\begin{aligned} p_t^* &= \Phi(\theta_t^*), \\ \theta_t^* &= w_d + \boldsymbol{\delta}'_d \mathbf{x}_{t-1} + \beta m_t, \end{aligned} \tag{13}$$

so that  $\beta$  captures the effect of the realized magnitude  $m_t$  on  $\text{sign}(R_t)$ . The corresponding PMF is  $f_{S_t|M_t, \mathbf{X}_{t-1}}(v | m_t, \mathbf{x}_{t-1}) = p_t^{*v} (1 - p_t^*)^{1-v}$ , for  $v \in \{0, 1\}$ .

With this approach, the conditional mean of  $R_t$  given that  $\mathbf{X}_{t-1} = \mathbf{x}_{t-1}$  becomes

$$\mathbb{E}(R_t | \mathbf{x}_{t-1}) = 2\xi_t^* - \psi_t, \tag{14}$$

where  $\psi_t = \mathbb{E}(M_t | \mathbf{x}_{t-1})$  is given in (4), and the term  $\xi_t^* = \mathbb{E}(M_t S_t | \mathbf{x}_{t-1})$  is obtained via iterated expectations as

$$\begin{aligned} \mathbb{E}(M_t S_t | \mathbf{x}_{t-1}) &= \mathbb{E}(M_t \mathbb{E}(S_t | M_t, \mathbf{x}_{t-1}) | \mathbf{x}_{t-1}) \\ &= \int_0^{+\infty} m \Phi(w_d + \boldsymbol{\delta}'_d \mathbf{x}_{t-1} + \beta m) f_{M_t|\mathbf{X}_{t-1}}(m | \mathbf{x}_{t-1}) dm, \end{aligned}$$

which must be evaluated numerically given the absence of a closed-form solution. Upon the change of variable  $v = F_{M_t|\mathbf{X}_{t-1}}(m | \mathbf{x}_{t-1})$ , this integral can be rewritten as

$$\xi_t^* = \int_0^1 q_t(v) \Phi(w_d + \boldsymbol{\delta}'_d \mathbf{x}_{t-1} + \beta q_t(v)) dv, \tag{15}$$

where  $q_t(v) = F_{M_t|\mathbf{X}_{t-1}}^{-1}(v | \mathbf{x}_{t-1})$  is the conditional quantile function of  $M_t$  appearing in (10).

Here again we use Monte Carlo integration to evaluate (15), which then serves to compute (14).

The parameters entering the CSM model specification are  $\boldsymbol{\theta}^* = (\boldsymbol{\theta}'_v, \boldsymbol{\theta}'_d)'$ , where  $\boldsymbol{\theta}_d^* = (w_d, \boldsymbol{\delta}'_d, \beta)'$ . In light of (12), the sample log-likelihood function is expressed as

$$L(\boldsymbol{\theta}^*) = L_v(\boldsymbol{\theta}_v) + L_d(\boldsymbol{\theta}_d^*), \quad (16)$$

where  $L_v(\boldsymbol{\theta}_v)$  is given in (11) and  $L_d(\boldsymbol{\theta}_d^*) = \sum_{t=1}^T s_t \log p_t^* + (1 - s_t) \log(1 - p_t^*)$ . The parameter estimates needed to make the CSM model operational can be computed as  $\hat{\boldsymbol{\theta}}^* = \arg \max_{\boldsymbol{\theta}^*} L(\boldsymbol{\theta}^*)$ . Since  $L_v(\boldsymbol{\theta}_v)$  and  $L_d(\boldsymbol{\theta}_d^*)$  have no common parameters, the values  $\hat{\boldsymbol{\theta}}^* = (\hat{\boldsymbol{\theta}}'_v, \hat{\boldsymbol{\theta}}'_d)'$  that maximize (16) can be equivalently found by computing  $\hat{\boldsymbol{\theta}}_v = \arg \max_{\boldsymbol{\theta}_v} L_v(\boldsymbol{\theta}_v)$  and  $\hat{\boldsymbol{\theta}}_d^* = \arg \max_{\boldsymbol{\theta}_d^*} L_d(\boldsymbol{\theta}_d^*)$ , separately, with no loss in statistical efficiency.

We also note that the maximization of  $L_d(\boldsymbol{\theta}_d^*)$  is not sensitive to the choice of starting values. For probit models with a log-likelihood function of that form, the Hessian matrix is negative definite for all values of  $\boldsymbol{\theta}_d^*$ , ensuring global concavity and a unique maximum (see, e.g., Maddala, 1983, p. 26). As a result, a Newton-Raphson procedure will typically converge to the unique maximum likelihood estimator from any reasonable starting point.

The baseline specification in (13) can be extended to allow for more flexible modeling, including interaction terms between  $m_t$  and elements of  $\boldsymbol{x}_{t-1}$ , as well as nonlinear effects of  $m_t$  itself. To illustrate the latter, we consider a polynomial specification in which the index function

$$\theta_t^* = w_d + \boldsymbol{\delta}'_d \boldsymbol{x}_{t-1} + \beta m_t + \beta_2 m_t^2 + \beta_3 m_t^3$$

includes higher-order terms in  $m_t$ . The additional components  $\beta_2 m_t^2$  and  $\beta_3 m_t^3$  introduce curvature into the probit index, allowing the probability of a positive return to respond nonlinearly to changes in return magnitude. This extension increases the model's flexibility while maintaining parsimony. We refer to this nonlinear index specification as the CSM

(Poly) model, whereas the version in which  $\theta_t^*$  is linear in  $(\mathbf{x}_{t-1}, m_t)$  is referred to as the CSM (Baseline) model.

While more flexible functional forms, such as splines or kernel-based methods, could model nonlinearities in the sign-magnitude relationship, they typically require tuning and regularization to prevent overfitting. In contrast, the polynomial expansion used in the CSM (Poly) model offers a simple, interpretable alternative that captures moderate nonlinearities without the computational burden of nonparametric methods. Finally, one could also augment (4) and (13) with lagged values of  $m_t$ ,  $\psi_t$ ,  $s_t$ , and  $\theta_t^*$ , to get dynamic autoregressive specifications; see Kauppi and Saikkonen (2008), Anatolyev and Gospodinov (2010), Nyberg (2011), Liu and Luger (2015), Anatolyev et al. (2017), and Pönkä (2017) for examples. We explored such extensions but found that none consistently outperformed the baseline CSM model, in line with AG’s finding that forecasting gains stem primarily from predictor variables rather than autoregressive structures (p. 240).

**Remark.** (CSM vs. copula-based decomposition: Two non-nested models). Although both the CSM and copula-based decomposition models build on the decomposition in (2), they represent distinct and, under the parametric specifications we consider, non-nested approaches. The copula-based framework models the marginals of  $M_t$  and  $S_t$  separately (each conditional on past information) and then captures their contemporaneous dependence via a bivariate copula. Because the copula operates on the marginal distributions, any dependence of the sign on the magnitude is induced implicitly through the joint distribution, rather than through explicit conditioning on  $m_t$  in the sign equation. In contrast, the CSM model embeds dependence by conditioning the sign directly on the contemporaneous magnitude, without requiring a copula specification for the joint distribution. The two approaches coincide only under conditional independence: in the copula framework, this corresponds to using the independence copula; in CSM, this corresponds to excluding  $m_t$  and any function of it from the probit equation.

### 3. Empirical application

In this section, we examine the predictability of monthly excess returns  $r_t$  on the S&P 500 value-weighted index. We begin with ten financial predictors that are widely used in the return predictability literature, including several proposed by Welch and Goyal (2008): the log dividend-price ratio ( $dp$ ), the log earnings-price ratio ( $ep$ ), the book-to-market ratio ( $btm$ ), the default yield spread ( $dfy$ ), the term spread ( $tms$ ), the short-term interest rate ( $tbl$ ), the long-term government bond return ( $ltr$ ), the default return spread ( $dfr$ ), net equity expansion ( $ntis$ ), and inflation ( $infl$ ). These variables are a subset of those in Welch and Goyal (2008), updated through 2021. The monthly data span January 1948 to December 2021 and are obtained from Amit Goyal’s website.

To avoid multicollinearity and improve model stability, we exclude two predictors ( $ep$  and  $btm$ ) due to their high correlation (above 0.80) with other variables. The final predictor set thus includes eight variables.<sup>3</sup> We deliberately work with a moderate set of predictors to facilitate transparent subset selection and interpretation. This design differs from studies using substantially larger predictor sets (e.g., Bianchi et al., 2025; Goyal et al., 2024), which can change the absolute magnitude of economic performance measures and may increase the scope for gains from flexible nonlinear methods.

Table 1 reports sample correlations between lagged predictors and excess returns, as well as the sign and magnitude components of returns. Panel A shows that correlations between  $r_t$  and the lagged predictors are generally weak, with the largest absolute correlation occurring for the short-term interest rate  $tbl_{t-1}$  ( $-0.097$ ). Panel B shows that several predictors exhibit stronger associations with individual components than with returns themselves. In particular,  $tbl_{t-1}$  exhibits the strongest correlation with the sign component  $s_t$  ( $-0.143$ ), while

---

<sup>3</sup>We also consider an extended specification that adds a time-series momentum indicator following Moskowitz et al. (2012). Section D of the Supplementary material shows that the main conclusions are unchanged.

the default yield spread  $dfy_{t-1}$  shows the strongest correlation with the magnitude component  $m_t$  (0.166), followed closely by  $dfr_{t-1}$  ( $-0.146$ ). Panel C reports correlations among predictors within the final eight-variable set; the largest absolute correlation is about 0.50, suggesting no evidence of severe multicollinearity in the specification used in the empirical analysis.

To ensure a fair comparison across subset sizes  $k \in \{1, \dots, 8\}$ , we fix the number of predictors at  $k$  for all methods and, for methods that involve subset selection, use a common in-sample selection criterion. We split the sample into two parts after aligning lagged predictors and returns: the first 400 monthly return observations (February 1948 to May 1981) are used for in-sample predictor selection (when applicable) and model estimation, consistent with AG’s rolling-window design, and the remaining 487 observations are reserved for out-of-sample forecast evaluation.

For each  $k$ , we enumerate all  $n_{k,8} = \binom{8}{k}$  candidate predictor combinations using the first 400 observations and generate a sequence of one-step-ahead in-sample forecasts for each combination. Methods that select a best  $k$ -predictor subset (the linear model, GARCH-M, MS, the AG copula-based models, and the CSM models) retain the single  $k$ -predictor combination that maximizes directional accuracy. Directional accuracy is measured by the area under the (correct classification) curve (AUC), following Jordà and Taylor (2011, 2012). This choice is motivated by the fact that the limited success in return predictability is largely attributable to the difficulty of forecasting the sign of returns, rather than their magnitude. Since the AUC captures the ability of forecasts to discriminate between positive- and negative-return periods, it provides a natural and informative criterion for predictor selection in this context. In contrast, the CSR approach does not select a single best subset; for each  $k$ , it forms forecasts by averaging across all linear regressions that use exactly  $k$  predictors drawn from the full set of eight candidates (Elliott et al., 2013). The historical average serves as an additional benchmark.

Let  $\hat{r}_t$  denote the one-step-ahead in-sample forecast of return  $r_t$ . Define  $N_n = \sum_{t=1}^{400} \mathbb{1}\{\text{sign}(r_t) = -1\}$ , and let  $v_j$  be the value of  $\hat{r}_t$  when  $\text{sign}(r_t) = -1$ , for  $j = 1, \dots, N_n$ . Similarly, define  $N_p = \sum_{t=1}^{400} \mathbb{1}\{\text{sign}(r_t) = 1\}$ , and let  $u_i$  be the value of  $\hat{r}_t$  when  $\text{sign}(r_t) = 1$ , for  $i = 1, \dots, N_p$ . The AUC is then computed as

$$\text{AUC} = \frac{1}{N_n N_p} \sum_{j=1}^{N_n} \sum_{i=1}^{N_p} \left( \mathbb{1}\{v_j < u_i\} + \frac{1}{2} \mathbb{1}\{v_j = u_i\} \right),$$

where the second term serves to break ties. The AUC estimates  $\Pr(v < u)$ , which equals 0.5 when the forecasts are uninformative and approaches 1 when they assign systematically higher values to positive-return periods than to negative-return periods. See Jordà and Taylor (2011) for further discussion.

As a robustness check, we also perform an alternative best  $k$ -predictor subset selection using the MSE criterion. Although the resulting subsets differ from those based on AUC, they overlap substantially for most models and subset sizes. Out-of-sample evaluation shows that MSE-based selection underperforms relative to AUC-based selection, particularly for the linear model. We therefore proceed with the AUC-based subsets for the remainder of the analysis. Additional details and results for the MSE-based selection are provided in Section A of the Supplementary material.

Table 2 reports the variables selected in the best subset for each value of  $k$ . Panel A shows the linear and GARCH-M models, which select identical subsets at all values of  $k$ . Panel B shows the MS model, and Panel C shows the decomposition-based models (CSM and copula-based), for which selection results are identical across all copula specifications.

Across models, there is substantial overlap in variable selection. The short-term interest rate ( $tbl_{t-1}$ ) is consistently chosen at  $k = 1$  in all specifications, reflecting its strong predictive power for return direction. This variable remains in every selected subset for the linear and decomposition-based models from  $k = 1$  to  $k = 8$ . In the MS model,  $tbl_{t-1}$  is selected for

$k = 1-3$  and again from  $k = 6$  onward.

Moreover, the linear and decomposition-based models produce identical subsets for most values of  $k$ , differing only at  $k = 2$ ,  $k = 5$ , and  $k = 6$ , where a single variable accounts for the discrepancy. Notably,  $dp_{t-1}$  enters the decomposition-based models earlier (at  $k = 5$ ) than in the linear model (at  $k = 7$ ), which is consistent with decomposition-based models exploiting predictive content from valuation ratios at smaller subset sizes. The MS model exhibits greater variation at small  $k$ , with inflation ( $infl_{t-1}$ ) entering already at  $k = 2$ . All models converge at  $k = 8$ , where all predictors are included by construction. Overall, the consistency in selected variables across specifications enhances the credibility of the comparative analysis by ensuring that performance differences are driven primarily by model structure, rather than predictor choice.

### 3.1. Forecasting performance

We assess out-of-sample forecasting performance using a one-step-ahead rolling-window scheme with a fixed length of  $L = 400$ , consistent with AG. This window provides the initial sample for model estimation and yields 487 out-of-sample return forecasts. At each step, the oldest observation is dropped, the next available data point is added, and the model is re-estimated. This procedure continues until the end of the sample, allowing forecasts to adapt to new information. By re-estimating the model each month, we accommodate potential changes in the data-generating process and mitigate parameter instability in a simple but practical manner.

Let  $\hat{r}_{t+1} = \hat{\mathbb{E}}(R_{t+1} | \mathbf{x}_t)$  denote the forecast obtained from a given conditional model, and let  $\bar{r}_{t+1}$  denote the historical average estimate of  $\mathbb{E}(R_{t+1})$ , both computed using the  $L$  observations in the rolling window. It is worth noting that all historical-average forecasts were positive.

