Portfolio return prediction and risk price heterogeneity

Nick Taylor*
University of Bristol

Abstract

A model of portfolio return dynamics is considered in which the price of risk is permitted to be heterogeneous. In doing this, a novel method is proposed that delivers improved out-of-sample forecasts of portfolio returns. The main innovation is the use of a set of predictors that account for variation in risk prices across (segmented) markets. These predictors are the conditional covariances between the returns to the components of the portfolio under consideration and commonly used state variables (that is, French-French factor returns). The results indicate that the proposed method dominates competing methods (including those that assume homogeneous risk prices) when applied to domestic and international data – a finding that is robust to the sample period, performance measure and the state variables used. The use of clustered conditional covariances leads to further improvements in out-of-sample performance.

Key Words: Prediction, risk price, portfolio returns, ICAPM, disaggregation information.

JEL Classification Codes: C10, C22, C53, C58, G12, G17.

*University of Bristol Business School, Bristol, BS8 1TN, UK. `Email: nick.taylor@bristol.ac.uk`.

1 Introduction

There is enormous interest in understanding the dynamics of stock returns. The ability to predict future returns brings obvious benefits to investors in terms of optimal decision making from a speculator and risk management perspective, companies in terms of strategic decisions relating to resource allocation, and to society in general in terms of financial stability, higher economic growth, and prosperity. This interest has produced a vast literature that debates whether indeed (market) portfolio returns are predictable (or not); see, e.g., Goyal et al. (2023) for a recent assessment.[1] Any empirical finding of predictability can be rationalised in one of three ways: it reflects the misuse of statistical methods leading to data snooping (Bossaerts and Hillion, 1999), and/or it reflects behavioural factors such as irrational investor behaviour (Cutler et al., 1989), and/or it reflects time-varying risk premia (Campbell and Cochrane, 1999). The current paper is consistent with the latter rationale by demonstrating that improved forecasts of portfolio returns are possible via a simple augmentation of a widely used (rational) economic model such that the time-varying risk premia is expanded to include those associated with components of the portfolio under investigation. The motivation for this expansion of the predictor set rests on the frictions to trade (limited arbitrage) that exist such that the price of risk is allowed to vary over assets and markets (and hence is applicable to markets that are segmented).

A characteristic feature of integrated (non-segmented) markets is the existence of a single price of risk across assets (and markets). That is, compensation for taking on a unit of risk is the same irrespective of the asset (or market). For instance, the capital asset price model (CAPM) predicts that the price of beta risk is given by the (ex-ante) risk premium to the market portfolio (that is, the conditional expectation of the market portfolio return over the risk-free rate). There is good reason to believe that this homogeneous risk price should exist: differences in risk prices across markets are expected to be arbitraged to ensure a single price of risk. However, impediments to trade such as information frictions (Merton, 1987) or margin constraints (Garleanu and Pedersen, 2011) and differences in investors' preferences across asset classes (Koijen et al., 2022) mean that arbitrage is often limited (Shleifer and Vishny, 1997). Indeed, there is ample empirical evidence that such risk

[1] Seminal studies in the area include Campbell and Thompson (2008), Welch and Goyal (2008), and Rapach et al. (2010). See Rapach and Zhou (2013) for an extensive review. The proliferation of predictors in recent years has led to several studies employing highly sophisticated techniques that involve machine learning methods to produce more parsimonious predictor sets; see, e.g., Rapach and Zhou (2020).

prices appear heterogeneous. For instance, Patton and Wellor (2022) are able to apply a statistical technique in order to identify clusters of stocks to maximise the cross-sectional variation in risk prices (cf. Fama and French, 1993, Foerster and Karolyi, 1999, and Griffin, 2002, who identify groups using a priori information). Irrespective of the approach, cross-sectional variation in risk prices appears significant. The current paper considers how this feature affects the time-series dynamics of portfolio returns (that is, returns obtained via cross-sectional aggregation across the components of the portfolio).

A seminal model of ex ante returns is the intertemporal CAPM (ICAPM) introduced by Merton (1973). This model allows expected returns to each asset to be determined by the conditional covariances of the asset's return with a set of latent state variables (which, in turn, capture changes in the opportunity set faced by investors). The sensitivities of each expected return to the covariances is permitted to vary over time, but not to vary over the assets at each point in time – as the sensitivities (that is, risk prices) are assumed to be homogeneous (see, e.g., Scruggs, 1998, Guo and Whitelaw, 2006, Guidolin and Timmermann, 2008, Bali and Engle, 2010, and Rossi and Timmermann, 2015, for applications in which this assumption is maintained). A key implication of this model is that when one aggregates in the cross-sectional dimension (over the component assets within a portfolio), the expected portfolio return will be a function of the conditional covariances between the portfolio return and each of the state variables. If, however, risk prices are heterogeneous, then it follows that this aggregation will produce a specification for expected portfolio returns in which all the conditional covariances between the component asset returns and the state variables are present.

To illustrate the proposed approach, consider the highly simplified case. Assume that an equal-weighted portfolio consists of two assets (with respective returns r_1 and r_2), and a single state variable (with return s). The portfolio return (r) is our variable of interest and is given by the (cross-sectional) mean of the returns to the assets (hence $r = (r_1 + r_2)/2$). Ex ante returns to these assets (with information set $\mathcal{F}$) are given by

$$\mathrm{E}(r_1|\mathcal{F}) = b_1 \operatorname{cov}(r_1, s|\mathcal{F}), \quad \mathrm{E}(r_2|\mathcal{F}) = b_2 \operatorname{cov}(r_2, s|\mathcal{F}),$$

where b_1 and b_2 represent the risk prices associated with the two assets. Ex ante portfolio returns will depend on risk prices such that two different specifications are implied:

$$\mathrm{E}(r|\mathcal{F}) = \begin{cases} b_1 \operatorname{cov}(r, s|\mathcal{F}), & \text{if } b_1 = b_2 \text{ (homogeneous risk prices)}, \\ b_1 \operatorname{cov}(r, s|\mathcal{F}) + (b_2 - b_1) \operatorname{cov}(r_2, s|\mathcal{F})/2, & \text{if } b_1 \neq b_2 \text{ (heterogeneous risk prices)}. \end{cases}$$

Thus, in the presence of heterogeneous risk prices (that is, $b_1 \neq b_2$), a predictor equation based on the first specification (which coincides with the standard ICAPM approach to predictability) will be misspecified.[2] The proposed approach uses a predictor equation based on the second of these specifications. This amounts to including component conditional covariances (hence disaggregated information) in the predictor equation (that is, $\operatorname{cov}(r_2, s|\mathcal{F})$ in this simple case). It is the statistical and economic significance of this specification (generalised to more than two assets, more than one state variable, and inclusion of commonly used predictors) that we investigate. To the best of our knowledge this is the first such study to investigate the time-series implications of risk price heterogeneity on aggregate returns (cf. Cong et al., 2022, and Patton and Wellor, 2022, study cross-sectional asset pricing implications, and Evgeniou et al., 2023, investigate firm-level return predictability). In doing this, the results also add to the ongoing debate on the predictability of the equity premium.

The results also contribute to the wider debate on the benefits of using disaggregated information in the context of predictability (see Lutkepohl, 2006, and Hendry and Hubrich, 2011, for theoretical treatments on the conditions that determine the benefits). A key message in these papers is that because of the complexity of the conditions, the issue can only be settled by empirical investigation. In the current context we highlight the use of disaggregated conditional covariances to forecast portfolio returns. This finding is consistent with other empirical studies in other settings. There is a growing number of papers that show that empirically, (aggregate) economic variables such as GDP and inflation can be more accurately predicted using disaggregated predictors (see, e.g., Hernandez-Murillo and Owyang, 2006, and Owyang et al. 2015, Aparicio and Bertolotto, 2020, and Joseph et al., 2022).[3] A key issue within these studies is that the number of disaggregated

[2]The ICAPM does allow that price of risk to vary over time. Indeed, there is ample evidence in this regard; see, e.g., Chan and Marsh (2022) who document a positive relationship between returns and beta risk only on earnings announcement days. This variation is considered in the subsequent analysis via use of various sub-periods in which the models are estimated.