To assess the performance of the predictive methods relative to the historical average, we follow AG and consider the out-of-sample  $R^2$  statistic (Campbell and Thompson, 2008) computed as

$$R_{\text{OOS}}^2 = 1 - \frac{\sum_{t=L}^{T-1} \mathcal{L}(r_{t+1} - \hat{r}_{t+1})}{\sum_{t=L}^{T-1} \mathcal{L}(r_{t+1} - \bar{r}_{t+1})}, \quad (17)$$

where the loss function is  $\mathcal{L}(e) = e^2$  when using squared forecast errors and  $\mathcal{L}(e) = |e|$  when using absolute forecast errors. A positive  $R_{\text{OOS}}^2$  indicates that the conditional model outperforms the historical average, while a negative value suggests the opposite. Additionally, we report the results of the well-known Diebold and Mariano (1995) (DM) test of equal predictive ability. The null hypothesis under test is  $H_0 : \mathbb{E}(\Delta\mathcal{L}_{t+1}) = 0$ , where  $\Delta\mathcal{L}_{t+1} = \mathcal{L}(r_{t+1} - \hat{r}_{t+1}) - \mathcal{L}(r_{t+1} - \bar{r}_{t+1})$  is the loss function differential (using either squared or absolute loss functions). This differential is considered statistically significant if the DM  $t$ -ratio statistic exceeds the critical values from the standard normal distribution.

Table 3 reports the  $R_{\text{OOS}}^2$  values (in %) across models and subset sizes  $k$ , under both squared and absolute loss functions. We consider the linear predictive regression, the CSR approach, the two nonlinear benchmarks (MS and GARCH-M), and the decomposition-based models (copula-based and CSM). Figure 1 complements the table by plotting the  $R_{\text{OOS}}^2$  profiles for a selected subset of models (Linear, CSR, GARCH-M, Clayton and FGM copulas, and CSM Baseline), facilitating a visual comparison of how performance changes with  $k$ .

Under squared loss, all decomposition-based models deliver positive  $R_{\text{OOS}}^2$  values for  $k \leq 5$ , with the strongest gains concentrated at intermediate subset sizes (roughly  $k = 3-5$ ); see Table 3. Figure 1(a) illustrates this pattern for the decomposition-based specifications included in the plot (Clayton, FGM, and CSM Baseline). The CSR approach performs reasonably well at low dimensionality (positive for  $k \leq 4$ ) but deteriorates once additional predictors are included. In contrast, the linear predictive regression and the MS model yield

negative  $R_{\text{OOS}}^2$  values across all  $k$ , indicating persistent underperformance relative to the historical average benchmark; this is consistent with evidence that regime-switching models often deliver strong in-sample fit but weaker out-of-sample performance due to uncertainty around regime forecasts (Boot and Pick, 2018). The GARCH-M model delivers gains only at very low dimensionality (notably  $k = 1$ ), and its performance declines steadily as  $k$  increases. For larger subset sizes ( $k \geq 6$ ), squared-loss performance weakens across specifications, including the decomposition-based class, yet these models remain the most competitive overall in this metric.

Under absolute loss, a similar ranking emerges, but performance is more robust for the decomposition-based models. In particular, all decomposition-based models deliver positive  $R_{\text{OOS}}^2$  values across all subset sizes and thus consistently outperform the historical average under absolute loss; see Table 3. Figure 1(b) shows that the plotted decomposition-based models (Clayton, FGM, and CSM Baseline) stay above zero throughout, while CSR and the linear model deliver improvements only for very small subset sizes (CSR for  $k \leq 3$ , with values close to zero at  $k = 3$ , and the linear model for  $k \leq 2$ ) before turning clearly negative. The GARCH-M model also produces positive gains only for small  $k$  (up to  $k = 3$ ) and then deteriorates, and the MS model continues to perform poorly with large negative values (Table 3).

Taken together, these results indicate that the CSM and copula-based decomposition models provide the most robust out-of-sample performance across loss functions and subset sizes. More broadly, the evidence highlights that increasing the number of predictors does not necessarily improve forecast accuracy: for the benchmark methods (linear, CSR, GARCH-M, and MS), performance is maximized at small subset sizes, reinforcing the role of parsimony emphasized by Elliott et al. (2013). The linear model delivers negative  $R_{\text{OOS}}^2$  values under squared loss for all  $k$ , and under absolute loss it improves on the historical average only at very small subset sizes ( $k \leq 2$ ), consistent with evidence that linear predictive regressions

often fail to outperform the historical average in equity premium forecasting (Campbell and Thompson, 2008; Welch and Goyal, 2008).

The results so far indicate that the decomposition-based approaches (both copula-based and CSM) generally outperform the historical average benchmark as well as the linear predictive regression and the MS specification. However, an important question remains: do the decomposition-based models differ significantly in performance, and if so, which specification performs best?

To address this question, we adopt the model confidence set (MCS) methodology of Hansen et al. (2011), which provides a joint comparison across many competing models and avoids the limitations of relying on numerous pairwise tests. For a given subset size  $k$ , exhaustively comparing the 11 candidate models would require  $\binom{11}{2} = 55$  pairwise tests, raising multiple-comparison concerns and inflating the scope for spurious rejections; see Harvey et al. (2020) for a discussion of multiple testing in finance.

The MCS procedure circumvents these limitations by identifying a subset of models that are statistically indistinguishable in terms of predictive performance. For each model, we compute the sequence of out-of-sample losses under a pre-specified loss function and construct test statistics to compare predictive performance. The weakest model is iteratively removed, and elimination continues until the remaining set contains only models that cannot be statistically differentiated in terms of predictive ability.<sup>4</sup>

Tables 4 and 5 report the results of the 80% MCS ( $\text{MCS}_{80\%}$ ) procedure, based on squared and absolute loss functions, respectively. The 80% confidence level applies a more aggressive elimination criterion than higher levels, so  $\text{MCS}_{80\%} \subseteq \text{MCS}_{90\%} \subseteq \text{MCS}_{95\%}$ , and a model belongs to  $\text{MCS}_{80\%}$  if its MCS  $p$ -value is at least 0.20. For each subset size  $k$ , the tables report the average loss (MSE or MAE, scaled by 100) together with the corresponding MCS

---

<sup>4</sup>We base the MCS on the  $T_{\max, \mathcal{M}}$  statistic of Hansen et al. (2011, Section 3.1.2) and implement it using the MCS package in R (Bernardi and Catania, 2018).

$p$ -values. Under squared loss (Table 4), the MS model is excluded from the MCS for every  $k$  (MCS  $p$ -values equal to zero), and the linear model is excluded for several subset sizes, most notably for  $k = 3$ – $6$  and  $k = 8$ . The GARCH-M specification remains competitive only at small subset sizes, but is excluded at  $k = 4$ – $6$  and receives comparatively low MCS  $p$ -values at larger  $k$ . The CSR approach is retained for most values of  $k$ , but is excluded at  $k = 8$ , consistent with its deterioration at high dimensionality. Overall, under squared loss, differences in average MSE are relatively small, which limits the extent to which the MCS can separate closely performing models.

Under absolute loss (Table 5), the MCS delivers sharper differentiation among the benchmark specifications. The linear and MS models are excluded for several intermediate and large subset sizes (notably for  $k = 3$ – $5$  and  $k = 8$ ). The CSR approach is retained for  $k = 1$ – $7$  but is excluded at  $k = 8$ , and its MCS  $p$ -values are sometimes close to the 0.20 cutoff (e.g.,  $k = 1$  and  $k = 5$ ). The GARCH-M benchmark is competitive at low dimensions (included with MCS  $p$ -values equal to one at  $k = 1$  and  $k = 2$ , and still retained at  $k = 3$ ), suggesting that incorporating volatility feedback may yield modest gains when the predictor dimension is limited; however, it is excluded at  $k = 4$ ,  $k = 5$ , and  $k = 8$ , indicating sensitivity to  $k$ .

In contrast, the decomposition-based models (the copula-based specifications and the CSM specifications) are retained in the MCS across all values of  $k$  under both loss functions, with MCS  $p$ -values typically close to one. This shows that these models are consistently among the statistically best-performing specifications out of sample. Empirically, the MCS does not meaningfully distinguish between the CSM models and the copula-based decomposition models, consistent with the view in AG that the dependence between the sign and magnitude of monthly returns is weak, so that modeling this dependence flexibly yields limited incremental gains in overall loss.

Finally, while the MCS results suggest broadly similar performance within the decomposition class under standard loss functions, the CSM framework is designed to target directional

predictability more directly by modeling the return sign conditional on predictors and contemporaneous magnitude. The next section therefore turns to the economic value of these directional forecasts.

### 3.2. Economic value

What matters to investors is not necessarily predictive accuracy as measured by statistical criteria (Jordà and Taylor, 2011), but rather the ability to correctly anticipate whether returns will be positive or negative, especially when significant profits or losses are at stake. Merton (1980) likewise argues that fund managers place greater emphasis on the direction of returns than on their magnitude. In this section, we therefore assess the economic significance of the models through their ability to predict return direction. Specifically, we examine their value for market timing by comparing the returns from a passive buy-and-hold strategy, under which the risky asset is held throughout the sample period, with those from market timing strategies constructed from forecasts generated by the various model specifications.

Following Breen et al. (1989), Pesaran and Timmermann (1995), Guo (2006), and AG, among others, we use a simple trading rule known as a switching strategy, which only requires the direction of the forecasts. This rule involves investing the current wealth in the stock market by buying shares of the index if the predicted excess return is positive, or investing the current wealth in the risk-free asset if the predicted excess return is negative. Once the investor makes a decision following this trading strategy, the funds remain invested until the end of the month, before a new decision is made at the beginning of the next month. The trading begins with an initial wealth of \$1, and the portfolio value is recalculated and reinvested every month.

To make the economic evaluation more realistic, we introduce proportional transaction costs of  $c = 10$  basis points of the portfolio's value when the investor makes a switch.<sup>5</sup> Let

---

<sup>5</sup>French (2008) documents a sharp decline in trading costs over this period, providing the basis for our

$r_{p,t+1}$  denote the out-of-sample portfolio return realized at time  $t + 1$ , and let  $R_{p,t+1}$  denote the corresponding random portfolio return. The investor’s wealth then evolves from month  $t$  to month  $t + 1$  according to  $W_{t+1} = W_t(1 + r_{p,t+1})(1 - c)$ , when a switch is made, and  $W_{t+1} = W_t(1 + r_{p,t+1})$ , otherwise. The portfolios formed using the switching rule based on forecasts from the different candidate models are compared to the benchmark buy-and-hold strategy. We note that the historical average model resulted in a strategy that is equivalent to the buy-and-hold strategy, as all the historical-average forecasts turned out positive.

The active strategies compared with the buy-and-hold benchmark include those based on the previously discussed forecasting models. In addition, we consider momentum-based switching strategies, where the forecasted sign of next month’s return is taken to be the sign of the cumulative return over a recent historical period. These backward-looking rules offer a useful contrast to the forward-looking model-based strategies. We report results for momentum strategies based on cumulative returns over the past 3, 6, and 12 months.

For each portfolio strategy, the following out-of-sample performance metrics are computed: (i) TW, the terminal wealth in dollars at the end of the trading period; (ii) AV, the average return (annualized by multiplying by 12) in percent; (iii) SD, the standard deviation of returns (annualized by multiplying by the square root of 12) in percent; (iv) SR, the Sharpe ratio; and (v) MDD, the maximum drawdown in percent over the trading period. To assess statistical significance, we test whether the Sharpe ratio of each strategy differs from that of the buy-and-hold benchmark using the two-sided  $p$ -value from the prewhitened  $HAC_{PW}$  method of Ledoit and Wolf (2008), which tests the null hypothesis of equal Sharpe ratios.

We evaluate out-of-sample portfolio performance across predictor subset sizes  $k = 1$  to 8. Table 6 focuses on four informative cases ( $k = 1, 2, 3, 7$ ) that capture early-stage gains and behavior at higher dimensions, with the remaining results ( $k = 4, 5, 6, 8$ ) reported in 

---

choice of 10 basis points as a conservative estimate.

the Supplementary material (Table B1). Over the out-of-sample trading period (May 1981–December 2021), the buy-and-hold benchmark delivers solid performance (annualized return 12.65%, standard deviation 15.00%, Sharpe ratio 0.17, terminal wealth \$104.63), despite encompassing major crisis episodes.<sup>6</sup>

Momentum strategies show mixed performance: while 12-month momentum proves competitive (TW \$100.21, SR 0.20), short-horizon variants (3-month and 6-month) substantially underperform, with terminal wealth below \$50. Figure 2 plots wealth growth for CSM (Baseline) at  $k = 3$  alongside buy-and-hold and momentum-based switching strategies, showing that CSM accumulates wealth more rapidly over time and ends above both buy-and-hold and all momentum horizons, with the gap widening in the later part of the sample.

The linear predictive regression exhibits mixed performance at very small subset sizes but deteriorates sharply as predictor dimension increases (terminal wealth \$71.58 at  $k = 4$  and \$62.31 at  $k = 5$ ; Supplementary material, Table B1), highlighting the instability of traditional linear models with many predictors. The CSR approach partially stabilizes performance relative to the individual linear model at moderate  $k$  (reaching \$126.80 at  $k = 5$ ; Supplementary material, Table B1), but it does not dominate uniformly and coincides with the linear model at  $k = 8$  by construction. The nonlinear benchmarks are also sensitive to predictor dimension: GARCH-M performs best at low to moderate  $k$  (TW \$151.01 at  $k = 2$  and \$147.63 at  $k = 3$ ), whereas the MS strategy collapses at  $k = 3$  (TW \$25.65).

Overall, the decomposition-based strategies deliver the strongest economic performance, particularly at intermediate and larger subset sizes. At  $k = 3$ , the copula-based strategies (Gaussian/Frank/FGM) reach TW \$165.23, well above CSR (\$98.22) and the linear model (\$112.79), while the proposed CSM strategy (Baseline and Poly coincide at this dimension) attains the highest terminal wealth (TW \$181.68) with a statistically significant Sharpe-

---

<sup>6</sup>We conduct a dedicated sub-sample analysis of the Global Financial Crisis and the COVID-19 crisis to assess model performance during turbulent periods. Results are reported in Section C of the Supplementary material.

ratio improvement over buy-and-hold (SR 0.21, 5% level). Figure 3 shows that CSM and the Frank copula track each other closely for much of the sample, with CSM pulling ahead in the later years; by contrast, CSR and the linear strategy lag behind after major drawdowns, resulting in substantially lower terminal wealth. This temporal perspective indicates that the superior performance of decomposition-based strategies reflects sustained accumulation rather than isolated episodes of outperformance.

At higher dimensions, decomposition-based strategies continue to perform strongly. At  $k = 7$ , copula-based strategies deliver terminal wealth of \$176–\$178 with Sharpe ratios of 0.22 (statistically significant at the 10% level), slightly outperforming CSM (Baseline) at this dimension (TW \$168.05, SR 0.22). Within the cases reported in Table 6, statistically significant Sharpe-ratio improvements at the 10% level are also achieved by multiple strategies at  $k = 2$  (GARCH-M, Frank copula, CSM Baseline) and at  $k = 3$  (Gaussian, Frank, and FGM copulas). Complementary evidence for the remaining subset sizes, including significance at  $k = 4$  for CSM (Poly), is reported in the Supplementary material (Table B1).

Figure 4 illustrates how terminal wealth varies with the number of predictors across selected strategies, revealing starkly different dimensional-stability properties. CSM (Baseline) is clearly non-monotone in  $k$ , peaking at  $k = 3$  and remaining highly competitive at  $k = 7–8$  while being weaker at  $k = 4–6$ . In contrast, the linear model exhibits severe instability (e.g., dropping from \$112.79 at  $k = 3$  to \$62.31 at  $k = 5$ ), while CSR offers only partial stabilization relative to the linear benchmark. The Clayton copula also displays pronounced non-monotonicity, with strong performance at  $k = 7–8$ .

We further gauge the economic value of the models by calculating their certainty equivalent returns (CER), as in Campbell and Thompson (2008), Cenesizoglu and Timmermann (2012), Ciner (2022), Christoffersen and Langlois (2013), and Borup et al. (2024), among many others. The CER is the guaranteed return that the investor would accept instead of a risky investment, such that the utility derived from the CER equals the expected utility

of the risky position. Typically, the average utility is used as a consistent estimate of the expected utility.

We compute the CER associated with each strategy by assuming two different types of investor preferences. First, we consider mean-variance preferences summarized by the mean-variance criterion

$$\mathbb{E}(R_{p,t+1}) - \frac{1}{2}\gamma\text{Var}(R_{p,t+1}),$$

where  $\gamma$  is the risk aversion parameter. The associated CER is computed as

$$\text{CER} = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}_p^2,$$

where  $\hat{\mu}_p$  and  $\hat{\sigma}_p^2$  are the sample mean and variance, respectively, of the realized portfolio returns  $\{r_{p,t+1}\}_{t=L+1}^T$  over the out-of-sample evaluation period.

Second, we consider constant relative risk aversion (CRRA) preferences,

$$U(R_{p,t+1}) = \frac{(1 + R_{p,t+1})^{1-\gamma}}{1 - \gamma}.$$

The advantage of CRRA utility is that it incorporates preferences toward higher-order moments (Brandt et al., 2009). In this case, the average utility over the out-of-sample period is

$$\bar{U} = \frac{1}{T-L} \sum_{t=L}^{T-1} \frac{(1 + r_{p,t+1})^{1-\gamma}}{1 - \gamma}.$$

By solving for CER in  $(1+\text{CER})^{1-\gamma} = (T-L)^{-1} \sum_{t=L}^{T-1} (1+r_{p,t+1})^{1-\gamma}$ , we obtain the expression

$$\text{CER} = \left[ \frac{1}{T-L} \sum_{t=L}^{T-1} (1 + r_{p,t+1})^{1-\gamma} \right]^{\frac{1}{1-\gamma}} - 1.$$

The CER gain is defined as the difference between the CER for the predictive methods and the CER for the historical average benchmark. We annualize the CER gain so that it can

be interpreted as the annual portfolio management fee that an investor would be willing to pay to access the information provided by the predictive methods compared to simply using the historical average.

The results are summarized in Table 7, which reports the CER gains relative to the historical average benchmark for an investor with moderate risk aversion ( $\gamma = 5$ ; results for  $\gamma \in \{3, 10\}$  were qualitatively similar to those discussed here). For brevity, we present results for  $k = 1, 2, 3, 7$ ; the remaining cases are provided in Section B of the Supplementary material. Panel A reports results for a mean-variance investor, while Panel B corresponds to CRRA preferences. In both cases, the CSM (Baseline) model delivers strong CER gains at  $k = 2-3$  and remains competitive at larger subset sizes (e.g.,  $k = 7$ , and also  $k = 8$  in Supplementary Table B2), though performance is non-monotone in  $k$  with weaker gains at intermediate dimensions (e.g.,  $k = 4$  and  $k = 6$ ; Supplementary Table B2). For example, at  $k = 2$ , it yields CER gains of 1.122 (Panel A) and 1.108 (Panel B), outperforming the CSR approach (0.596 and 0.303, respectively). A similar pattern holds at  $k = 3$ , where the CSM (Baseline) achieves 1.878 and 1.939, which are higher than most competing models. At  $k = 7$ , the CSM (Baseline) continues to perform strongly with gains of 2.302 and 2.612 under mean-variance and CRRA preferences, respectively.

The copula-based models deliver competitive results and frequently outperform the linear and CSR benchmarks, though their relative performance varies with  $k$  and the specific copula specification. For instance, under mean-variance preferences and  $k = 2$ , the Frank copula achieves the highest CER gain among the copulas (1.122), while at  $k = 7$ , the Gaussian copula attains 2.442, slightly exceeding the CSM (Baseline) model (2.302). Under CRRA preferences, copula-based strategies remain competitive at larger  $k$  values, with the Gaussian copula reaching 2.748 at  $k = 7$  (versus 2.612 for CSM Baseline). Momentum strategies based on 12-month signals perform consistently well, while 3-month momentum substantially underperforms (CER gains approximately  $-0.71\%$  and  $-0.56\%$  under mean-variance

and CRRA preferences, respectively) and 6-month momentum delivers only modest positive gains. The linear model exhibits non-monotone performance in  $k$ : it delivers positive CER gains at small subset sizes and again at larger sizes (e.g.,  $k = 7-8$ ), but turns negative at intermediate dimensions (e.g.,  $-0.408$  at  $k = 5$  under mean-variance preferences; Supplementary Table B2).

The nonlinear benchmark models exhibit more mixed performance. GARCH-M performs well at smaller subset sizes, for example yielding CER gains of 1.075 (Panel A) and 1.062 (Panel B) at  $k = 2$ , and remains positive at  $k = 7$  with 0.412 and 0.675, respectively. The MS model performs strongly only at  $k = 1$ , with CER gains above 1.6 in both panels, but its performance declines rapidly thereafter, becoming negative by  $k = 3$  and remaining negative at  $k = 7$ . Overall, while some copula specifications deliver the highest CER gains at particular  $k$ , the CSM framework delivers substantial economic value and remains consistently competitive across subset sizes.

Nevertheless, a natural question arises as to whether an investor stands to benefit from transitioning away from the CSM (Baseline) model in favor of forecasts generated by alternative copula-based decomposition strategies. Table 8 addresses this question directly by reporting CER gains computed relative to the CSM (Baseline) benchmark, using the same setup and investor preferences as in Table 7. For brevity, we report results for  $k = 1, 2, 3, 7$ ; the remaining cases are provided in Section B of the Supplementary material.

Across both preferences, copula-based strategies generally offer no improvement in utility over the CSM (Baseline) model when the number of predictors is relatively low ( $k \leq 3$ ). For example, under mean-variance preferences, adopting a Gaussian copula instead of the CSM (Baseline) implies annual CER losses of 8.7, 50.4, and 28.8 basis points for  $k = 1, 2, 3$ , respectively. Under CRRA preferences, the corresponding losses are 6.4, 45.0, and 29.6 basis points. These negative CER differentials indicate that, for modest predictor specifications, the CSM (Baseline) model remains difficult to beat (at best, the Frank copula essentially

ties the baseline at  $k = 2$ ).

At  $k = 7$ , some copula specifications modestly improve upon the baseline: under mean-variance preferences, the Gaussian and Clayton copulas deliver gains of 14.0 and 11.0 basis points relative to the baseline, respectively, while under CRRA preferences they deliver 13.6 and 10.7 basis points. At  $k = 8$  (Supplementary material, Table B2), these differentials remain small overall: Gaussian/Frank/FGM are essentially neutral (or slightly negative), while the Clayton copula continues to provide a modest positive gain of about 9 basis points. At intermediate dimensions, however, the scope for improvements over CSM (Baseline) can be economically meaningful: for example, at  $k = 4$  and  $k = 6$ , several copula specifications and, in particular, the CSM (Poly) variant deliver sizable positive CER differentials relative to the baseline (Supplementary material, Table B2). Overall, these results underscore that CSM (Baseline) is difficult to beat for low-dimensional specifications ( $k \leq 3$ ), while at larger dimensions ( $k = 7-8$ ) copula-based alternatives offer at most modest incremental utility, and the Poly variant provides an effective way to strengthen robustness at intermediate subset sizes.

Finally, we stress that the levels of certainty equivalents reported here are not meant to be directly compared to large-scale applications with more extensive predictor sets, such as those in Bianchi et al. (2025) and Goyal et al. (2024). Our focus is on relative performance across methods within our moderate predictor set and under the same portfolio construction and evaluation framework.

## 4. Conclusion

In this paper, we introduced the CSM decomposition approach, a new method for return prediction that builds on the AG framework by decomposing returns into sign and magnitude components. The expected return is modeled using the joint distribution of these compo-

nents, combining a multiplicative error model for the magnitude with a binary choice model for the sign conditional on the contemporaneous magnitude. By avoiding the need to explicitly specify or estimate a copula function, the CSM framework simplifies implementation while preserving the ability to capture nonlinear return dynamics. This structure exploits the well-documented relationship between return magnitude (i.e., volatility) and direction, as emphasized by Christoffersen and Diebold (2006).

In our empirical application, return decomposition-based strategies often outperform traditional benchmarks such as the historical average and the linear predictive regression, as well as the CSR forecast-combination method. Relative to economically motivated nonlinear benchmarks that incorporate time-varying risk and regime dependence (GARCH-in-mean and Markov-switching specifications), the CSM approach remains highly competitive. Compared with copula-based decomposition methods, CSM provides a direct way to exploit sign-magnitude dependence, and the CSM (Baseline) specification delivers consistently strong CER gains across investor preferences and predictor dimensions. Copula-based methods are also competitive, with performance that varies across copula choices and subset sizes.

Economically, these findings translate into high terminal wealth outcomes and robust utility gains in the market-timing exercise, with CSM typically among the best-performing specifications even as performance differences narrow at larger predictor dimensions. Overall, the results highlight the value of modeling direction and magnitude jointly and position CSM as a practical and effective tool for return forecasting and portfolio allocation.

## References

- Algieri, B. and A. Leccadito (2019). Ask CARL: Forecasting tail probabilities for energy commodities. *Energy Economics* 84, 104497.
- Anatolyev, S. and N. Gospodinov (2010). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics* 28(2), 232–245.
- Anatolyev, S. and N. Gospodinov (2019). Multivariate return decomposition: Theory and implications. *Econometric Reviews* 38(5), 487–508.
- Anatolyev, S., N. Gospodinov, I. Jamali, and X. Liu (2017). Foreign exchange predictability and the carry trade: A decomposition approach. *Journal of Empirical Finance* 42, 199–211.
- Ang, A. and G. Bekaert (2002). International asset allocation with regime shifts. *Review of Financial Studies* 15(4), 1137–1187.
- Ang, A. and G. Bekaert (2004). How regimes affect asset allocation. *Financial Analysts Journal* 60(2), 86–99.
- Bekiros, S. D. and D. A. Georgoutsos (2008). Non-linear dynamics in financial asset returns: the predictive power of the CBOE volatility index. *European Journal of Finance* 14(5), 397–408.
- Bernardi, M. and L. Catania (2018). The model confidence set package for R. *International Journal of Computational Economics and Econometrics* 8(2), 144–158.
- Bianchi, D., A. Rubesam, and A. Tamoni (2025). It takes two to tango: Economic theory and model uncertainty for equity premium prediction. *SSRN Working paper*.

- Boot, A. and A. Pick (2018). Optimal forecasting for Markov-switching models. *Journal of Business & Economic Statistics* 36(4), 628–642.
- Borup, D., J. N. Eriksen, M. M. Kjær, and M. Thyrsgaard (2024). Predicting bond return predictability. *Management Science* 70(2), 931–951.
- Bossaerts, P. and P. Hillion (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies* 12(2), 405–428.
- Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies* 22(9), 3411–3447.
- Breen, W., L. R. Glosten, and R. Jagannathan (1989). Economic significance of predictable variations in stock index returns. *Journal of Finance* 44(5), 1177–1189.
- Brown, D. P. and R. H. Jennings (1989). On technical analysis. *Review of Financial Studies* 2(4), 527–551.
- Butler, A. W., G. Grullon, and J. P. Weston (2005). Can managers forecast aggregate market returns? *Journal of Finance* 60(2), 963–986.
- Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21(4), 1509–1531.
- Campisi, G., S. Muzzioli, and B. De Baets (2024). A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices. *International Journal of Forecasting* 40(3), 869–880.
- Cederburg, S., T. L. Johnson, and M. S. O’Doherty (2023). On the economic significance of stock return predictability. *Review of Finance* 27(2), 619–657.

- Cenesizoglu, T. and A. Timmermann (2012). Do return prediction models add economic value? *Journal of Banking & Finance* 36(11), 2974–2987.
- Cespa, G. and X. Vives (2012). Dynamic trading and asset prices: Keynes vs. Hayek. *Review of Economic Studies* 79(2), 539–580.
- Chevapatrakul, T. (2013). Return sign forecasts based on conditional risk: Evidence from the UK stock market index. *Journal of Banking & Finance* 37(7), 2342–2353.
- Christoffersen, P. and H. Langlois (2013). The joint dynamics of equity market factors. *Journal of Financial and Quantitative Analysis* 48(5), 1371–1404.
- Christoffersen, P. F. and F. X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52(8), 1273–1287.
- Christoffersen, P. F., F. X. Diebold, R. S. Mariano, A. S. Tay, and Y. K. Tse (2007). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: International evidence. *Journal of Financial Forecasting* 1(2), 1–22.
- Ciner, C. (2022). Predicting the equity market risk premium: A model selection approach. *Economics Letters* 215, 110448.
- Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Edmans, A., I. Goldstein, and W. Jiang (2015). Feedback effects, asymmetric trading, and the limits to arbitrage. *American Economic Review* 105(12), 3766–3797.
- Elliott, G., A. Gargano, and A. Timmermann (2013). Complete subset regressions. *Journal of Econometrics* 177(2), 357–373.

- Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17(5), 425–446.
- Engle, R. F., D. M. Lilien, and R. P. Robins (1987). Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica* 55(2), 391–407.
- Frazier, D. T. and X. Liu (2016). A new approach to risk-return trade-off dynamics via decomposition. *Journal of Economic Dynamics and Control* 62, 43–55.
- French, K. R. (2008). Presidential address: The cost of active investing. *Journal of Finance* 63(4), 1537–1573.
- Goyal, A., I. Welch, and A. Zafirov (2024). A comprehensive 2022 look at the empirical performance of equity premium prediction. *Review of Financial Studies* 37(11), 3490–3557.
- Granger, C. W. and C.-Y. Sin (2000). Modelling the absolute returns of different stock indices: exploring the forecastability of an alternative measure of risk. *Journal of Forecasting* 19(4), 277–298.
- Guo, H. (2006). On the out-of-sample predictability of stock market returns. *Journal of Business* 79(2), 645–670.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57(2), 357–384.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* 79(2), 453–497.
- Harvey, C. R., Y. Liu, and A. Saretto (2020). An evaluation of alternative multiple testing methods for finance applications. *Review of Asset Pricing Studies* 10(2), 199–248.

- Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics* 99(3), 560–580.
- Hong, H. and J. C. Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance* 54(6), 2143–2184.
- Joe, H. and J. J. Xu (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.
- Jordà, Ò. and A. M. Taylor (2011). Performance evaluation of zero net-investment strategies. Technical report, National Bureau of Economic Research.
- Jordà, Ò. and A. M. Taylor (2012). The carry trade and fundamentals: Nothing to fear but FEER itself. *Journal of International Economics* 88(1), 74–90.
- Judd, K. L. (1998). *Numerical Methods in Economics*. Cambridge, Massachusetts: MIT Press.
- Kauppi, H. and P. Saikkonen (2008). Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics* 90(4), 777–791.
- Ledoit, O. and M. Wolf (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance* 15(5), 850–859.
- Liu, X. (2015). Modeling time-varying skewness via decomposition for out-of-sample forecast. *International Journal of Forecasting* 31(2), 296–311.
- Liu, X. (2017). Unfolded risk-return trade-offs and links to macroeconomic dynamics. *Journal of Banking & Finance* 82, 1–19.