[3]The benefits of using disaggregated information has now been demonstrated beyond the point forecast setting; see, e.g., Paulsen et al.(2022), for evidence in the context of density forecasts of economic aggregates.

predictors can be extremely large. This, in turn, has led to the use of machine learning, shrinkage, clustering, and/or factor models to reduce the curse of dimensionality (see, e.g., Joseph et al., 2022, for the use of machine learning techniques). To avoid the use of complex techniques we consider a decomposition of a large portfolio of returns into a parsimonious set of smaller widely used portfolios (and hence disaggregated predictors). By adopting this kitchen sink approach, we deliberately adopt a simple approach and are still able to demonstrate the out-of-sample virtues of disaggregate predictors. Moreover, we avoid the criticism that the results are reliant on techniques that were not available when the forecasts were generated (we use data covering the period from 1964), or are data snooping induced using a range of model selection techniques (Cremers, 2002).

The unrestricted specification is estimated using a wide range of domestic and international portfolio returns and state variables. For all datasets considered, the results are unequivocal: improved measures of out-of-sample fit, and economic significance are always achieved over competing specifications. For instance, using sixty years of monthly frequency returns to 10 US industry portfolios and market returns as the state variable in the covariance predictor, an out-of-sample R^2 statistic of 2.556% is obtained. This compares to out-of-sample R^2 statistic values of 0.883% and 1.419% when using a common set of predictors (as used in Welch and Goyal, 2008) and the aggregated ICAPM specification in which homogeneous risk prices are assumed, respectively. Similar dominance is observed when the economic significance of the proposed approach is considered, when the components of the portfolio are changed, or when the sample period is changed. The results are consistent with other studies that use disaggregated information to forecast the equity premium. For instance, Lou et al. (2022) show that improved forecasts of quarterly frequency market returns are obtained via decomposed measures of past smoothed daytime and night-time returns. The implications of such findings (and ours) are not in breach of the notion that prices fully reflect all information in a timely fashion. Instead, returns are more predictable because of a richer set of time-varying risk premia than previously considered.

The rest of the paper is organised as follows. The next section lays out the proposed and benchmark forecasting methods, and is followed by an empirical investigation of the out-of-sample quality of these methods. The final section concludes.

2 Methodology

The forecasting methods are described within the context of the following data generating process (DGP).

2.1 The DGP

A panel of data exists consisting of disaggregated dependent components, and a set of scaled predictors (some of which are specific to a particular dependent component, and some that are common to all components). Let r_t be a single time series of interest from a forecasting perspective (henceforth the *variable of interest*), which can be decomposed into n (> 1) individual random variables (henceforth *components*). In the current context the variable of interest will be (excess) returns to the portfolio under investigation (henceforth referred to as the *main portfolio*), and the components will be the (excess) returns to the ith asset (or portfolio) within the main portfolio (henceforth referred to the *component portfolios*).

In the spirit of the ICAPM, we assume that the dynamics of the ith (excess) return (that make up the main portfolio) are a linear function of a set of predictors that are common to each equation within the panel (henceforth the *common predictors*), and the conditional covariances between each return and a set of state variables designed to measure changes in the opportunity set (henceforth the *component-specific predictors*), that is,

$$r_{i,t+h} = c_i + \sum_{j=1}^{m} a_{i,j} x_{j,t} + \sum_{k=1}^{q} b_{i,k} \text{cov}(r_{i,t+h}, s_{k,t+h} | \mathcal{F}_t) + \eta_{i,t+h}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \quad (1)$$

where $a_{i,j}$ is the coefficient associated with the ith asset and the jth common predictor, $b_{i,k}$ is the price of risk associated with the ith asset and kth state variable, $\text{cov}(r_{i,t+h}, s_{k,t+h}|\mathcal{F}_t)$ is the conditional covariance between the ith asset return and kth state variable, and η_{t+h} is a suitably defined zero-mean error term. In contrast to the homogeneous price of risk assumed in the ICAPM, we allow the price of risk (given by $b_{i,k}$) to vary over the assets.

The variable of interest is the equal-weighted return to the main portfolio (henceforth referred to as the *main portfolio return*). The DGP associated with this variable (henceforth the aggregated

DGP) is obtained by taking the mean across $r_{i,t+h}$ to give

$$r_{t+h} = \frac{1}{n}\sum_{i=1}^{n} c_i + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} a_{i,j}x_{j,t} + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{q} b_{i,k}\text{cov}(r_{i,t+h}, s_{k,t+h}|\mathcal{F}_t) + \frac{1}{n}\sum_{i=1}^{n}\eta_{i,t+h}. \quad (2)$$

The conditional covariance associated with the first asset in the main portfolio can be written as

$$\text{cov}(r_{1,t+h}, s_{k,t+h}|\mathcal{F}_t) = n\,\text{cov}(r_{t+h}, s_{k,t+h}|\mathcal{F}_t) - \sum_{i=2}^{n}\text{cov}(r_{i,t+h}, s_{k,t+h}|\mathcal{F}_t), \quad (3)$$

where $\text{cov}(r_{t+h}, s_{k,t+h}|\mathcal{F}_t)$ is the conditional covariance between the main portfolio return and the kth state variable (henceforth referred to as *aggregated component-specific predictors*).

Substituting (3) into (2) we obtain

$$\begin{aligned} r_{t+h} = {} & \frac{1}{n}\sum_{i=1}^{n} c_i + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} a_{i,j}x_{j,t} + \sum_{k=1}^{q} b_{1,k}\text{cov}(r_{t+h}, s_{k,t+h}|\mathcal{F}_t) \\ & + \frac{1}{n}\sum_{i=2}^{n}\sum_{k=1}^{q}(b_{i,k} - b_{1,k})\text{cov}(r_{i,t+h}, s_{k,t+h}|\mathcal{F}_t) + \frac{1}{n}\sum_{i=1}^{n}\eta_{i,t+h}. \quad (4) \end{aligned}$$

By expressing the DGP in this way we are able to distinguish between a predictor equation implied by the ICAPM and the proposed predictor equation. Under the assumptions of the ICAPM, the price of risk associated with the kth state variable is the same for all assets in the main portfolio. This amounts to assuming that $b_{1,k}$ equals $b_{i,k}$ such that all coefficients on the component-specific conditional covariances equal zero. Our conjecture is that these restrictions lead to a deterioration in out-of-sample performance.

2.2 Forecasting methods and models

The objective of all methods is to generate h-step ahead forecasts of the variable of interest (that is, the main portfolio return r_{t+h}). Three models are considered: a highly restricted version of the ICAPM (in which the price of risk is set to zero); an unrestricted version of the ICAPM (in which a homogeneous price of risk is assumed); and a generalised version of the ICAPM (in which heterogeneous prices of risk are permitted).

The first method that we consider uses the common predictors only and generates forecasts that are the fitted values from a model based on a restricted version of the aggregated DGP. Specifically, the following model is estimated via ordinary least squares (OLS):

$$r_{t+h} = c + \sum_{j=1}^{m} a_j x_{j,t} + \epsilon_{t+h}, \tag{5a}$$

where c and a_j are free parameters. This method of generating forecasts is based on a restricted version of the ICAPM and is henceforth referred to as method $\mathcal{A}$. Here, $p_{\mathcal{A}} = m + 1$ parameters are estimated using T observations. In the absence of common predictors this method coincides with a naive method in which only the historical mean is used as the predictor.

The second method that we consider uses the common predictors and the aggregated component-specific predictors (thus no knowledge of the disaggregated information) and generates forecasts that are the fitted values from a model based on a restricted version of the aggregated DGP, with parameters estimated by OLS. Specifically,

$$r_{t+h} = c + \sum_{j=1}^{m} a_j x_{j,t} + \sum_{k=1}^{q} b_k^a \text{cov}(r_{t+h}, s_{k,t+h}|\mathcal{F}_t) + \epsilon_{t+h}, \tag{5b}$$

where c, a_j and b_k^a are free parameters. This method of generating forecasts is based on an unrestricted version of the ICAPM and is henceforth referred to as method $\mathcal{B}$. Here, $p_{\mathcal{B}} = m + q + 1$ parameters are estimated using T observations.

The third method assumes full knowledge of the disaggregated predictors, and generates forecasts that are the fitted values from a unrestricted version of a model based on the aggregated DGP. Specifically, the following model is estimated via OLS:

$$r_{t+h} = c + \sum_{j=1}^{m} a_j x_{j,t} + \sum_{k=1}^{q} b_j^a \text{cov}(r_{t+h}, s_{k,t+h}|\mathcal{F}_t) + \sum_{i=2}^{n} \sum_{k=1}^{q} b_{i,k}^d \text{cov}(r_{i,t+h}, s_{k,t+h}|\mathcal{F}_t) + \epsilon_{t+h}, \tag{5c}$$

where c, a_j, b_k^a and $b_{i,j}^d$ are free parameters. This method of generating forecasts is henceforth referred to as method $\mathcal{C}$. It is based on a generalised version of the ICAPM in which $p_{\mathcal{C}} = m + nq + 1$ parameters are estimated using T observations.

The models, on which the above methods are based, are nested such that the error variances monotonically decrease as we move from method $\mathcal{A}$ to $\mathcal{C}$. However, the number of parameters in these models monotonically increase over this space. It follows that the out-of-sample performance of the methods will depend on how these effects trade-off. This is examined in the next section.

2.3 Out-of-sample performance measures

It is a well-established fact that the true quality of a forecasting method can only be assessed by examining its out-of-sample performance. To this end, we consider a range of out-of-sample performance measures that are both statistical and economic in nature.

2.3.1 Traditional performance measures

The forecasts generated by each method are evaluated by considering functions of the *conditional* mean squared forecast error (MSFE). In all cases a sample of size T is assumed. For the kth method above, under the OLS and normal error assumptions, the conditional MSFE is given by

$$v_k \equiv \mathrm{E}(\epsilon^2_{T+h}|\mathcal{F}_{k,T}) = \frac{V_k}{T} \times \frac{(T+\theta_1)(\theta_{k,2}+\theta_{k,3})}{\theta_{k,3}}, \quad (6a)$$

where

$$\theta_1 \sim \chi^2_1, \quad \theta_{k,2} \sim \chi^2_{p_k-1}, \quad \theta_{k,3} \sim \chi^2_{T-p_k+1}. \quad (6b)$$

Here V_k is the error variance associated with the kth method, and $\mathcal{F}_{k,T}$ is the T-size data sample and model employed by users of the kth method. The expressions are based on a standard result in the literature; see Leeb (2009).

The moments associated with the conditional MSFE are relatively simple to calculate: the ith unconditional moment about zero for the kth method is given by

$$\mathrm{E}(v_k^j) = g(j, p_k, T)V_k^j, \quad (7a)$$

where

$$g(j,p_k,T) = \left(\frac{1}{2T}\right)^j \frac{\Gamma((T-p_k-2j+1)/2)}{\Gamma((T-p_k+1)/2)} \prod_{i=1}^{j}(T-2i) \sum_{i=0}^{j} 2^i T^{j-i} \binom{j}{i} \frac{\Gamma(i+1/2)}{\Gamma(1/2)}. \quad (7b)$$

Here $\Gamma(.)$ is the Euler gamma function, and $\binom{j}{i}$ is the binomial coefficient. For instance, the *unconditional* MSFE is given by the following first moment:

$$\mathrm{E}(v_k) = g(1, p_k, T)V_k, \tag{8a}$$

where

$$g(1, p_k, T) = \frac{T-2}{T-1-p_k}\left(\frac{T+1}{T}\right). \tag{8b}$$

As we move from method $\mathcal{A}$ to $\mathcal{C}$, $g(1, p_k, T)$ increases (less parsimony), while V_k decreases (better fit). Thus we have the usual trade-off in order to yield to the best out-of-sample performance in terms of the unconditional MSFE.

2.3.2 Distributional performance measures

It is possible to go further and derive an expression for the unconditional distribution of the conditional MSFE. To this end we first note from (6a) that the conditional MSFE consists of the product of $T + \theta_1$ and $V_k(\theta_{k,2} + \theta_{k,3})/\theta_{k,3}T$; enabling us to apply standard techniques to yield the distribution of the product of two independent random variables. Given the (chi-squared) distributions of θ_1, $\theta_{k,2}$ and $\theta_{k,3}$, the cumulative density function for the conditional MSFE can be simplified to

$$F_{v_k}(v) = \begin{cases} \mathrm{erf}\left(\sqrt{\frac{(v-V_k)T}{2V_k}}\right), & \text{if } p_k = 1, \\ \frac{\Gamma(T/2)}{\Gamma((p_k-1)/2)\Gamma((T-p_k+1)/2)} \int_1^{v/V_k} u^{-T/2}(u-1)^{(p_k-3)/2}\mathrm{erf}\left(\sqrt{\frac{(v-uV_k)T}{2uV_k}}\right) du, & \text{if } p_k > 1, \end{cases} \tag{9}$$

where erf(.) is the Gauss error function. In both cases, the functions are defined over the domain $v > V_k$. Use of this function extends the notion of out-of-sample performance beyond the usual approach.

The traditional approach to measuring forecasting performance is based on the first moment of the conditional MSFE, that is, the unconditional MSFE. A set of forecasts that yield the smallest unconditional MSFE is deemed to be the best. This approach does, however, ignore the variation of MSFE values (*performance risk*) due to the particular sample of data used. To incorporate this element of performance we consider the best set of forecasts to be those that coincide with the least chance of obtaining a poor (that is, high) MSFE value. Given this risk metric, we introduce the following *performance-at-risk* (PaR) measure analogous to the value-at-risk (VaR) performance measure:

$$\mathrm{PaR}_\alpha(v_k) = \inf\{v \in \mathbb{R} : \Pr(v_k \geq v) < \alpha\} = F_{v_k}^{-1}(1-\alpha). \tag{10}$$

where $F_{v_k}^{-1}(1-\alpha)$ is the $(1-\alpha)$-quantile of v_k. Typical values of α would be 0.01 (1% PaR measure), and 0.05 (5% PaR measure). The cumulative density in (9) can be used to calculate these PaR measures.

2.3.3 An economic significance measure

In addition, a measure of out-of-sample economic value is possible. The (out-of-sample) performance fee associated with a dynamic asset allocation strategy based on the kth method (relative to a static strategy in which the mean and variance of returns are known) can be shown to be given by

$$\delta_k = -\frac{\Psi_k \mathrm{SR}^2 + \ln(\Psi_k / R_k^4)}{2\gamma}, \tag{11a}$$

where

$$\Psi_k = \frac{R_k^4}{2 + g(1, p_k, T)(R_k^2 - 1) + R_k^4}. \tag{11b}$$

Here γ is the Arrow-Pratt relative risk aversion parameter associated with an exponential utility function, SR is the Sharpe ratio associated with the buy-and-hold strategy, and R_k^2 is the in-sample goodness-of-fit associated with the kth method. See the Appendix for a full description of the assumptions and subsequent derivation of this measure.

2.3.4 Efficiency

To isolate the benefits of risk price heterogeneity (that is, using component-specific predictors), we calculate the efficiency of method j with respect to method $\mathcal{A}$ (with no common predictor included). Efficiency is defined as

$$\mathrm{eff}_j = 1 - \frac{M_j}{M_{\mathcal{A}}}, \tag{12}$$

where M_j represents a measure of forecasting (non-)performance. If we use the unconditional MSFE, then as the sample size grows (that is, $T \to \infty$), this particular efficiency approaches the in-sample goodness of fit (R^2) associated with method j. In this sense, for finite T values, the efficiency represents the out-of-sample goodness of fit (OOS-R^2). In a similar vein, the MSFE value can be replaced by the PaR value, to produce the relative efficiency of performance risk (denoted ∇PaR).

3 Empirical results

The dataset used, and the results associated with application of forecasting methods $\mathcal{A}$ to $\mathcal{C}$ to these data are described in this section.

3.1 Data

The data used consists of three separate parts: component portfolio returns, state variables and common predictors. For all parts, monthly frequency data are used.

In terms of US data, we make use of industry portfolios and portfolios sorted on firm characteristics. Specifically, we consider the 10-industry portfolio, and 25-component portfolios based on various bivariate sorts on size, book-to-market value, operating profitability, investment, momentum, short-term reversal, long-term reversal, accruals, market beta, net share issues, return variance (total), and return variance (residual).[4] This yields 14 different main portfolios (each with 25-component portfolios).[5] In addition, we consider 6-component portfolios based on various bivariate sorts on size, book-to-market value, operating profit, investment, momentum, short-term reversal, long-term reversal, earnings, cashflow, and dividend yield. This yields nine different main portfolios (each with 6-component portfolios).[6] Within each component portfolio we use value-weighted returns (inclusive of dividends) over the period January 1964 to December 2022. These data were downloaded from the Kenneth French data library at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[4]The industries in the 10-industry portfolio are consumer non-durables, consumer durables, manufacturing, energy, hi-tech, telecoms, shops, health, utilities and other (mines, construction, building materials, transport, hotels, business services, entertainment, and finance).

[5]The 14 bivariate sorts are: size/book-to-market value, size/operating profitability, size/investment, size/momentum, size/short-term reversal, size/long-term reversal, size/accruals, size/market beta, size/net share issues, size/return variance (total), size/return variance (residual), book-to-market value/investment, book-to-market value/operating profitability, and operating profitability/investment.