- Liu, X. and R. Luger (2015). Unfolded GARCH models. *Journal of Economic Dynamics and Control* 58, 186–217.
- Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs. Cambridge University Press.
- Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8(4), 323–361.
- Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012). Time series momentum. *Journal of Financial Economics* 104(2), 228–250.
- Nyberg, H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting* 27(2), 561–578.
- Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *Journal of Finance* 50(4), 1201–1228.
- Pönkä, H. (2017). Predicting the direction of US stock markets using industry returns. *Empirical Economics* 52(4), 1451–1480.
- Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* 23(2), 821–862.
- Rapach, D. E. and G. Zhou (2013). Forecasting stock returns. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Chapter 6, pp. 328–383.
- Robert, C. and G. Casella (2010). *Introducing Monte Carlo Methods with R*. Use R! Springer.
- Sanchez-Espigares, J. A. and A. Lopez-Moreno (2021). *MSwM: Fitting Markov Switching Models in R*. The R Project for Statistical Computing. R package version 1.5.

- Shih, J. H. and T. A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* 51(4), 1384–1399.
- Sklar, M. (1959). Fonctions de répartition à  $N$  dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8, 229–231.
- Treynor, J. L. and R. Ferguson (1985). In defense of technical analysis. *Journal of Finance* 40(3), 757–773.
- Trivedi, P. K. and D. M. Zimmer (2007). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics* 1(1), 1–111.
- Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21(4), 1455–1508.

Table 1: Sample correlations

Panel A: Between $r_t$ and $\mathbf{x}_{t-1}$										
	$r_t$	$dp_{t-1}$	$dfy_{t-1}$	$tms_{t-1}$	$tbl_{t-1}$	$ltr_{t-1}$	$dfr_{t-1}$	$ntis_{t-1}$	$infl_{t-1}$	
$r_t$	1.000	0.059	0.022	0.058	-0.097	0.045	0.038	-0.018	-0.069	
Panel B: Between $s_t$ , $m_t$ , and $\mathbf{x}_{t-1}$										
	$s_t$	$m_t$	$dp_{t-1}$	$dfy_{t-1}$	$tms_{t-1}$	$tbl_{t-1}$	$ltr_{t-1}$	$dfr_{t-1}$	$ntis_{t-1}$	$infl_{t-1}$
$s_t$	1.000	-0.011	0.006	-0.025	0.059	-0.143	0.026	0.049	-0.046	-0.118
$m_t$		1.000	0.029	0.166	-0.015	0.044	0.093	-0.146	-0.062	0.024
Panel C: Among $\mathbf{x}_{t-1}$										
	$dp$	$dfy$	$tms$	$tbl$	$ltr$	$dfr$	$ntis$	$infl$		
$dp$	1.000	0.125	0.322	0.344	-0.006	0.389	0.169	0.168		
$dfy$		1.000	0.260	0.139	0.067	-0.361	0.069	0.063		
$tms$			1.000	-0.374	0.076	-0.261	-0.197	-0.198		
$tbl$				1.000	-0.045	-0.085	0.407	0.499		
$ltr$					1.000	-0.457	0.078	0.102		
$dfr$						1.000	0.003	-0.004		
$ntis$							1.000	0.102		
$infl$								1.000		

*Notes:* This table reports sample correlations between the variables used in the analysis over the full sample from January 1948 to December 2021. Panel A shows correlations between the current excess return ( $r_t$ ) and the lagged predictors:  $dp_{t-1}$  (dividend-price ratio),  $dfy_{t-1}$  (default yield spread),  $tms_{t-1}$  (term spread),  $tbl_{t-1}$  (short-term interest rate),  $ltr_{t-1}$  (long-term return),  $dfr_{t-1}$  (default return spread),  $ntis_{t-1}$  (net equity expansion), and  $infl_{t-1}$  (inflation). Panel B presents correlations between the sign ( $s_t$ ) and magnitude ( $m_t$ ) components of the excess return and the lagged predictors. Panel C reports the correlations among the lagged predictors themselves.

Table 2: Best subset selection based on AUC criterion

	const	dp	dfy	tms	tbl	ltr	dfr	ntis	infl
Panel A: Linear and GARCH-M models									
$k = 1$	*				*				
$k = 2$	*				*			*	
$k = 3$	*				*		*	*	
$k = 4$	*		*	*	*			*	
$k = 5$	*		*	*	*		*	*	
$k = 6$	*		*	*	*		*	*	*
$k = 7$	*	*	*	*	*		*	*	*
$k = 8$	*	*	*	*	*	*	*	*	*
Panel B: MS model									
$k = 1$	*				*				
$k = 2$	*				*				*
$k = 3$	*		*	*	*				
$k = 4$	*	*	*	*			*		
$k = 5$	*	*	*	*			*	*	
$k = 6$	*		*	*	*		*	*	*
$k = 7$	*		*	*	*	*	*	*	*
$k = 8$	*	*	*	*	*	*	*	*	*
Panel C: Decomposition-based models (CSM and copula-based)									
$k = 1$	*				*				
$k = 2$	*				*		*		
$k = 3$	*				*		*	*	
$k = 4$	*		*	*	*			*	
$k = 5$	*	*	*	*	*			*	
$k = 6$	*	*	*	*	*			*	*
$k = 7$	*	*	*	*	*		*	*	*
$k = 8$	*	*	*	*	*	*	*	*	*

*Notes:* A star (\*) indicates that the corresponding variable is included in the best subset of size  $k$ . The subsets are selected based on the AUC criterion using one-step-ahead in-sample forecasts. An empty cell indicates exclusion of the variable from the selected subset. Panel A shows the selected subsets for the linear and GARCH-M models, which select identical variables for all values of  $k$ . Panel B shows the selected subsets for the MS model. Panel C shows the selected subsets for the decomposition-based models (CSM and copula-based); results are identical across all copula specifications. At  $k = 8$ , all predictors are included by construction in all models, so no subset selection is required.

Table 3: Out-of-sample  $R^2$  values by subset size ( $k$ ) and loss function

	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	Squared	Absolute	Squared	Absolute	Squared	Absolute	Squared	Absolute
Linear model	-0.24	0.91	-0.56	0.42	-1.52	-0.26	-2.28	-1.00
CSR approach	0.37	0.24	0.51	0.18	0.40	0.02	0.11	-0.21
GARCH-M model	0.40	1.35**	-0.09	0.72	-0.99	0.14	-2.09	-0.84
MS model	-5.15*	-0.93	-4.49	-0.24	-7.60***	-2.68	-9.61***	-4.26***
Copula-based approach								
Gaussian	0.20	1.39**	0.47	1.27*	0.97	1.34	0.75	0.95
Frank	0.18	1.47**	0.47	1.55*	0.94	1.37	0.71	0.99
Clayton	0.28	1.54**	0.75	1.39*	1.03	1.18	0.90	1.20
FGM	0.17	1.46**	0.59	1.36*	0.94	1.37	0.70	0.99
CSM approach								
Baseline	0.19	1.25**	0.77	1.41*	0.83	1.18	0.57	0.74
Poly	0.04	1.32**	0.23	1.16	0.67	1.19	0.28	0.73
	$k = 5$		$k = 6$		$k = 7$		$k = 8$	
	Squared	Absolute	Squared	Absolute	Squared	Absolute	Squared	Absolute
Linear model	-3.52	-1.57	-3.76	-1.03	-5.80*	-2.78	-5.15	-2.87
CSR approach	-0.40	-0.56	-1.23	-1.07	-2.68	-1.83	-5.15	-2.87
GARCH-M model	-2.92	-1.22	-3.29	-0.96	-4.80	-2.33	-3.64	-2.69
MS model	-10.69***	-5.08***	-7.30*	-2.23	-7.77*	-2.87	-16.93***	-7.21***
Copula-based approach								
Gaussian	0.93	0.92	-0.34	0.37	-0.11	0.52	-0.39	0.35
Frank	1.35	1.51	-0.27	0.46	-0.04	0.56	-0.40	0.34
Clayton	1.04	1.05	-0.09	0.57	0.08	0.65	-0.28	0.42
FGM	0.90	0.95	-0.20	0.47	-0.04	0.56	-0.49	0.32
CSM approach								
Baseline	0.88	0.78	-0.38	0.26	-0.42	0.34	-0.68	0.14
Poly	1.04	0.94	0.72	0.80	-0.96	0.47	-0.99	0.35

Notes: This table reports  $R^2_{\text{OOS}}$  statistics (in %) for squared and absolute loss functions, relative to the historical-average benchmark. Positive values indicate improvement over the historical average, while negative values indicate deterioration. Stars indicate rejection of equal predictive accuracy (DM test) against the historical average benchmark: \*, \*\*, and \*\*\* denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Out-of-sample model performance comparison using MCS procedure with squared loss

	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	MSE	MCS $p$ -value	MSE	MCS $p$ -value	MSE	MCS $p$ -value	MSE	MCS $p$ -value
Historical average	0.189	0.999	0.189	0.970	0.189	0.953	0.189	0.946
Linear model	0.189	0.292	0.189	0.593	0.191	0.000	0.193	0.000
CSR approach	0.188	1.000	0.188	1.000	0.188	1.000	0.188	0.959
GARCH-M model	0.188	1.000	0.188	0.938	0.190	0.222	0.193	0.000
MS model	0.198	0.000	0.197	0.000	0.203	0.000	0.206	0.000
Copula-based approach								
Gaussian	0.188	1.000	0.187	1.000	0.187	1.000	0.187	1.000
Frank	0.188	1.000	0.187	1.000	0.187	1.000	0.187	1.000
Clayton	0.188	1.000	0.187	1.000	0.186	1.000	0.187	1.000
FGM	0.188	1.000	0.188	1.000	0.187	1.000	0.187	1.000
CSM approach								
Baseline	0.188	1.000	0.187	1.000	0.187	1.000	0.188	1.000
Poly	0.189	0.952	0.188	1.000	0.187	1.000	0.188	0.891
	$k = 5$		$k = 6$		$k = 7$		$k = 8$	
	MSE	MCS $p$ -value	MSE	MCS $p$ -value	MSE	MCS $p$ -value	MSE	MCS $p$ -value
Historical average	0.189	0.895	0.189	1.000	0.189	1.000	0.189	1.000
Linear model	0.196	0.000	0.196	0.000	0.195	0.665	0.198	0.000
CSR approach	0.189	0.743	0.191	0.825	0.194	0.506	0.198	0.000
GARCH-M model	0.194	0.000	0.195	0.000	0.198	0.320	0.195	0.243
MS model	0.208	0.000	0.201	0.000	0.204	0.000	0.220	0.000
Copula-based approach								
Gaussian	0.187	1.000	0.189	0.995	0.189	1.000	0.189	1.000
Frank	0.187	1.000	0.189	1.000	0.189	1.000	0.189	0.829
Clayton	0.187	1.000	0.189	1.000	0.188	1.000	0.189	1.000
FGM	0.187	1.000	0.189	1.000	0.189	1.000	0.190	0.829
CSM approach								
Baseline	0.187	1.000	0.189	0.982	0.189	1.000	0.189	1.000
Poly	0.187	1.000	0.187	1.000	0.190	1.000	0.190	0.829

*Notes:* This table reports MSE values multiplied by 100, along with the associated MCS  $p$ -values. The MCS  $p$ -value indicates the lowest confidence level at which a given model would still be included in the MCS. In other words, it reflects whether the model is among those not significantly worse than the best model in terms of forecasting accuracy. A  $p$ -value of zero indicates that the model has been excluded from the MCS.

Table 5: Out-of-sample model performance comparison using MCS procedure with absolute loss

	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	MAE	MCS $p$ -value	MAE	MCS $p$ -value	MAE	MCS $p$ -value	MAE	MCS $p$ -value
Historical average	3.269	0.000	3.269	0.559	3.268	0.322	3.268	0.731
Linear model	3.239	0.805	3.255	0.878	3.277	0.000	3.301	0.000
CSR approach	3.261	0.225	3.262	0.543	3.268	0.388	3.276	0.296
GARCH-M model	3.224	1.000	3.247	1.000	3.264	0.472	3.296	0.000
MS model	3.299	0.000	3.276	0.500	3.356	0.000	3.402	0.000
Copula-based approach								
Gaussian	3.223	1.000	3.227	1.000	3.224	1.000	3.237	1.000
Frank	3.221	1.000	3.218	1.000	3.224	1.000	3.234	1.000
Clayton	3.218	1.000	3.224	1.000	3.230	1.000	3.229	1.000
FGM	3.220	1.000	3.222	1.000	3.224	1.000	3.236	1.000
CSM approach								
Baseline	3.228	1.000	3.222	1.000	3.230	1.000	3.244	1.000
Poly	3.237	0.982	3.231	1.000	3.230	1.000	3.245	1.000
	$k = 5$		$k = 6$		$k = 7$		$k = 8$	
	MAE	MCS $p$ -value	MAE	MCS $p$ -value	MAE	MCS $p$ -value	MAE	MCS $p$ -value
Historical average	3.267	0.800	3.269	1.000	3.269	0.984	3.269	0.952
Linear model	3.320	0.000	3.302	0.649	3.301	0.924	3.362	0.000
CSR approach	3.287	0.247	3.303	0.753	3.328	0.421	3.362	0.000
GARCH-M model	3.308	0.000	3.300	0.552	3.345	0.445	3.356	0.000
MS model	3.434	0.000	3.328	0.503	3.362	0.326	3.504	0.000
Copula-based approach								
Gaussian	3.239	1.000	3.256	1.000	3.252	1.000	3.253	1.000
Frank	3.219	1.000	3.253	1.000	3.250	1.000	3.250	1.000
Clayton	3.234	1.000	3.250	1.000	3.248	1.000	3.250	1.000
FGM	3.237	1.000	3.253	1.000	3.250	1.000	3.259	1.000
CSM approach								
Baseline	3.234	1.000	3.260	1.000	3.258	1.000	3.248	0.659
Poly	3.243	1.000	3.242	1.000	3.253	1.000	3.257	1.000

Notes: This table reports mean absolute error (MAE) values multiplied by 100, along with the associated MCS  $p$ -values. See Table 4 for more details.