[6]The nine bivariate sorts are: size/book-to-market value, size/operating profitability, size/investment, size/momentum, size/short-term reversal, size/long-term reversal, size/cash flow, size/earnings, and size/dividend yield.

For international region data we consider the following markets: Asia-Pacific (excluding Japan), developed, developed (excluding the US), Europe, Japan, and North American. In each case we consider 6-component and 25-component portfolios based on various bivariate sorts on size, book-to-market value, operating profit, investment and momentum. This yields four main portfolios per region (each with 6- and 25-component portfolios).[7] Within each component portfolio we use value-weighted returns (inclusive of dividends) over the period January 1991 to December 2022. These data were downloaded from the Kenneth French data library at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

The state variables in the conditional covariance predictors consist of the Fama-French five factors plus the momentum factor; see Bali and Engle (2010) for use of these factors within the context of the ICAPM. That is, returns to the market (MK), small-minus-big (SMB), high-minus-low (HML), robust-minus-weak (RMW), conservative-minus-aggressive (CMA), and momentum (MOM) factors. These factors are collected for the US, and each of the international regions previously described. In all cases, we use value-weighted returns (inclusive of dividends) over the period January 1964 to December 2022 (US data only) or January 1991 to December 2022 (international regions). These data were downloaded from the Kenneth French data library at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Four common predictors are employed (see, e.g., Brandt and Santa-Clara, 2006, for use of these 'classic' predictors): the short term interest rate (given by the 3-month Treasury bill rate) denoted TB; the term spread (the difference between the long term yield on government bonds and the short term interest rate) denoted TMS; the default yield (the difference between BAA- and AAA-rated corporate bond yields) denoted DFY; and the dividend price ratio (the difference between the log of 12-month sum of dividends paid to stocks in the S&P500 index and the log of the S&P500 index) denoted DP. These data are collected over the period from January 1964 to December 2021 and were downloaded from Amit Goyal's website at `https://sites.google.com/view/agoyal145/?redirpath=/`.

The objective is to generate accurate forecasts of main portfolio returns, that is, forecasts of the equal-weighted average of the returns to the component portfolios. As such, these returns represents

[7] The four bivariate sorts are: size/book-to-market value, size/operating profitability, size/investment, and size/momentum.

a measure of market returns. To see this we calculate the (contemporaneous) correlation between each main portfolio return and various measures of market returns. Several US portfolios are considered: eight different industry portfolios, 14 different 25-component bivariate-sort portfolios, and nine different 6-component bivariate-sort portfolios. Market returns are given by returns to the value-weighted CRSP market portfolio index (CRSP-VW), the equal-weighted CRSP market portfolio index (CRSP-EW), the S&P 500 index, and a S&P 500 ETF (that is, the SPDR ETF) – all of which were collected from the CRSP database (accessed via WRDS). The correlations are provided in Table 1 and are based on monthly frequency data observed over the period January 1994 to December 2022.

Insert Table 1 here

The correlations are (unsurprisingly) close to unity. For instance, the correlation between the return to the aggregate 25-component portfolio based on the size/book-to-market sort and the CSRP-VW return is 0.945. This correlation falls when considering the less broad S&P 500 index and S&P 500 ETF returns, though remains above 0.9. Similar results hold for the other portfolios. Thus, our variables of interest (that is, main portfolio returns) represent a reasonable measure of the market return, and thus the subsequent results can be viewed within the wider context of the equity premium predictability literature.

3.2 In-sample results

We begin with an examination of the in-sample fit of the models associated with methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, with a one-month forecasting horizon. The main portfolio consists of the 10-industry component portfolio. The dependent variable is the main portfolio return, which in turn is the equal-weighted average of the returns to each of the 10 industry returns. It follows that the dependent variable is a (pseudo) market return. The component predictors are the conditional covariances between the component portfolio returns and the state variable. Conditional covariances are constructed using the 60-month moving average (MA) of the product of the component portfolio return and the state variable return and using the exponentially weighted moving average (EWMA) method with a decay factor of 0.97 (as recommended by RiskMetrics). The state variable is the MK return. Using this state variable means that the aggregated covariance between the component portfolio returns and the MK return will be approximately equal to the conditional variance of MK returns. The common predictors are the TB, TMS, DFY and DP variables. Restricted, unrestricted, and generalised versions of the ICAPM as given by (5a), (5b) and (5c), respectively, are estimated by OLS. Results are provided in Table 2.

Insert Table 2 here

The standard prediction model based on the TB, TMS, DFY and DP predictors achieves an adjusted R^2 of 0.884%. The fit of the unrestricted ICAPM depends on the method used to construct the conditional variances: the MA method delivers an adjusted R^2 of 0.870%, while the EWMA method delivers an adjusted R^2 of 1.420%. When all the component-specific conditional covariances are included in the model the adjusted R^2 value reaches 2.569%. More importantly, we test whether the coefficients on the component-specific conditional covariances jointly equal zero via a likelihood ratio test. The p-values equal 0.015 and 0.044 when using the MA and EWMA conditional covariance construction methods. This lends support to the generalised ICAPM (over the unrestricted ICAPM) in which component-specific conditional covariances (disaggregated information) play a role in the prediction of market returns.

3.3 Out-of-sample results (US data)

To assess the 1-step ahead out-of-sample performance of methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, we estimate the models using the 10-industry component portfolio. The true error variance (V_k) is replaced by its sample counterpart and the formulae laid out in section 2.3 are applied. This calibrated approach is adopted in all subsequent analysis.

Method $\mathcal{A}$ is based on a constant and the classic predictors, method $\mathcal{B}$ is based on a constant, the classic predictors, and the conditional covariance of main portfolio returns with state variable returns, method $\mathcal{C}$ is based on a constant, the classic predictors, the conditional covariance of main portfolio returns with state variable returns and the conditional covariances of component portfolio returns with the state variable. We use MK returns as the state variable, and the MA and EWMA conditional covariance construction methods. The models are also estimated without the classic predictors. The R^2, OOS-R^2 and 5% ∇PaR values are provided in Table 3. To examine the impact of sample size on out-of-sample performance, the OOS-R^2 statistics are calculated for the following sample sizes: $T = \{120, 240, 360, 480\}$, and the sample size that corresponds to the actual number of observations in the full sample (which is, $T = 708$). To avoid the use of model selection techniques that could bias the results, method $\mathcal{C}$ assumes that all covariances terms are included in the associated model. In doing this we hope to avoid the data snooping critique.

Insert Table 3 here

The results indicate that if the sample size is sufficiently high (that is, $T \geq 360$) then method $\mathcal{C}$ is the dominant forecasting strategy. For instance, this method delivers an OOS-R^2 statistic of 2.566% when common predictors are included, the EWMA method is used to construct the covariances, and the maximum number of observations are used. By contrast, methods $\mathcal{A}$ and $\mathcal{B}$ have OOS-R^2 statistics of 0.883 and 1.419, respectively. Regarding the ∇PaR values, the respective results are 0.476%, 0.915%, and 1.688%, for methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. Collectively, the results are supportive of forecasting methods that make use of the disaggregated information in the form of the component conditional covariances (and thus the generalised ICAPM in which heterogeneous risk prices are present). The caveat is that the sample size must be sufficiently large.

Turning to economic significance, we calculate the performance fees where the benchmark portfolio is either one in which wealth is equally split between the risk-free and risky asset, or one in which all wealth is allocated to the risky asset (that is, the buy-and-hold portfolio benchmark). The data and estimation assumptions are the same as those used above. Results are provided in Table 4.

Insert Table 4 here

It is clear from the results that method $\mathcal{C}$ remains the dominant forecasting strategy. For instance, this method delivers an annualised performance fee of 4.462% when common predictors are included, the EWMA method is used to construct the covariances, the maximum number of observations are used, and the buy-and-hold benchmark strategy is adopted. This compares to fees of 1.306% and 2.260% when methods $\mathcal{A}$ and $\mathcal{B}$ are considered, respectively. Again, sample size plays an important role in ensuring that this conclusion holds.

3.4 Out-of-sample results (international data)

The results thus far pertain to only one main portfolio (the 10-industry component portfolio), one state variable (that is, MK returns), and one sample period with which the error variances are estimated. We generalise the out-of-sample performance results in three ways. First, we consider the forecasting methods in the context of 6- and 25-component US main portfolios based on various bivariate sorts of firm-specific information (such as size). This yields nine different 6-component main portfolios and 14 different 25-component main portfolios. Second, we consider the following international region data: Asia-Pacific (excluding Japan), developed, developed (excluding the US), Europe, Japan, and North America. In each case, we consider 6-component and 25-component portfolios based on various bivariate sorts. This yields four main portfolios per region (each with 6- and 25-component portfolios). Finally, to allow for time variation in prediction equation parameters, we consider all sub-samples of the (maximum) sample period (with different start and end points), subject to the restriction that the sample size is equal to 60 observations or greater, and that the start of the sample is January, and the end of the sample is December.[8] In doing this, we consider 1540 sub-samples (when the maximum sample period is 1964 to 2022) and 406 sub-samples (when the maximum sample period is 1991 to 2022).