Table 6: Out-of-sample performance statistics for the trading strategies for  $k = 1, 2, 3, 7$ 

	$k = 1$					$k = 2$				
	TW	AV	SD	SR	MDD	TW	AV	SD	SR	MDD
Buy and hold	104.63	12.65	15.00	0.17	0.50	104.63	12.65	15.00	0.17	0.50
Momentum										
3-month	32.39	9.18	10.69	0.15	0.23	32.39	9.18	10.69	0.15	0.23
6-month	48.71	10.26	11.30	0.17	0.23	48.71	10.26	11.30	0.17	0.23
12-month	100.21	12.15	12.15	0.20	0.30	100.21	12.15	12.15	0.20	0.30
Linear model	93.90	12.29	14.33	0.17	0.50	109.38	12.69	14.50	0.18	0.50
CSR approach	104.63	12.65	15.00	0.17	0.50	119.54	12.91	14.54	0.18	0.44
GARCH-M model	112.08	12.73	14.40	0.18	0.50	151.01	13.52	14.72	0.19*	0.50
MS model	116.96	12.59	12.56	0.20	0.29	59.60	10.92	12.58	0.17	0.48
Copula-based approach										
Gaussian	120.26	12.93	14.57	0.18	0.50	105.28	12.51	13.96	0.18	0.44
Frank	122.55	12.98	14.57	0.18	0.50	154.85	13.59	14.75	0.19*	0.50
Clayton	118.17	12.87	14.44	0.18	0.50	89.80	12.09	13.78	0.18	0.44
FGM	122.55	12.98	14.57	0.18	0.50	111.75	12.66	13.97	0.19	0.44
CSM approach										
Baseline	123.55	13.00	14.57	0.18	0.50	154.85	13.59	14.75	0.19*	0.50
Poly	122.55	12.98	14.57	0.18	0.50	96.89	12.27	13.71	0.18	0.44
	$k = 3$					$k = 7$				
	TW	AV	SD	SR	MDD	TW	AV	SD	SR	MDD
Buy and hold	104.63	12.65	15.00	0.17	0.50	104.63	12.65	15.00	0.17	0.50
Momentum										
3-month	32.39	9.18	10.69	0.15	0.23	32.39	9.18	10.69	0.15	0.23
6-month	48.71	10.26	11.30	0.17	0.23	48.71	10.26	11.30	0.17	0.23
12-month	100.21	12.15	12.15	0.20	0.30	100.21	12.15	12.15	0.20	0.30
Linear model	112.79	12.67	13.84	0.19	0.44	101.16	12.17	12.27	0.20	0.44
CSR approach	98.22	12.39	14.32	0.18	0.44	62.91	11.00	12.27	0.17	0.45
GARCH-M model	147.63	13.39	14.24	0.19	0.44	75.29	11.51	12.76	0.18	0.50
MS model	25.65	8.73	11.77	0.12	0.41	54.04	10.69	12.78	0.16	0.44
Copula-based approach										
Gaussian	165.23	13.67	14.22	0.20*	0.44	178.45	13.61	12.98	0.22*	0.44
Frank	165.23	13.67	14.22	0.20*	0.44	176.61	13.66	13.05	0.22*	0.44
Clayton	117.55	12.91	14.79	0.18	0.50	178.48	13.66	13.04	0.22*	0.44
FGM	165.23	13.67	14.22	0.20*	0.44	176.61	13.66	13.05	0.22*	0.44
CSM approach										
Baseline	181.68	13.89	14.13	0.21**	0.44	168.05	13.53	12.96	0.22	0.44
Poly	181.68	13.89	14.13	0.21**	0.50	149.16	13.24	12.99	0.21	0.44

*Notes:* This table presents out-of-sample performance statistics for the various trading strategies. TW represents the terminal wealth in dollars at the end of the trading period. AV is the annualized average return in percent, SD is the annualized standard deviation of returns in percent, SR is the Sharpe ratio, and MDD is the maximum drawdown in percent. Stars indicate statistical significance of the SR differential compared to the benchmark buy-and-hold strategy: \* and \*\* denote significance at the 10% and 5% levels, respectively.

Table 7: CER gains against historical average for  $\gamma = 5$ 

	$k = 1$	$k = 2$	$k = 3$	$k = 7$
Panel A: Mean-variance preferences				
Momentum				
3-month momentum	-0.706	-0.706	-0.706	-0.706
6-month momentum	0.042	0.042	0.042	0.042
12-month momentum	1.436	1.436	1.436	1.436
Linear model	0.127	0.403	0.854	1.384
CSR approach	0.000	0.596	0.239	0.207
GARCH-M model	0.526	1.075	1.365	0.412
MS model	1.616	-0.067	-1.607	-0.322
Copula-based approach				
Gaussian	0.603	0.618	1.590	2.442
Frank	0.649	1.122	1.590	2.380
Clayton	0.636	0.320	0.414	2.412
FGM	0.649	0.759	1.590	2.380
CSM approach				
Baseline	0.690	1.122	1.878	2.302
Poly	0.649	0.546	1.878	1.993
Panel B: CRRA preferences				
Momentum				
3-month momentum	-0.555	-0.555	-0.555	-0.555
6-month momentum	0.142	0.142	0.142	0.142
12-month momentum	1.557	1.557	1.557	1.557
Linear model	0.090	-0.259	0.098	0.558
CSR approach	0.303	0.303	0.303	0.544
GARCH-M model	0.485	1.062	1.365	0.675
MS model	1.714	0.085	-1.597	-0.187
Copula-based approach				
Gaussian	0.573	0.658	1.643	2.748
Frank	0.618	1.108	1.643	2.686
Clayton	0.595	0.350	0.407	2.719
FGM	0.618	0.798	1.643	2.686
CSM approach				
Baseline	0.637	1.108	1.939	2.612
Poly	0.618	0.584	1.939	2.279

*Notes:* This table reports the CER gains for each forecasting strategy relative to the historical average benchmark, under two types of investor preferences: mean-variance (Panel A) and constant relative risk aversion (CRRA; Panel B). The parameter  $k$  denotes the number of predictors used in each model. The CER gain is interpreted as the annual management fee (in percentage points) an investor would be willing to pay to switch from the benchmark to the corresponding strategy.

Table 8: CER gains against CSM (Baseline) for  $\gamma = 5$

	$k = 1$	$k = 2$	$k = 3$	$k = 7$
Panel A: Mean-variance preferences				
Copula-based approach				
Gaussian	-0.087	-0.504	-0.288	0.140
Frank	-0.041	0.000	-0.288	0.078
Clayton	-0.054	-0.802	-1.464	0.110
FGM	-0.041	-0.363	-0.288	0.078
CSM approach				
Poly	-0.041	-0.576	0.000	-0.309
Panel B: CRRA preferences				
Copula-based approach				
Gaussian	-0.064	-0.450	-0.296	0.136
Frank	-0.019	0.000	-0.296	0.074
Clayton	-0.042	-0.758	-1.532	0.107
FGM	-0.019	-0.310	-0.296	0.074
CSM approach				
Poly	-0.019	-0.524	0.000	-0.333

*Notes:* This table reports CER gains relative to the CSM (Baseline) benchmark, using the same setup and preferences as in Table 7. See the notes to that table for details on interpretation.

Figure 1: Out-of-sample  $R^2$  as a function of the number of predictors ( $k$ )

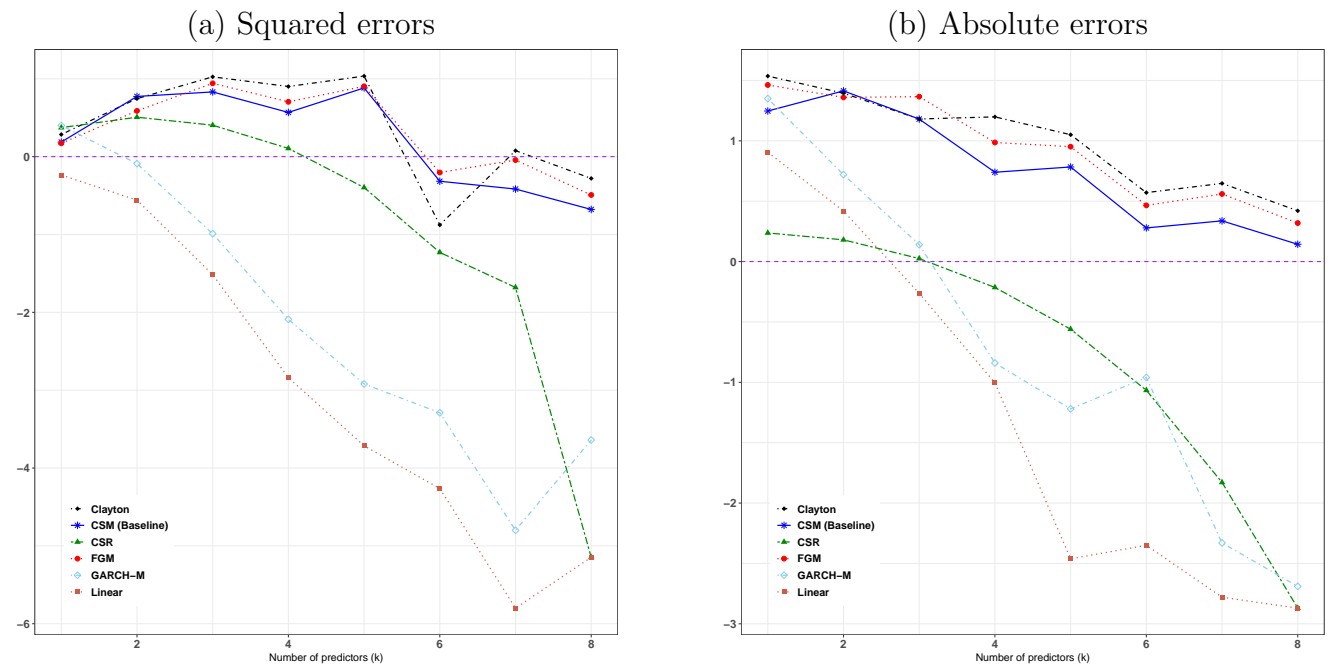


Figure 2: Wealth growth for the CSM (Baseline,  $k = 3$ ), buy-and-hold, and momentum strategies

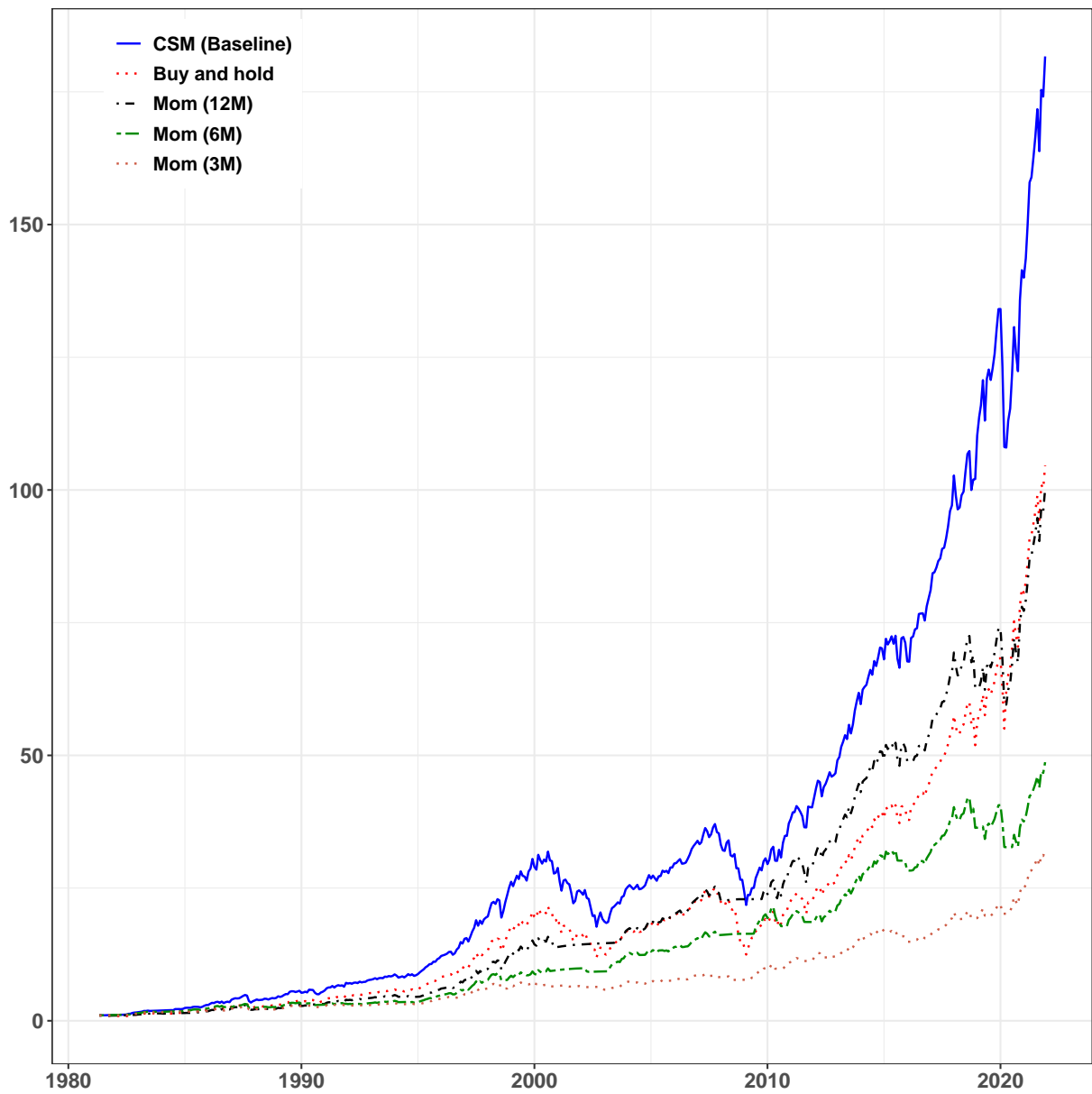


Figure 3: Wealth growth for the CSM (Baseline), copula-based, GARCH-M, CSR, and linear strategies, each with  $k = 3$