In terms of state variables, we consider the Fama-French five factors plus the momentum factor. These are combined in various ways to yield a variety of conditional covariances. We consider MK returns (denoted FF1); MK, SMB and HML returns (denoted FF3); MK, SMB, HML and MOM returns (denoted FF4); and MK, SMB, HML, RMW and CMA returns (denoted FF5). The use of these groups corresponds to the use of groups of conditional covariances (and hence a larger number of predictors). For instance, use of a 25-component main portfolio and the FF3 factors will involve 75 separate conditional covariance predictors. This is clearly over- parametrised. To overcome this issue, we invoke the following assumption. When using method $\mathcal{B}$ we use the conditional covariances associated the main portfolio return and each of the above state variable combinations. By contrast, for method $\mathcal{C}$, we limit the conditional covariances such that we only use the MK return state variable (hence we only consider the conditional covariances between component portfolio returns and MK returns).

[8] See Kolev and Karapandza (2017) for use of sub-sample analysis in the context of out-of-sample equity premium prediction.

The use of such a variety of data means that there is a huge expansion in the number of different method comparisons. For instance, when using US data over the period 1964 to 2022, and 25-component main portfolios, we make 21,560 (14×1540) method performance comparisons. Moreover, as we consider four different state variable combinations, this amounts to 86,240 comparisons for the US data alone. To summarise this information, we present the number of successes per dataset and state variable grouping. In particular, the results in Tables 5 to 8 provide the percentage of times each method is superior to all other methods within each region and state variable combination. A success (that is, when a method performs better than all other methods) is registered for each main portfolio, and each sub-sample, and then aggregated over all these dimensions. Performance is defined in terms of the OOS-R^2 and 5% PaR values, and two different methods of constructing the conditional covariances (that is, the MA and EWMA methods). Tables 5 and 6 contain the aggregated success rates for the 6-component main portfolios, for all regions, all state variable combinations, and the MA and EWMA conditional covariance methods, while Tables 7 and 8 contain the aggregated success rates for the 25-component main portfolios, for all regions, all state variable combinations, and the MA and EWMA conditional covariance methods. The international data is only available over the period 1991 to 2022, while the US data is available over the period 1964 to 2022. For comparison purposes we also consider US data over the period 1991 to 2022.

Insert Tables 5, 6, 7 and 8 here

The results highlight the superiority of method $\mathcal{C}$ over the other methods. For instance, when using US data over the period 1991 to 2022, 25-component main portfolios, the EWMA covariance method, the FF5 state variable grouping, and the OOS-R^2 performance criterion, method $\mathcal{C}$ achieves a success rate of 96.622% (the corresponding success rates for methods $\mathcal{A}$ and $\mathcal{B}$ are 0.616% and 2.762%). Similar findings are observed over all the non-US datasets. For instance, the corresponding success rates when using European data for methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are 0.308%, 5.788% and 93.904%, respectively. Thus, use of component conditional covariances is a sensible general approach if one seeks improved out-of-sample forecasts of market returns.

While method $\mathcal{C}$ is, in general, the dominant method, the results do reveal interesting variations in the degree of dominance over three dimensions. First, when considering the PaR performance criteria there is a drop in the method $\mathcal{C}$ success rates. For instance, when using US data over the period 1991 to 2022, 25-component main portfolios, the EWMA covariance method, the FF5 state variable combination, and the PaR performance criterion, the success rates for methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are 6.175%, 7.477% and 86.348%, respectively. Thus, distributional aspects of the performance of the methods suggest that method $\mathcal{C}$ can occasionally deliver poor performance (albeit still to a lesser extent than the other methods). Second, the number of components in the main portfolio is a determinant of performance. Specifically, when using the 6-component main portfolio there is a drop in the dominance of method $\mathcal{C}$. For instance, when using US data over the period 1991 to 2022, 6-component main portfolios, the EWMA covariance method, the FF5 state variable combination, and the PaR performance criterion, the success rates for methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are 24.001%, 25.780% and 50.219%, respectively. Third, replacing the EWMA with the MA covariance method, we see that these values become 40.394%, 30.679%, and 28.927%, respectively. Thus, the results indicate that the choice of the main portfolio size and the conditional covariance construction method are important considerations when using method $\mathcal{C}$ to forecast market returns.

3.5 Predicting global stock returns

The variable of interest in the above analysis is the main portfolio (market) return within each international region. By contrast, we now consider the case where one seeks a forecast of global portfolio returns using a selection of international portfolios. To this end, the main portfolio consists of the Asia-Pacific (excluding Japan), Japan, Europe, North America, US, and developed (excluding US) market portfolios, and is thus a 6-component portfolio. The global stock return is defined as the equal-weighted average of these market portfolio returns. The MK return is used as the state variable.

A graphical representation of the results is provided in Figure 1. We plot the OOS-R^2 and ∇PaR efficiency values of method $\mathcal{C}$ against those associated with method $\mathcal{B}$ (efficiencies are relative to method $\mathcal{A}$). Each point in the plots represents a result from each sub-sample. Both conditional covariance construction methods are considered. Points above the 45-degree line indicate cases where method $\mathcal{C}$ is superior to method $\mathcal{B}$, and above the x-axis indicate cases where method $\mathcal{C}$ is superior to method $\mathcal{A}$. The proportion of points above both lines represents the success rate of method $\mathcal{C}$. Use of the plots also provides an indication of the magnitudes of the efficiencies.

Insert Figure 1 here

The results confirm the dominance of method $\mathcal{C}$, with efficiencies up to 10% common. Moreover, when method $\mathcal{C}$ fails to be dominant, the corresponding inefficiency is close to zero. The plots also indicate that the dominance of method $\mathcal{C}$ increases when using the EWMA method to construct the conditional covariances and diminishes when using the PaR efficiency criteria.

The results thus far have focused entirely on the relative quality of one-step ahead monthly forecasts. To examine longer horizons, we present results for 3-month and 12-month forecasting horizons. Results based on the 6-component international market portfolio and the EWMA conditional covariance construction methods are provided in Figure 2.

Insert Figure 2 here

The results indicate that large OOS-R^2 method $\mathcal{C}$ efficiencies are possible when the 3-month forecasting horizon is considered. These efficiencies fall when using the PaR criteria, though method $\mathcal{C}$ remains dominant. Somewhat unsurprisingly the dominance of method $\mathcal{C}$ falls over the longer 12-month forecasting horizon. Indeed, when using the PaR criterion over this horizon, method $\mathcal{A}$ appears to dominate. Thus, the dominance of method $\mathcal{C}$ appears to be confined to forecasting horizons of up to three months.

3.6 Method refinement

Up to this point we have assumed that the methods have either no conditional covariance predictors (method $\mathcal{A}$), one conditional covariance predictor (method $\mathcal{B}$), or all available conditional covariance predictors (method $\mathcal{C}$). As such, no attempt has been made to differentiate amongst the quality of the conditional covariance predictors. To address this issue, we perform cluster analysis to group together similar conditional covariance predictors, while maintaining a degree of heterogeneity across the groups. By grouping in this way, one is able to aggregate the conditional covariances within each group such that a smaller number of conditional covariances are used. The ambition is that this will reduce the number of parameters required, but will maintain a high level of fit.

The k-means clustering technique is used to group together the conditional covariances with similar attributes. The k-means algorithm is applied as follows. The full model is estimated by

OLS using all component conditional covariances for each state variable. The estimated coefficients represent the attribute (that is, the observation) that is grouped in the algorithm. Then the following steps are taken. First, in the assignment step, each observation is assigned to the cluster with the nearest mean (where distance is measured by the squared Euclidean distance). Second, in the update step, the mean of each cluster is recalculated. This process is repeated until convergence occurs. The sum of the conditional covariances within each group are then constructed and form part of the new predictor equation, which is estimated by OLS. Henceforth we refer to the forecasting method based on clustering as *method $\mathcal{C}$ (with clustering)*.

To demonstrate the virtues of clustering we use 14 different US main portfolios each consisting of 25-component portfolios (as described in the data section), all available sub-samples (taken from the January 1964 to December 2022 sample), and the MK return state variable. Component conditional covariances are constructed using the MA and EWMA methods, and out-of-sample efficiencies are measured via OOS-R^2 and ∇PaR. Box-and-whisker plots of the out-of-sample efficiencies against the number of clusters are provided in Figure 3. At the two extremes we have method $\mathcal{B}$ (one cluster) and method $\mathcal{C}$ (25 'clusters', where each cluster consists of just one conditional covariance). Between these we have the results associated with method $\mathcal{C}$ (with clustering).

Insert Figure 3 here

Each box contains the out-of-sample efficiencies with respect to method $\mathcal{A}$ based on 1540 different sub-samples, and each of the 14 main portfolios. Thus, for each box we have 21,560 ($= 1540 \times 14$) out-of-sample efficiency measures. As we consider 25 different cluster numbers, each plot in Figure 3 represents 539,000 ($= 21,560 \times 25$) efficiencies. Moreover, as there are four plots, the figure depicts over two million efficiencies in total.

The results support the use of clustered conditional covariances. Consistent with results in the previous tables, the EWMA method of constructing the conditional covariances is superior to the MA method, and the OOS-R^2 results are slightly more supportive of the proposed methods. In terms of clustered predictors, the results show a clear improvement over methods $\mathcal{B}$ and $\mathcal{C}$, with a sweet spot occurring when around 12 to 16 clusters are used. For instance, using the EWMA conditional covariances, methods $\mathcal{B}$ and $\mathcal{C}$ deliver median OOS-R^2 values of around 1.5% and 10%, respectively. By contrast, use of 14 clustered predictors delivers median OOS-R^2 values of around

16%. The whiskers in the plots also show some variation in the spread of efficiencies when using different numbers of clusters. For method $\mathcal{B}$, the spread is very small. This spread increases as the number of clusters increases (indicating variation in efficiencies across the 1540 sub-samples and 14 main portfolios). However, it is noticeable that the highest OOS-R^2 values achieved by method $\mathcal{C}$ are still below the lowest OOS-R^2 values delivered by method $\mathcal{C}$ (with clustering) when using 12 to 16 clustered predictors. Overall, these results give credence to the use of clustered predictors within the context of the proposed forecasting method.

4 Conclusion

The statistical and economic benefits of using component conditional covariances when forecasting market returns are robust to the use of data from different international regions. This novel finding does not rely on use of a sophisticated forecasting method, does not require a complex method to construct conditional covariances, and is not confined to particular markets. Thus, taking advantage of the heterogeneity in risk prices across markets is a worthy pursuit when an out-of-sample prediction of aggregate (market) returns is the objective. It does not violate any notion of market efficiency, as the predictors are (component-specific) proxies for the time-varying variation in risk premia – a finding that is fully consistent with heterogeneous risk prices over segmented markets (and a generalised version of the ICAPM).

The approach taken in this paper has been fairly simple, and hence is easy to implement by users. Moreover, the benchmark methods have been allowed to use a richer set of variables (for instance, method $\mathcal{B}$ permits use of a richer set of state variables, while method $\mathcal{C}$ is restricted to the MK return state variable). Notwithstanding this bias towards to competing methods, the proposed method is successful and robust to variation in the design of the experiment.

The size (that is, the number of component assets or portfolios) of the main portfolio is an important consideration as this determines the number of predictors in the proposed method. The virtues of dimension reduction via clustering have been demonstrated on a moderately sized portfolio. In future research one could consider a very large main portfolio, perhaps consisting individual stocks. This would then amount to a big data problem, in which the number of predictors is potentially greater than the number of (time-series) observations. In this context, one could use factor models such that one achieves a reduction in the number of component conditional covariances used in the predictor equation.

Appendix A Measuring economic significance

Economic significance is measured by the performance fee an investor is willing to pay to switch from a static buy-and-hold trading strategy to a dynamic trading strategy. The assumptions and derivation of this measure are laid out below.

Assumption 1 (Return dynamics)*.* The returns to the risky asset (or portfolio) under management are given by

$$r_t = \mathrm{E}(r_t|\mathcal{F}_{k,t-1}) + \epsilon_t, \tag{A.1}$$

where $\mathrm{E}(r_t|\mathcal{F}_{k,t-1}) \sim N(\mu, \sigma_e^2)$ is the conditional expectation of 1-step ahead returns held by the kth investor, and $\epsilon_t \sim N(0, \sigma^2 - \sigma_e^2)$ is an error term. It follows that $r_t \sim N(\mu, \sigma^2)$.

Remark. Here predictability is controlled by the magnitude of σ_e^2. Moreover, in-sample predictability is given by the R^2 statistic. Specifically, $R^2 = 1 - (\sigma^2 - \sigma_e^2)/\sigma^2$.

Assumption 2 (Trading strategies returns)*.* The kth trading strategy involves taking positions in a risk-free asset (earning r_f) and a risky asset (earning r_t) to maximise the conditional expectation of next period utility. The returns to the kth trading strategy portfolio (based on the kth information set) are given by

$$r_{k,t} = 1 + r_f + w_{k,t}(r_t - r_f). \tag{A.2}$$

where $w_{k,t}$ is the portfolio weight used by the kth strategy at time t based on information at time $t-1$.

Two trading strategies are considered. The assumptions underlying these strategies are given next.

Assumption 3 (The static trading strategy)*.* The static trading strategy adopts a buy-and-hold strategy and has full knowledge of the true mean and variance of the risky asset returns. This amounts to assuming that conditional mean of risky asset returns is given by μ and the conditional variance of risky asset returns is given by σ^2.

Assumption 4 (The dynamic trading strategy)*.* The dynamic strategy takes a time-varying position in the asset to maximise the conditional expectation of next period utility. Moreover, the dynamic strategy considers the parameter uncertainty associated with use of a sample size T. This amounts to assuming that the conditional mean of risky asset returns is given by $\mathrm{E}(r_t|\mathcal{F}_{k,t-1})$ and the conditional variance of risky asset returns is given by $g(1,p_k,T)(\sigma^2-\sigma_e^2)$, where $g(1,p_k,T)$ is the adjustment factor to reflect use of a method that employs a model with p_k parameters and a sample of size T.

Assumption 5 (Utility). The benefits of consuming the portfolio return to the kth strategy are measured via the negative exponential utility function. Specifically, utility at time t is given by

$$u(r_{k,t},\gamma)=-\exp(-\gamma r_{k,t}),\quad t=1,\ldots,T, \tag{A.3}$$

where γ is the Arrow-Pratt relative risk aversion parameter.

We are now ready to define the measure of the benefits of using a dynamic strategy with respect to a static strategy. Specifically, the formal definition of economic value in this regard is given in the following definition.

Definition (The performance fee set)**.** The performance fee set contains the fees the user of the dynamic strategy is willing to pay to enjoy the unconditional expected utility over that enjoyed by the static strategy. Specifically, this set is given by $\{\delta\in\mathbb{R}|\Phi[\delta]\}$ with

$$\Phi[\delta]=\mathrm{E}(u(r_{1,t}-\delta,\gamma)-u(r_{0,t},\gamma))\geq 0. \tag{A.4}$$

Here $r_{1,t}$ is the dynamic trading strategy return, and $r_{0,t}$ is the static trading strategy return.

Proposition 1 (The maximum performance fee)**.** *Under Assumptions 1 to 5, the maximum performance fee an investor is willing to pay to maintain use of the dynamic trading strategy (instead of the static trading strategy) is given by*

$$\delta=-\frac{\Psi SR^2+\ln(\Psi/R^4)}{2\gamma},$$

where

$$\Psi=\frac{R^4}{2+g(1,p,T)(R^2-1)+R^4},$$

where SR is the Sharpe ratio associated with the buy-and-hold strategy, and R^2 is the in-sample goodness-of-fit associated with the dynamic trading strategy.

Proof. The portfolio weights to each strategy are obtained by differentiating the conditional expectation of next period utility, setting to zero, and solving for the weight. Doing this under the distribution assumptions associated with (A.1), we obtain the following expressions for the weights to the static and dynamic strategies:

$$w_0 = \frac{\mu - r_f}{\gamma \sigma^2}, \quad w_{1,t} = \frac{\mathrm{E}(r_t|\mathcal{F}_{k,t-1}) - r_f}{\gamma g(1,p,T)(\sigma^2 - \sigma_e^2)}. \tag{A.5}$$

Substituting these expressions into (A.2) we obtain expressions for the returns to each strategy portfolio, which are then substituted into the utility function expression in (A.3). The final step is to calculate the unconditional expectation of the performance fee set defined in (A.4) and solve for the performance fee. □

Remark. Using the expression in (A.5), and the distribution assumptions in (A.1), it follows that the weights associated with the dynamic strategy are normally distributed with mean and variance given by

$$\mathrm{E}(w_{1,t}) = \frac{\mu - r_f}{\gamma g(1,p,T)(\sigma^2 - \sigma_e^2)}, \quad \mathrm{var}(w_{1,t}) = \frac{\sigma_e^2}{(\gamma g(1,p,T)(\sigma^2 - \sigma_e^2))^2}, \tag{A.6}$$

respectively. From here we see that variation in the dynamic strategy weights depends on the degree of predictability.