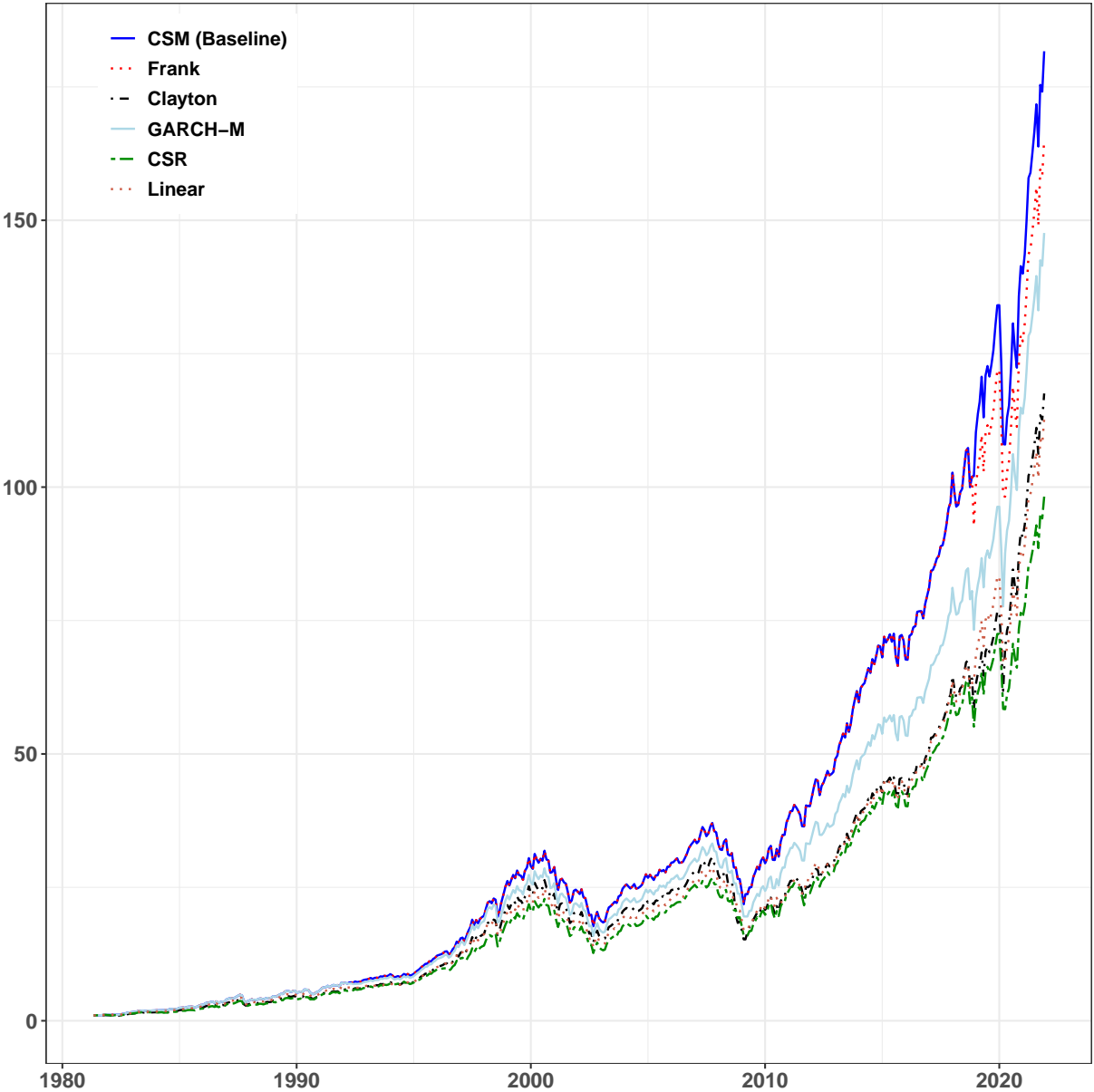
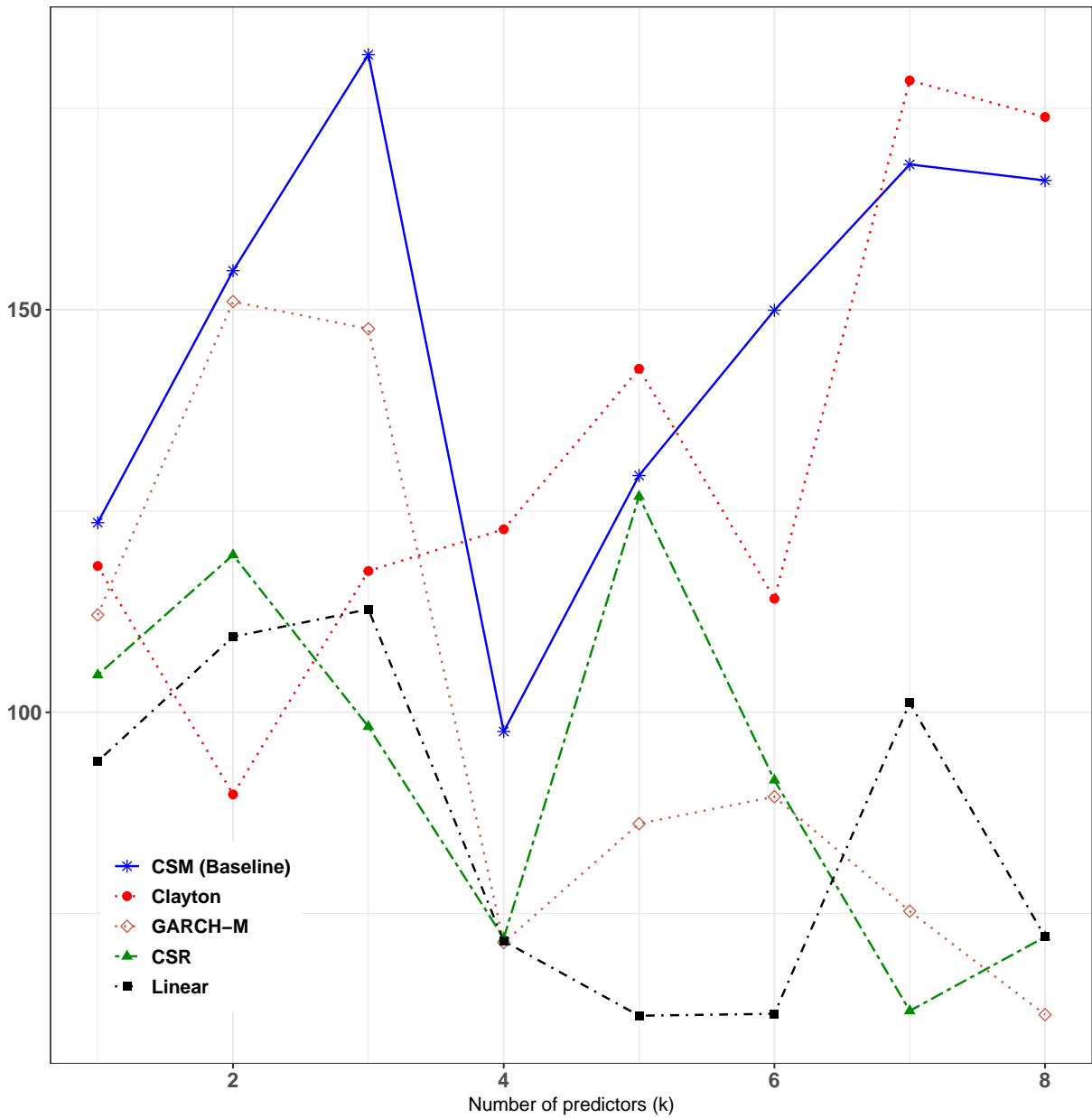


Figure 4: Terminal wealth as a function of the number of predictors ( $k$ ) for CSM (Baseline), Clayton, GARCH-M, CSR, and linear models



## Appendix. Copulas and $\varrho_t(z)$

This appendix presents the copula functions and the corresponding formulas for the deformed probability of success,  $\varrho_t(z)$ , for the Gaussian, Frank, Clayton, and FGM copulas. In each case,  $\varrho_t(z)$  is obtained by differentiating the copula function with respect to its first argument.

**Gaussian copula.** The Gaussian copula is

$$C(w_1, w_2) = \Phi_2(\Phi^{-1}(w_1), \Phi^{-1}(w_2)), \quad (18)$$

where  $\Phi_2(\cdot, \cdot)$  is the bivariate standard normal CDF with correlation parameter  $\theta_c \in (-1, +1)$ . When  $\theta_c = 0$ , the Gaussian copula collapses to the independence copula:  $C(w_1, w_2) = w_1 w_2$ . The deformed probability function  $\varrho_t(z)$  appearing in (9) is given as

$$\varrho_t(z) = \Phi\left(\frac{\Phi^{-1}(p_t) + \theta_c \Phi^{-1}(z)}{\sqrt{1 - \theta_c^2}}\right),$$

which equals the conditional probability  $p_t$  when  $\theta_c = 0$ .

**Frank copula.** The Frank copula is

$$C(w_1, w_2) = -\frac{1}{\theta_c} \log\left(1 + \frac{(e^{-\theta_c w_1} - 1)(e^{-\theta_c w_2} - 1)}{e^{-\theta_c} - 1}\right), \quad (19)$$

where  $-\infty < \theta_c < +\infty$ ,  $\theta_c \neq 0$ . The Frank copula tends to the independence copula as  $\theta_c \rightarrow 0$ . The deformed probability function  $\varrho_t(z)$  appearing in (9) is given by

$$\varrho_t(z) = \left(1 - \frac{1 - e^{-\theta_c(1-p_t)}}{1 - e^{\theta_c p_t}} e^{\theta_c(1-z)}\right)^{-1}, \quad \theta_c \neq 0,$$

and  $\varrho_t(z) \rightarrow p_t$  as  $\theta_c \rightarrow 0$ , recovering the marginal success probability under independence.

**Clayton copula.** The Clayton copula is

$$C(w_1, w_2) = (w_1^{-\theta_c} + w_2^{-\theta_c} - 1)^{-1/\theta_c}, \quad (20)$$

where  $\theta_c > 0$ . The Clayton copula tends to the independence copula as  $\theta_c \rightarrow 0^+$ . The function  $\varrho_t(z)$  from (9) is given by

$$\varrho_t(z) = 1 - \left( 1 + \frac{(1 - p_t)^{-\theta_c} - 1}{z^{-\theta_c}} \right)^{-1/\theta_c - 1}, \quad \theta_c > 0,$$

and  $\varrho_t(z) \rightarrow p_t$  as  $\theta_c \rightarrow 0^+$ , as expected under independence.

**FGM copula.** The FGM copula is

$$C(w_1, w_2) = w_1 w_2 (1 + \theta_c (1 - w_1)(1 - w_2)), \quad (21)$$

where  $\theta_c \in [-1, +1]$ . The FGM copula collapses to the independence copula when  $\theta_c = 0$ .

The function  $\varrho_t(z)$  in (9) is given by

$$\varrho_t(z) = 1 - (1 - p_t)(1 + \theta_c p_t (1 - 2z)), \quad \theta_c \in [-1, +1],$$

which equals  $p_t$  when  $\theta_c = 0$ .

Supplementary material for:

# A new decomposition approach to modeling financial returns: conditioning sign on magnitude

Arsène Brou and Richard Luger

## A. Results under MSE-based selection

This section reports additional results when best  $k$ -predictor subset selection uses the MSE criterion. The main text focuses on AUC-based selection; MSE-based selection provides a complementary robustness check.

Table A1 reports the selected predictor subsets for the linear model (Panel A), GARCH-M (Panel B), MS (Panel C), and the decomposition-based models (Panel D) across  $k = 1, \dots, 8$ . Selection overlap is substantial across models, even though the subsets are chosen separately for each  $k$  and need not be nested. In particular,  $dp$  is selected at  $k = 1$  in all cases, and  $dfy$  and  $tbl$  appear frequently at intermediate dimensions. The linear and decomposition-based selections coincide for  $k = 2-5$ , and differ only at  $k = 6-7$  before converging to the full predictor set at  $k = 8$ . At  $k = 6$ , the linear model selects  $infl$  (with  $dp$ ,  $dfy$ ,  $tbl$ ,  $ltr$ , and  $ntis$ ), whereas the decomposition-based models select  $dfr$  instead. At  $k = 7$ , the linear model includes  $tms$  and  $infl$  but excludes  $dfr$ , while the decomposition-based models include  $dfr$  and  $infl$  but exclude  $tms$ . Importantly, the decomposition-based models select identical subsets across copula specifications (Panel D); GARCH-M and MS broadly converge to similar selections at larger  $k$ , while differing more at small  $k$ .

Out-of-sample performance, measured by  $R_{\text{OOS}}^2$  (in %) under squared and absolute loss, is reported in Table A2 and summarized in Figure A1. The figure shows that the  $R_{\text{OOS}}^2$  profiles under squared loss (Panel A) for decomposition-based models peak at intermediate  $k$  before declining, whereas conventional benchmarks remain weak and tend to deteriorate at larger  $k$ .

Under absolute loss (Panel B), decomposition-based models deliver uniformly positive  $R_{\text{OOS}}^2$  values from  $k = 2$  onward, while the linear and GARCH-M benchmarks remain negative and CSR turns negative beyond  $k = 3$ .

Overall, MSE-based selection yields weaker performance than AUC-based selection. The deterioration is especially pronounced for conventional benchmarks: the linear model has uniformly negative  $R_{\text{OOS}}^2$  values (from  $-0.88\%$  to  $-5.15\%$  under squared loss and from  $-0.10\%$  to  $-2.87\%$  under absolute loss), and GARCH-M also produces negative  $R_{\text{OOS}}^2$  throughout (e.g., from  $-0.27\%$  to  $-3.97\%$  under squared loss). CSR achieves only modestly positive  $R_{\text{OOS}}^2$  at small  $k$  under squared loss, peaking at  $0.51\%$  at  $k = 2$ , and then deteriorates, ultimately coinciding with the linear model at  $k = 8$  by construction. The MS model performs particularly poorly, with squared-loss  $R_{\text{OOS}}^2$  values ranging from  $-6.23\%$  to  $-19.51\%$ ; frequent DM-test rejections against the historical average indicate statistically significant underperformance.

In contrast, the decomposition-based models are more resilient. Under squared loss, they typically produce positive  $R_{\text{OOS}}^2$  values for  $k = 2-5$ , with the Clayton copula reaching  $0.84\%$  at  $k = 3$  and CSM (Baseline) attaining  $0.58\%$  at  $k = 3$ . These gains attenuate at higher dimensions, with several specifications turning negative for  $k \geq 6$ , consistent with the declining profiles in Panel (a) of Figure A1.

Taken together, these results support our reliance on AUC-based subset selection in the main analysis: it delivers stronger and more stable out-of-sample performance across model classes, and it is particularly important for benchmark specifications (linear and GARCH-M) that deteriorate markedly under MSE-based selection.

Table A1: Best subset selection based on MSE criterion

	const	dp	dfy	tms	tbl	ltr	dfr	ntis	infl
Panel A: Linear model									
$k = 1$	*	*							
$k = 2$	*		*		*				
$k = 3$	*	*	*		*				
$k = 4$	*	*	*		*	*			
$k = 5$	*	*	*		*	*		*	
$k = 6$	*	*	*		*	*		*	*
$k = 7$	*	*	*	*	*	*		*	*
$k = 8$	*	*	*	*	*	*	*	*	*
Panel B: GARCH-M model									
$k = 1$	*	*							
$k = 2$	*	*		*					
$k = 3$	*	*	*	*					
$k = 4$	*	*		*	*			*	
$k = 5$	*	*	*		*	*			*
$k = 6$	*	*	*		*	*	*	*	
$k = 7$	*	*	*	*	*	*		*	*
$k = 8$	*	*	*	*	*	*	*	*	*
Panel C: MS model									
$k = 1$	*	*							
$k = 2$	*		*		*				
$k = 3$	*		*		*			*	
$k = 4$	*	*	*	*			*		
$k = 5$	*	*	*	*			*	*	
$k = 6$	*	*	*	*	*	*	*		
$k = 7$	*	*	*	*	*	*		*	*
$k = 8$	*	*	*	*	*	*	*	*	*
Panel D: Decomposition-based models (CSM and copula-based)									
$k = 1$	*	*							
$k = 2$	*		*		*				
$k = 3$	*	*	*		*				
$k = 4$	*	*	*		*	*			
$k = 5$	*	*	*		*	*		*	
$k = 6$	*	*	*		*	*	*	*	
$k = 7$	*	*	*		*	*	*	*	*
$k = 8$	*	*	*	*	*	*	*	*	*

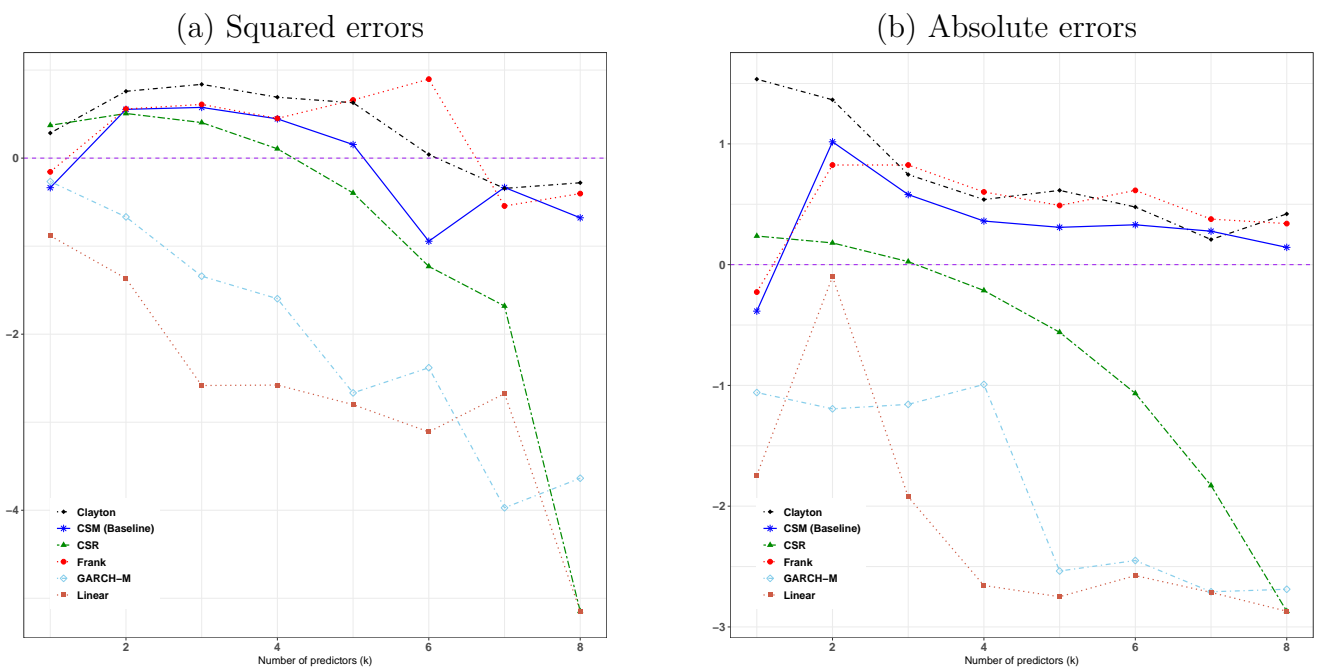
*Notes:* A star (\*) indicates that the corresponding variable is included in the best subset of size  $k$ . The subsets are selected based on the MSE criterion using one-step-ahead in-sample forecasts. An empty cell indicates exclusion of the variable from the selected subset.