Corollary 1 (The no-predictability case)**.** *When there is no in-sample predictability (as measured by the in-sample R^2 statistic), the maximum performance fee is negative in finite samples. Specifically,*

$$\lim_{R^2 \to 0} \delta = -\frac{\ln(1/(2 - g(1,p,T))}{2\gamma} \leq 0,$$

which equals zero only when an infinite sample is available (that is, when $g(1,p,T)$ equals unity).

Proof. The results follows directly from the expression given in Proposition 1. □

Remark. It follows that in-sample predictability is not a sufficient condition for economic significance. This result is consistent with an upper bound on predictability in efficient markets; see, e.g., Ross (2005), Huang and Zhou (2017) and Poti (2018).

References

Aparicio, D., and M. Bertolotto. 2020. Forecasting inflation with online prices. *International Journal of Forecasting* 36: 232–247.

Bali, T., and R. Engle. 2010. The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics* 57: 377–390.

Bossaerts, P., and P. Hillion. 1999. Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies* 12: 405–428.

Brandt, M., and P. Santa-Clara. 2006. Dynamic portfolio selection by augmenting the asset space. Journal of Finance 61: 2187–218.

Campbell, J., and J. Cochrane. 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107: 205–251.

Campbell, J., and S. Thompson. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21: 1509–1531.

Chan, K., and T. Marsh. 2022. Asset pricing on earnings announcement days. *Journal of Financial Economics* 144: 1022–1042.

Cong, L., G. Feng, J. He, and J. Li. 2022. Uncommon factors for Bayesian asset clusters. Available at `SSRN:https://ssrn.com/abstract=4219905orhttp://dx.doi.org/10.2139/ssrn.4219905`.

Cremers, M. 2002. Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies* 15: 1223–1249.

Cutler, D., J. Poterba, and L. Summers. 1989. What moves stock prices? *Journal of Portfolio Management* 15: 4–12.

Evgeniou, T., A. Guecioueur, and R. Prieto. 2023. Uncovering sparsity and heterogeneity in firm-level return predictability using machine learning. *Journal of Financial and Quantitative Analysis*, forthcoming.

Fama, E., and K. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.

Foerster, S., and A. Karolyi. 1999. The effects of market segmentation and investor recognition on asset prices: Evidence from foreign stocks listing in the United States. *Journal of Finance* 54: 981–1013.

Garleanu, N., and L. Pedersen. 2011. Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies* 24: 1980–2022.

Goyal, A., I. Welch, and A. Zafirov. 2023. A comprehensive 2021 look at the empirical performance of equity premium prediction II, Swiss Finance Institute Research Paper No. 21-85

Griffin, J. 2002. Are the Fama and French factors global or country specific? *Review of Financial Studies* 15: 783–803.

Guidolin, M., and A. Timmermann. 2008. International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies* 21: 889–935.

Guo, H., and R. Whitelaw. 2006. Uncovering the risk return relation in the stock market. *Journal of Finance* 61: 1433–1463.

Huang, D., and G. Zhou. 2017. Upper bounds on return predictability. *Journal of Financial and Quantitative Analysis* 52: 401–425.

Hendry, D., and K. Hubrich. 2011. Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics* 29: 216–227.

Hernandez-Murillo, R. and M. Owyang. 2006. The information content of regional employment data for forecasting aggregate conditions. Economics Letters 90: 335–339.

Joseph, A., G. Potjagailo, E. Kalamara, C. Chakraborty, and G. Kapetanios. 2022. Forecasting UK inflation bottom up, Bank of England Staff Working Paper No. 915.

Koijen, R., R. Richmond, and M. Yogo. 2022. Which investors matter for equity valuations and expected returns? University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2019-92.

Kolev, G., and R. Karapandza. 2017. Out-of-sample equity premium predictability and sample split–invariant inference. Journal of Banking and Finance 84: 188–201.

Leeb, H. 2009. Conditional predictive inference post model selection. *Annals of Statistics* 37: 2838–2876.

Lou, D., C. Polk, and S. Skouras. 2022. The day destroys the night, night extends the day: A clientele perspective on equity premium variation. London School of Economics Working Paper.

Lutkepohl, H., 2006. Forecasting with VARMA models. In: Handbook of Economic Forecasting. Elsevier, pp. 287–325.

Merton, R. 1973. An intertemporal capital asset pricing model. *Econometrica* 5: 867–87.

Merton, R. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42: 483–510.

Owyang, M., J. Piger, and H. Wall. 2015. Forecasting national recessions using state-level data. *Journal of Money, Credit and Banking* 47: 847–866.

Patton, A., and B. Weller. 2022. Risk price variation: The missing half of empirical asset pricing. *Review of Financial Studies* 35: 5127–5184.

Paulsen, K., T. Fastbo, and T. Ingebrigtsen, 2022. Aggregate density forecast of models using disaggregate data - A copula approach. Working Paper, No. 5/2022, Norges Bank, Oslo, `https://hdl.handle.net/11250/299750`.

Poti, V. 2018. A new tight and general bound on return predictability. *Economics Letters* 162: 140–145.

Rapach, D., and Zhou, G., 2013. Forecasting stock returns. In: Handbook of Economic Forecasting, vol. 2, pp. 328–383.

Rapach, D., Strauss, J. K., Tu, J., and Zhou, G., 2019. Industry return predictability: A machine learning approach. *Journal of Financial Data Science* 1: 9–28.

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

Rapach, D, and Zhou, G., 2020. Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods, Wiley, chap. 1, pp. 1–33.

Ross, S. 2005. Neoclassical Finance. Princeton University Press.

Rossi, A., and A. Timmermann. 2015. Modeling covariance risk in Merton's ICAPM. *Review of Financial Studies* 28: 1428–1461.

Scruggs, J. 1998. Resolving the puzzling intertemporal relation between the market risk premium and conditional market variance: A two-factor approach. *Journal of Finance* 53: 575–603.

Shleifer, A. and R. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52: 35–55.

Table 1 – The correlation between main portfolio returns and market returns

	Market Measure			
Component Portfolios	CRSP-VW	CRSP-EW	S&P 500	SPY ETF
Panel A. Industry portfolios				
5-industry portfolios	0.985	0.841	0.989	0.986
10-industry portfolios	0.972	0.845	0.974	0.973
12-industry portfolios	0.973	0.844	0.975	0.973
17-industry portfolios	0.956	0.853	0.948	0.948
30-industry portfolios	0.953	0.856	0.944	0.943
38-industry portfolios	0.941	0.852	0.929	0.929
48-industry portfolios	0.952	0.869	0.938	0.938
49-industry portfolios	0.954	0.871	0.940	0.940
Panel B. 6-component portfolios				
Size/book-to-market value sorted	0.952	0.924	0.916	0.915
Size/operating profitability sorted	0.953	0.899	0.928	0.927
Size/investment sorted	0.911	0.825	0.907	0.905
Size/momentum sorted	0.965	0.934	0.925	0.924
Size/short-term reversal sorted	0.954	0.895	0.930	0.929
Size/long-term reversal sorted	0.967	0.934	0.927	0.926
Size/cash flow sorted	0.967	0.946	0.927	0.927
Size/earnings sorted	0.956	0.943	0.922	0.923
Size/dividend yield sorted	0.951	0.924	0.921	0.920
Panel C. 25-component portfolios				
Size/book-to-market value sorted	0.945	0.946	0.918	0.919
Size/operating profitability sorted	0.969	0.908	0.960	0.961
Size/investment sorted	0.974	0.927	0.957	0.958
Size/momentum sorted	0.954	0.954	0.923	0.924
Size/short-term reversal sorted	0.950	0.943	0.923	0.924
Size/long-term reversal sorted	0.953	0.947	0.926	0.927
Size/accruals sorted	0.957	0.950	0.929	0.930
Size/market beta sorted	0.952	0.946	0.925	0.926
Size/net share issues sorted	0.956	0.963	0.926	0.928
Size/return variance (total) sorted	0.950	0.964	0.921	0.922
Size/return variance (residual) sorted	0.942	0.948	0.916	0.918
Book-to-market value/investment sorted	0.954	0.959	0.926	0.927
Book-to-market value/operating profitability sorted	0.952	0.957	0.924	0.925
Operating profitability/investment sorted	0.991	0.912	0.983	0.983

Notes: This table contains the correlation coefficients between each main portfolio return and each of four market returns (that is, value-weighted CRSP market portfolio index returns, equal-weighted CRSP market portfolio index returns, S&P 500 index returns, and S&P 500 (SPDR) ETF returns). Monthly frequency data observed over the period January 1994 to December 2022 are used.