Table A2: Out-of-sample  $R^2$  values (in %) for different models, loss functions, and subset sizes ( $k$ )

	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	Squared	Absolute	Squared	Absolute	Squared	Absolute	Squared	Absolute
Linear model	-0.88	-1.74	-1.37	-0.10	-2.58	-1.92	-2.58	-2.66
CSR approach	0.37	0.24	0.51	0.18	0.40	0.02	0.11	-0.21
MS	-13.66	-6.52	-6.23	-2.40	-7.31	-4.06	-9.61 <sup>***</sup>	-4.26 <sup>***</sup>
GARCH-M	-0.27	-1.06	-0.67	-1.19	-1.34	-1.16	-1.60	-0.99
Copula-based approach								
Gaussian	-0.22	-0.11	0.58	0.75	0.60	0.75	0.49	0.54
Frank	-0.16	-0.23	0.56	0.82	0.61	0.82	0.45	0.60
Clayton	0.29	1.54	0.76	0.93	0.84	0.99	0.69	0.69
FGM	0.17	1.46	0.55	0.80	0.60	0.80	0.42	0.59
CSM approach								
Baseline	-0.34	-0.38	0.55	0.58	0.58	0.58	0.45	0.36
Poly	-0.30	-0.25	0.38	0.75	0.58	0.75	0.26	0.46
	$k = 5$		$k = 6$		$k = 7$		$k = 8$	
	Squared	Absolute	Squared	Absolute	Squared	Absolute	Squared	Absolute
Linear model	-2.80	-2.75	-3.11	-2.57	-2.67	-2.71	-5.15	-2.87
CSR approach	-0.40	-0.56	-1.23	-1.07	-2.68	-1.83	-5.15	-2.87
MS	-10.69 <sup>***</sup>	-5.08 <sup>***</sup>	-14.71 <sup>***</sup>	-6.38 <sup>**</sup>	-19.51 <sup>***</sup>	-9.84 <sup>***</sup>	-16.93 <sup>***</sup>	-7.21 <sup>***</sup>
GARCH-M	-2.67	-2.54	-2.38	-2.45	-3.97	-2.71	-3.64	-2.69
Copula-based approach								
Gaussian	0.70	0.43	-0.02	0.48	-0.52	0.21	-0.39	0.35
Frank	0.66	0.49	0.90	0.62	-0.54	0.38	-0.40	0.34
Clayton	0.63	0.50	0.04	0.41	-0.34	0.28	-0.28	0.42
FGM	0.64	0.31	0.30	0.35	-1.12	0.30	-0.49	0.32
CSM approach								
Baseline	0.15	0.33	-0.94	0.36	-0.33	0.04	-0.68	0.14
Poly	-1.12	0.35	-1.12	0.30	-1.12	0.30	-0.99	0.35

*Notes:* This table reports  $R^2_{\text{OOS}}$  statistics (in %) for squared and absolute loss functions, relative to the historical-average benchmark. Positive values indicate improvement over the historical average, while negative values indicate deterioration. Stars indicate rejection of equal predictive accuracy (DM test) against the historical average benchmark: \*, \*\*, and \*\*\* denote significance at the 10%, 5%, and 1% levels, respectively.

Figure A1: Out-of-sample  $R^2$  as a function of the number of predictors ( $k$ ): MSE-based selection



## B. Additional economic value results

This appendix complements the main text by reporting results for additional predictor subset sizes ( $k = 4, 5, 6, 8$ ), extending the analysis beyond the main text’s focus on  $k = 1, 2, 3, 7$ . Table B1 reports out-of-sample performance statistics for these subset sizes. The CSM (Poly) specification performs strongly at intermediate dimensions, achieving terminal wealth of \$139.78 at  $k = 4$ , \$148.04 at  $k = 5$ , and \$137.85 at  $k = 6$ , which is substantially above the linear model (\$71.58, \$62.31, and \$65.92) and also above CSR at these  $k$  values. At larger  $k$ , the two CSM variants become very similar (TW \$166.06 for CSM Baseline and \$166.86 for CSM Poly at  $k = 8$ ). Copula-based strategies are highly competitive at larger subset sizes, with the Clayton specification delivering the highest terminal wealth among the copulas at  $k = 8$  (TW \$173.94). Overall, decomposition-based strategies dominate the conventional benchmarks (linear, GARCH-M, and MS) across these subset sizes, while the MS strategy remains comparatively weak.

Table B2 reports certainty equivalent return (CER) gains relative to the historical average benchmark for an investor with moderate risk aversion ( $\gamma = 5$ ), under mean-variance preferences (Panel A) and CRRA preferences (Panel B). Across  $k$ , decomposition-based strategies deliver positive and economically meaningful CER gains, with particularly large gains at  $k = 8$ . At this subset size, both CSM variants exceed 2% CER under each preference (e.g., 2.267–2.324 in Panel A and 2.579–2.615 in Panel B), and the copula-based strategies are also strongly positive, with Clayton among the top performers (especially under CRRA). In contrast, the MS benchmark produces negative CER gains throughout, while GARCH-M is negative at  $k = 4$  but becomes positive for  $k = 5, 6, 8$ , albeit with gains that remain modest relative to the decomposition-based strategies.

Finally, Table B3 reports CER differentials computed relative to the CSM (Baseline) benchmark, which can be interpreted as the additional annual fee (in percentage points) an

investor would be willing to pay to switch from the baseline to an alternative decomposition-based strategy. The CSM (Poly) specification delivers sizable incremental gains at  $k = 4$  and  $k = 6$ , and smaller but still positive improvements at  $k = 5$  and  $k = 8$ . Among the copula-based alternatives, Clayton delivers positive differentials across all subset sizes, whereas Gaussian provides smaller and at times negative differentials, particularly at  $k = 5$  and  $k = 8$ , while Frank turns negative only at  $k = 8$ .

Overall, these supplementary results confirm that the economic value of the decomposition framework remains strong across predictor dimensions. While copula-based strategies can be highly competitive at larger subset sizes, the CSM specifications provide a robust alternative, and the polynomial variant delivers clear incremental utility at intermediate dimensions.

Table B1: Out-of-sample performance statistics for the trading strategies for  $k = 4, 5, 6, 8$ 

	$k = 4$					$k = 5$				
	TW	AV	SD	SR	MDD	TW	AV	SD	SR	MDD
Buy and hold	104.63	12.65	15.00	0.17	0.50	104.63	12.65	15.00	0.17	0.50
Momentum										
3-month	32.39	9.18	10.69	0.15	0.23	32.39	9.18	10.69	0.15	0.23
6-month	48.71	10.26	11.30	0.17	0.23	48.71	10.26	11.30	0.17	0.23
12-month	100.21	12.15	12.15	0.20	0.30	100.21	12.15	12.15	0.20	0.30
Linear model	71.58	11.54	13.86	0.16	0.46	62.31	11.03	12.71	0.17	0.46
CSR approach	71.94	11.57	13.91	0.16	0.45	126.80	12.81	12.87	0.20	0.44
GARCH-M model	71.40	11.63	14.47	0.16	0.50	86.16	12.01	13.96	0.17	0.46
MS model	40.81	10.03	13.01	0.14	0.44	32.36	9.48	11.73	0.13	0.44
Copula-based approach										
Gaussian	110.36	12.70	14.46	0.18	0.45	127.65	12.96	13.82	0.19	0.50
Frank	110.66	12.74	14.67	0.18	0.51	142.66	13.18	13.41	0.20	0.45
Clayton	122.73	12.70	14.55	0.18	0.45	142.67	13.26	13.93	0.20	0.50
FGM	110.66	12.74	14.67	0.18	0.51	148.34	13.35	13.94	0.20	0.50
CSM approach										
Baseline	97.61	12.39	14.42	0.17	0.45	129.41	12.98	13.72	0.20	0.50
Poly	139.78	13.28	14.41	0.19*	0.50	148.04	13.36	14.03	0.20	0.50
	$k = 6$					$k = 8$				
	TW	AV	SD	SR	MDD	TW	AV	SD	SR	MDD
Buy and hold	104.63	12.65	15.00	0.17	0.50	104.63	12.65	15.00	0.17	0.50
Momentum										
3-month	32.39	9.18	10.69	0.15	0.23	32.39	9.18	10.69	0.15	0.23
6-month	48.71	10.26	11.30	0.17	0.23	48.71	10.26	11.30	0.17	0.23
12-month	100.21	12.15	12.15	0.20	0.30	100.21	12.15	12.15	0.20	0.30
Linear model	65.92	11.26	13.34	0.16	0.44	72.10	11.29	11.90	0.18	0.45
CSR approach	91.57	11.99	12.72	0.19	0.45	72.10	11.29	11.90	0.18	0.45
GARCH-M model	89.51	12.11	13.99	0.17	0.44	62.44	10.95	12.08	0.17	0.42
MS model	68.57	11.31	12.93	0.17	0.45	42.51	10.02	12.18	0.15	0.43
Copula-based approach										
Gaussian	102.82	12.38	13.47	0.19	0.45	167.11	13.52	13.01	0.22	0.44
Frank	105.97	12.47	13.56	0.19	0.45	168.24	13.54	13.04	0.22	0.44
Clayton	114.10	12.65	13.53	0.19	0.45	173.94	13.63	13.01	0.22	0.44
FGM	102.81	12.39	13.57	0.18	0.45	168.24	13.54	13.04	0.22	0.44
CSM approach										
Baseline	91.52	12.09	13.46	0.18	0.45	166.06	13.50	12.97	0.22	0.44
Poly	137.85	13.18	13.97	0.20	0.50	166.86	13.50	12.89	0.22	0.50

*Notes:* This table presents out-of-sample performance statistics for the various trading strategies. TW represents the terminal wealth in dollars at the end of the trading period. AV is the annualized average return in percent, SD is the annualized standard deviation of returns in percent, SR is the Sharpe ratio, and MDD is the maximum drawdown in percent. Stars indicate statistical significance of the SR differential compared to the benchmark buy-and-hold strategy: \* and \*\* denote significance at the 10% and 5% levels, respectively.

Table B2: CER gains against historical average for  $\gamma = 5$ 

	$k = 4$	$k = 5$	$k = 6$	$k = 8$
Panel A: Mean-variance preferences				
Momentum				
3-month momentum	-0.706	-0.706	-0.706	-0.706
6-month momentum	0.042	0.042	0.042	0.042
12-month momentum	1.436	1.436	1.436	1.436
Linear model	-0.286	-0.408	-0.209	0.726
CSR approach	-0.306	1.649	0.916	0.726
GARCH-M model	-0.640	0.112	0.199	0.278
MS model	-1.232	-1.850	-0.330	-0.670
Copula-based approach				
Gaussian	0.448	1.163	0.818	2.262
Frank	0.334	1.470	0.843	2.264
Clayton	0.659	1.377	1.048	2.360
FGM	0.334	1.470	0.762	2.264
CSM approach				
Baseline	0.170	1.252	0.530	2.267
Poly	1.064	1.414	1.270	2.324
Panel B: CRRA preferences				
Momentum				
3-month momentum	-0.555	-0.555	-0.555	-0.555
6-month momentum	0.142	0.142	0.142	0.142
12-month momentum	1.557	1.557	1.557	1.557
Linear model	-0.220	0.240	-0.131	1.073
CSR approach	0.987	0.987	1.263	1.073
GARCH-M model	-0.636	0.189	0.274	0.631
MS model	-1.112	-1.788	-0.251	-0.324
Copula-based approach				
Gaussian	0.445	1.365	1.029	2.570
Frank	0.325	1.672	1.057	2.572
Clayton	0.663	1.581	1.255	2.667
FGM	0.325	1.672	0.977	2.572
CSM approach				
Baseline	0.172	1.462	0.747	2.579
Poly	1.077	1.614	1.473	2.615

*Notes:* This table reports the CER gains for each forecasting strategy relative to the historical average benchmark, under two types of investor preferences: mean-variance (Panel A) and constant relative risk aversion (CRRA; Panel B). The parameter  $k$  denotes the number of predictors used in each model. The CER gain is interpreted as the annual management fee (in percentage points) an investor would be willing to pay to switch from the benchmark to the corresponding strategy.

Table B3: CER gains against CSM (Baseline) for  $\gamma = 5$

	$k = 4$	$k = 5$	$k = 6$	$k = 8$
Panel A: Mean-variance preferences				
Copula-based approach				
Gaussian	0.278	-0.089	0.288	-0.005
Frank	0.164	0.218	0.313	-0.003
Clayton	0.489	0.125	0.518	0.093
FGM	0.164	0.218	0.232	-0.003
CSM approach				
Poly	0.894	0.162	0.740	0.057
Panel B: CRRA preferences				
Copula-based approach				
Gaussian	0.273	-0.097	0.282	-0.009
Frank	0.153	0.210	0.310	-0.007
Clayton	0.491	0.119	0.508	0.088
FGM	0.153	0.210	0.230	-0.007
CSM approach				
Poly	0.905	0.152	0.726	0.036

*Notes:* This table reports CER gains relative to the CSM (Baseline) benchmark, using the same setup and preferences as in Table B2. See the notes to that table for details on interpretation.

## C. Economic performance during crisis periods

This appendix evaluates the economic performance of the forecasting strategies during two major crisis episodes: the Global Financial Crisis (GFC) and the COVID-19 crisis period. The GFC is defined as August 2007 to March 2009, while the COVID-19 crisis period spans December 2019 to December 2021 (the end of our sample). At the beginning of each crisis, models are estimated using the most recent 400 available observations, and portfolio performance is then evaluated from the start to the end of the crisis using a rolling window of 400 observations, as in the main analysis. We normalize initial wealth to \$1 (so TW is a gross wealth multiple). To ensure a consistent comparison across approaches, all models are estimated using the full set of eight predictors (i.e.,  $k = 8$  throughout this appendix).<sup>1</sup> Table C1 reports performance statistics over both periods, and Figures C1 and C2 plot the corresponding cumulative wealth paths.

During the GFC, most model-based switching strategies deliver negative average returns and Sharpe ratios (Table C1), although they generally attenuate losses relative to buy-and-hold. Buy-and-hold exhibits the weakest performance (TW = 0.59 and MDD = 0.50). Momentum strategies are comparatively resilient during this window, preserving capital and achieving modest gains (TW  $\approx$  1.04–1.09). The decomposition-based strategies (CSM and copula-based specifications) deliver very similar outcomes during the GFC (TW = 0.80), close to the linear/MS/CSR benchmarks (TW between 0.79 and 0.82). Figure C1 highlights that CSM (Baseline) and the FGM copula strategy follow essentially identical wealth paths throughout the crisis, declining substantially less than buy-and-hold but underperforming short-horizon momentum.

The COVID-19 crisis period shows a different pattern. All strategies deliver positive aver-

---

<sup>1</sup>Under  $k = 8$ , the CSR and linear strategies coincide by construction. In addition, several decomposition-based strategies yield identical switching decisions over these short windows, which is reflected by overlapping wealth paths in Figures C1 and C2.

Table C1: Out-of-sample performance statistics for the trading strategies during crisis periods

	Global Financial Crisis					COVID-19 Crisis				
	TW	AV	SD	SR	MDD	TW	AV	SD	SR	MDD
Buy and hold	0.59	-29.98	21.26	-0.43	0.50	1.57	23.51	19.05	0.35	0.19
Momentum										
3-month	1.09	5.42	5.84	0.17	0.04	1.95	33.37	14.32	0.66	0.06
6-month	1.09	5.47	3.14	0.32	0.05	1.64	24.84	12.44	0.57	0.06
12-month	1.04	2.25	4.79	0.02	0.05	1.70	27.41	17.70	0.44	0.12
Linear model	0.79	-12.15	16.18	-0.25	0.27	1.79	29.63	16.08	0.52	0.08
CSR approach	0.79	-12.15	16.18	-0.25	0.27	1.79	29.63	16.08	0.52	0.08
GARCH-M model	0.77	-14.04	16.32	-0.28	0.29	1.58	23.82	19.01	0.35	0.19
MS model	0.82	-10.31	16.62	-0.22	0.24	1.79	29.63	16.08	0.52	0.08
Copula-based approach										
Gaussian	0.80	-11.97	16.09	-0.25	0.26	1.57	23.51	19.05	0.35	0.19
Frank	0.80	-11.97	16.09	-0.25	0.26	1.57	23.51	19.05	0.35	0.19
Clayton	0.80	-11.97	16.09	-0.25	0.26	1.57	23.51	19.05	0.35	0.19
FGM	0.80	-11.97	16.09	-0.25	0.26	1.57	23.51	19.05	0.35	0.19
CSM approach										
Baseline	0.80	-11.97	16.09	-0.25	0.26	1.57	23.51	19.05	0.35	0.19
Poly	0.80	-11.97	16.09	-0.25	0.26	1.79	29.63	16.08	0.52	0.08

*Notes:* See Table B1 for details on notation.

age returns and Sharpe ratios, with 3-month momentum achieving the strongest performance (TW = 1.95). A second group of strategies—Linear/CSR/MS and CSM (Poly)—coincides in this episode and attains strong performance (TW = 1.79). Buy-and-hold, GARCH-M, and the remaining decomposition-based specifications exhibit more moderate but still positive outcomes (TW  $\approx$  1.57–1.58). Figure C2 illustrates the associated overlaps in wealth paths: the FGM copula and buy-and-hold coincide (TW = 1.57), as do CSR and CSM (Poly) (TW = 1.79), while 3-month momentum maintains a clear lead over the full window.

Overall, within these two crisis windows, momentum—particularly at short horizons—is the most consistently resilient strategy. Among the model-based switching approaches, performance is broadly similar during the GFC, whereas during COVID the strongest non-momentum performance is achieved by the Linear/CSR/MS and CSM (Poly) strategies, with buy-and-hold performing worst during the GFC.

Figure C1: Wealth growth during the Global Financial Crisis: 3-month momentum, CSR, GARCH-M, and buy-and-hold, together with the CSM (Baseline) and FGM copula strategies. The CSM (Baseline) and FGM wealth paths coincide and therefore overlap.

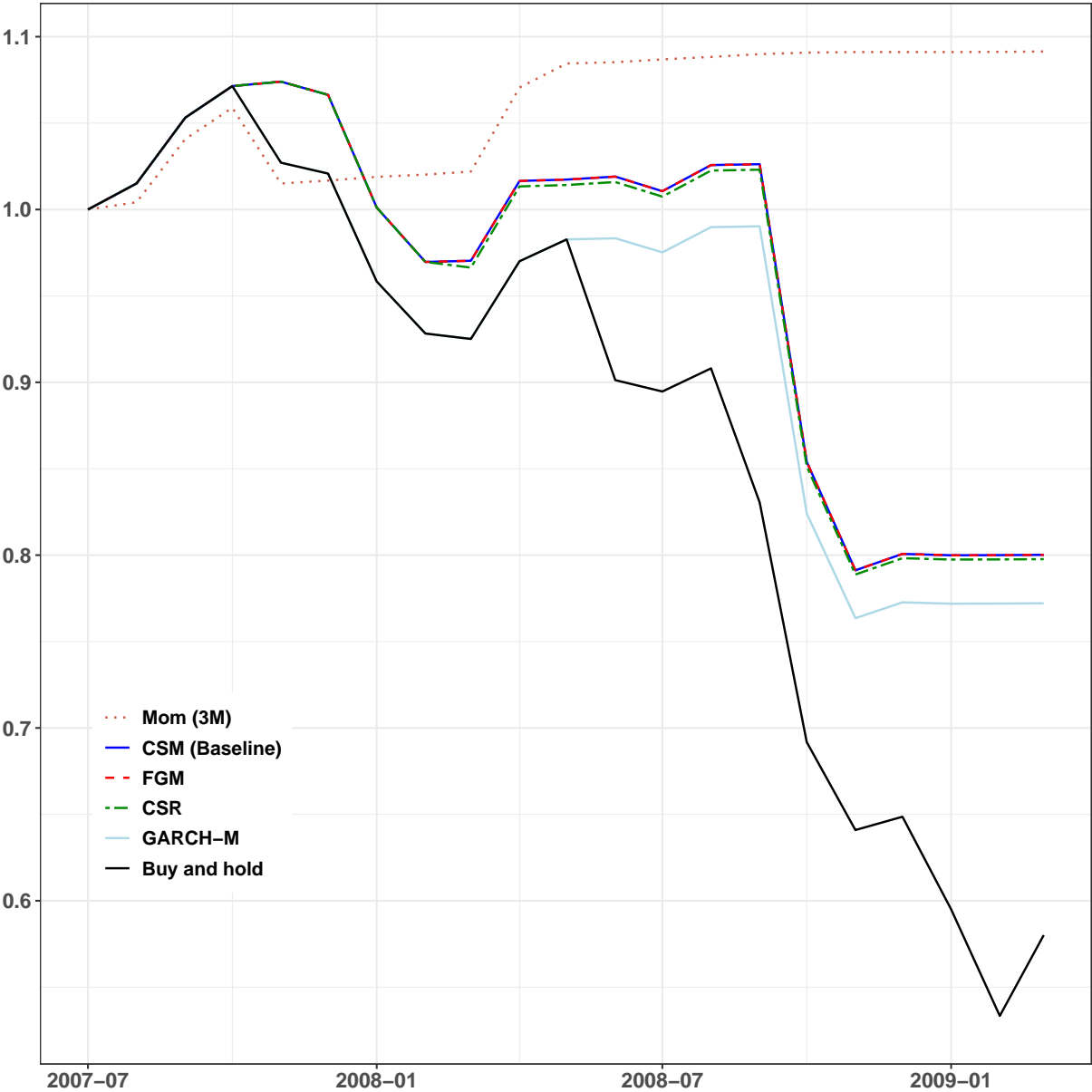
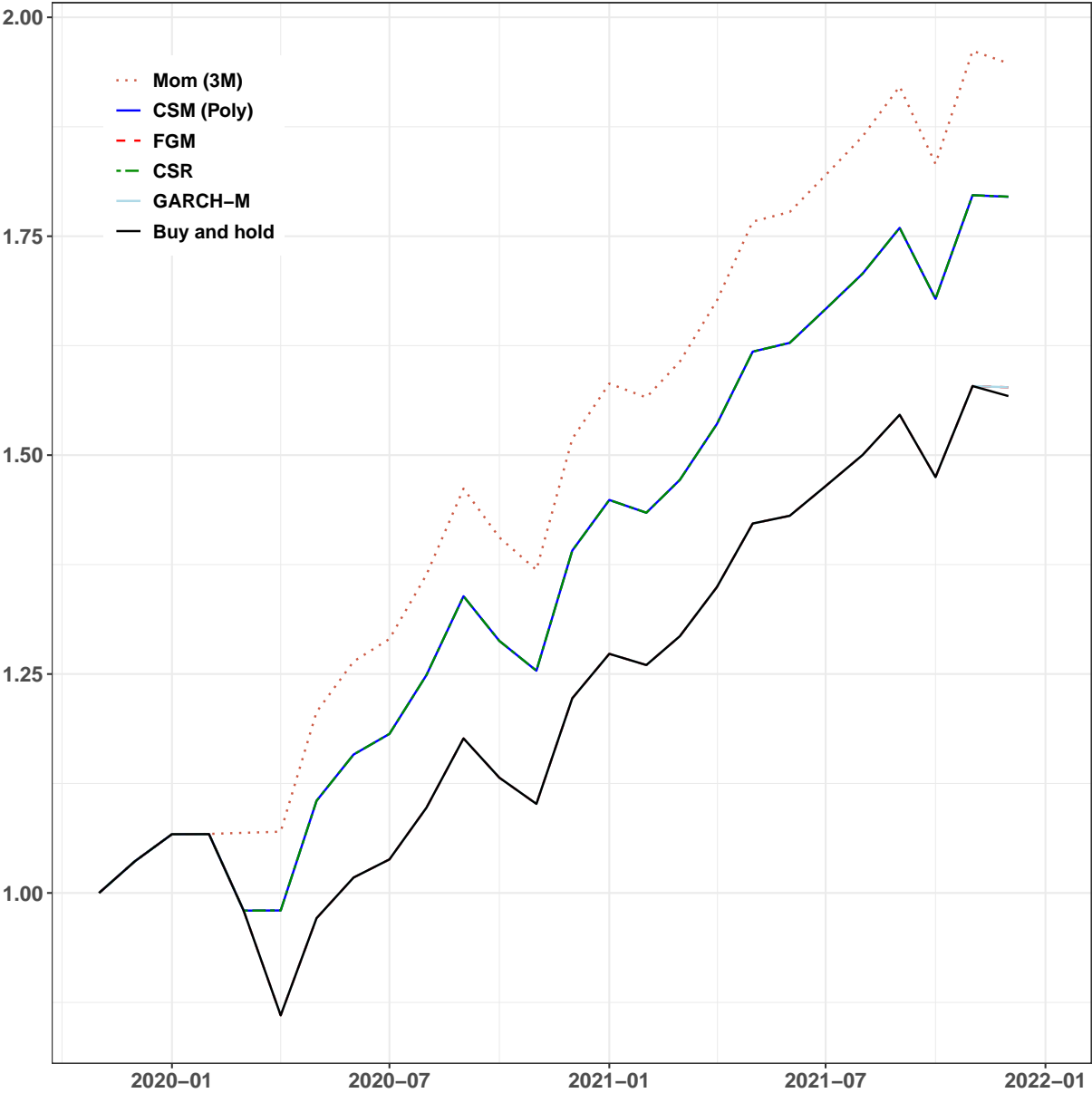


Figure C2: Wealth growth during the COVID crisis period: 3-month momentum, CSM (Poly), FGM copula, CSR, GARCH-M, and buy-and-hold. The FGM and buy-and-hold wealth paths coincide, as do the CSR and CSM (Poly) paths, and the corresponding curves therefore overlap.



## D. Additional results with momentum variable

This appendix reports results from an augmented predictor set that adds a standard time-series momentum indicator to the eight predictors used in the main analysis. We construct the signal following Moskowitz et al. (2012). Using the monthly S&P 500 returns, the momentum variable at time  $t - 1$  is defined as the cumulative return from month  $t - 12$  to  $t - 2$ , excluding the most recent month to mitigate short-term reversal effects (Jegadeesh and Titman, 1993; Moskowitz et al., 2012):

$$\text{Mom}_{t-1} = \prod_{j=t-12}^{t-2} (1 + r_j) - 1,$$

where  $r_j$  denotes the simple monthly return.

We then re-estimate all predictive models using the full set of nine predictors. Table D1 reports out-of-sample  $R^2$  statistics and the corresponding portfolio performance, while Figure D1 plots the associated cumulative wealth paths for a representative subset of strategies. Under squared loss, all models deliver negative  $R_{\text{OOS}}^2$ . The deterioration is severe for the linear/CSR and regime-switching benchmarks (e.g.,  $-5.15\%$  for the linear model and CSR, and  $-21.97\%$  for MS), whereas the decomposition-based models are substantially closer to the historical-average benchmark (ranging from  $-0.42\%$  to  $-1.12\%$ ). Under absolute loss, the decomposition-based approaches again provide the strongest relative statistical performance: the copula-based specifications attain small but positive  $R_{\text{OOS}}^2$  values ( $0.15\text{--}0.27\%$ ), while CSM with the polynomial sign specification performs comparably ( $0.23\%$ ); the baseline CSM specification is close to zero ( $-0.01\%$ ). Since the augmented specification uses all nine predictors, CSR equals the linear model here, explaining the identical  $R_{\text{OOS}}^2$  and portfolio statistics.

Figure D1 illustrates the temporal wealth evolution for selected strategies. The decomposition-based approaches (CSM Baseline, Clayton, Gaussian) track similar paths

throughout the sample period and end at markedly higher wealth levels than the benchmarks. CSR follows a broadly similar trajectory early in the sample but begins to lag behind the decomposition-based strategies around the late-1990s/early-2000s, with further relative slippage during subsequent turbulent episodes. The MS strategy exhibits the weakest overall wealth accumulation among the plotted benchmarks, consistent with its very poor  $R_{\text{OOS}}^2$  values.

The economic evaluation leads to a similar conclusion. Adding momentum does not materially improve the performance of the linear/CSR, GARCH-in-mean, or regime-switching strategies, which remain dominated in this exercise. Among decomposition-based strategies, the copula-based specifications deliver the highest terminal wealth (TW  $\approx$  \$118–\$123, with Clayton achieving \$122.78) with Sharpe ratios around 0.20, while CSM remains competitive (TW  $\approx$  \$110–\$111) and substantially improves on the non-decomposition benchmarks (Linear/CSR TW = \$72.10; GARCH-M TW = \$69.87; MS TW = \$62.86).

Overall, incorporating a momentum indicator does not change the main message: modeling returns through sign-magnitude decompositions remains the most reliable approach in this application, while linear and regime-switching benchmarks remain weak when evaluated out-of-sample.

## References

- Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(1), 65–91.
- Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012). Time series momentum. *Journal of Financial Economics* 104(2), 228–250.

Table D1: Out-of-sample  $R^2$  and economic performance across models

	$R^2_{\text{OOS}}$		Portfolio performance				
	Squared	Absolute	TW	AV	SD	SR	MDD
Linear model	-5.15	-2.87	72.10	11.29	11.90	0.18	0.45
CSR approach	-5.15	-2.87	72.10	11.29	11.90	0.18	0.45
GARCH-M model	-4.36	-2.77	69.87	11.44	12.19	0.18	0.42
MS model	-21.97	-10.52	62.86	10.94	11.83	0.18	0.38
Copula-based approach							
Gaussian	-0.60	0.15	120.66	12.72	13.10	0.20	0.46
Frank	-0.42	0.27	118.29	12.67	13.06	0.20	0.46
Clayton	-0.46	0.26	122.78	12.77	13.10	0.20	0.46
FGM	-0.59	0.16	118.29	12.67	13.06	0.20	0.46
CSM approach							
Baseline	-0.91	-0.01	110.32	12.50	13.07	0.19	0.46
Poly	-1.12	0.23	111.12	12.49	12.86	0.20	0.46

*Notes:*  $R^2_{\text{OOS}}$  is computed relative to the historical-average benchmark. TW denotes terminal wealth (initial wealth \$1), AV and SD are annualized in percent, SR is the Sharpe ratio, and MDD is maximum drawdown in percent.

Figure D1: Wealth growth for CSM (Baseline), copula-based (Clayton, Gaussian), CSR, and MS strategies with nine predictors (including momentum).