Table 2 – In-sample performance

			MA cov.		EWMA cov.	
Predictor/test	$\mathcal{A}1$	$\mathcal{A}2$	$\mathcal{B}$	$\mathcal{C}$	$\mathcal{B}$	$\mathcal{C}$
Constant	0.630***	3.731	3.327	8.750**	2.158	5.244
	(0.161)	(2.329)	(2.314)	(3.943)	(2.379)	(4.266)
Term spread		0.000	−0.068	0.227	−0.098	−0.021
		(0.194)	(0.214)	(0.242)	(0.202)	(0.223)
Default yield		0.306	0.270	0.210	0.147	0.087
		(0.296)	(0.318)	(0.355)	(0.335)	(0.414)
Dividend yield		0.342	0.322	1.109***	0.258	0.728*
		(0.212)	(0.214)	(0.406)	(0.216)	(0.438)
Treasury bill yield		−0.556**	−0.577**	−0.119	−0.521*	−0.259
		(0.283)	(0.275)	(0.428)	(0.277)	(0.388)
Cond. var. (pseudo MK)			0.024	0.278	0.068*	1.332
			(0.031)	(1.699)	(0.037)	(1.601)
Cond. cov. (non-durables/MK)				0.055		−0.111
				(0.470)		(0.391)
Cond. cov. (durables/MK)				−0.025		−0.183
				(0.158)		(0.190)
Cond. cov. (manufacturing/MK)				−0.324		−0.227
				(0.385)		(0.291)
Cond. cov. (energy/MK)				0.129		−0.026
				(0.217)		(0.170)
Cond. cov. (hi-tech/MK)				−0.036		−0.171
				(0.191)		(0.179)
Cond. cov. (telecoms/MK)				0.318		0.083
				(0.284)		(0.243)
Cond. cov. (shops/MK)				−0.498**		−0.416**
				(0.206)		(0.196)
Cond. cov. (health/MK)				0.433***		0.167
				(0.147)		(0.162)
Cond. cov. (utilities/MK)				−0.092		−0.294
				(0.201)		(0.204)
100 × Adj.R^2	0.000	0.884	0.870	2.468	1.420	2.569
LR p-value (all predictors)		0.037	0.049	0.005	0.014	0.004
LR p-value (disagg. predictors only)				0.015		0.044

Notes: This table contains the in-sample estimated coefficients associated with OLS regressions (and restricted variants thereof) of main portfolio returns on the term spread, default yield, dividend yield, Treasury bill yield, the conditional variance of (pseudo) market returns, and the conditional covariances of market returns with returns to each component of the main portfolio. Standard errors are in parentheses. Fit is measured by the adjusted R^2 statistic. The p-values associated with likelihood-ratio tests under the null that all coefficients on the predictors equal zero, and under the null that all coefficients on the disaggregated predictors (that is, the conditional covariances) equal zero. Conditional variance and covariance are measured using the MA and EWMA methods. The main portfolio consists of 10-industry portfolios. Monthly frequency data observed over the period January 1964 to December 2021 are used. Coefficient significance is indicated by *** (1%), ** (5%) and * (10%).

Table 3 – The statistical performance of forecasting methods

			Performance measure										
				OOS-R^2					∇PaR				
Method	Comm. pred.	p	R^2	T=120	T=240	T=360	T=480	T=full	T=120	T=240	T=360	T=480	T=full
Panel A. MA covariance													
$\mathcal{A}$	N	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\mathcal{B}$	N	2	0.447	−0.404	0.027	0.168	0.238	0.303	−1.482	−0.532	−0.045	−0.046	0.200
$\mathcal{C}$	N	11	2.831	−6.167	−1.431	0.038	0.754	1.410	−11.757	−3.854	−1.488	−0.296	0.660
$\mathcal{A}$	Y	5	1.455	−2.003	−0.230	0.341	0.623	0.883	−4.988	−1.693	−0.492	−0.009	0.476
$\mathcal{B}$	Y	6	1.583	−2.771	−0.529	0.189	0.543	0.869	−6.281	−2.044	−0.847	−0.122	0.362
$\mathcal{C}$	Y	15	4.432	−8.433	−1.541	0.543	1.549	2.465	−15.700	−4.725	−1.467	0.174	1.586
Panel B. EWMA covariance													
$\mathcal{A}$	N	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\mathcal{B}$	N	2	1.440	0.597	1.024	1.164	1.233	1.298	−0.469	0.471	0.953	0.952	1.195
$\mathcal{C}$	N	11	4.051	−4.833	−0.157	1.294	2.001	2.648	−10.353	−2.550	−0.213	0.964	1.908
$\mathcal{A}$	Y	5	1.455	−2.003	−0.230	0.341	0.623	0.883	−4.988	−1.693	−0.492	−0.009	0.476
$\mathcal{B}$	Y	6	2.129	−2.201	0.029	0.743	1.095	1.419	−5.691	−1.478	−0.287	0.434	0.915
$\mathcal{C}$	Y	15	4.531	−8.320	−1.436	0.646	1.651	2.566	−15.580	−4.617	−1.362	0.278	1.688

Notes: This table contains the percentage in-sample and out-of-sample R^2 and performance risk efficiencies (based on 5% PaR values) associated with methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. The methods either exclude or include the common predictors. The following sample sizes (T) are assumed: 120, 240, 360, 480, and 696 (that is the full sample available). Conditional covariance is measured using the MA and EWMA methods. The main portfolio consists of 10-industry component portfolios. Monthly frequency data observed over the period January 1964 to December 2021 are used.

Table 4 – The economic value of forecasting methods

			Performance measure									
			Fee (50/50 portfolio benchmark)					Fee (buy-and-hold portfolio benchmark)				
Method	Comm. pred.	p	T=120	T=240	T=360	T=480	T=full	T=120	T=240	T=360	T=480	T=full
Panel A. MA covariance												
$\mathcal{A}$	N	1	−0.720	−0.359	−0.239	−0.179	−0.124	−1.440	−0.719	−0.479	−0.359	−0.247
$\mathcal{B}$	N	2	−1.073	−0.335	−0.093	0.027	0.139	−2.146	−0.669	−0.185	0.055	0.278
$\mathcal{C}$	N	11	−6.219	−1.542	−0.139	0.537	1.151	−12.439	−3.083	−0.277	1.074	2.302
$\mathcal{A}$	Y	5	−2.472	−0.541	0.073	0.375	0.653	−4.945	−1.082	0.146	0.750	1.306
$\mathcal{B}$	Y	6	−3.158	−0.798	−0.055	0.309	0.644	−6.316	−1.596	−0.109	0.619	1.288
$\mathcal{C}$	Y	15	−8.252	−1.538	0.394	1.311	2.139	−16.503	−3.076	0.787	2.623	4.277
Panel B. EWMA covariance												
$\mathcal{A}$	N	1	−0.720	−0.359	−0.239	−0.179	−0.124	−1.440	−0.719	−0.479	−0.359	−0.247
$\mathcal{B}$	N	2	−0.181	0.542	0.779	0.897	1.006	−0.363	1.084	1.558	1.794	2.012
$\mathcal{C}$	N	11	−4.910	−0.357	1.010	1.668	2.266	−9.820	−0.714	2.019	3.336	4.533
$\mathcal{A}$	Y	5	−2.472	−0.541	0.073	0.375	0.653	−4.945	−1.082	0.146	0.750	1.306
$\mathcal{B}$	Y	6	−2.628	−0.296	0.439	0.799	1.130	−5.257	−0.591	0.879	1.598	2.260
$\mathcal{C}$	Y	15	−8.136	−1.438	0.490	1.405	2.231	−16.272	−2.876	0.979	2.811	4.462

Notes: This table contains the annualised percentage fees associated with methods $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. The methods either exclude or include the common predictors. The following sample sizes (T) are assumed: 120, 240, 360, 480, and 696 (that is the full sample available). Conditional covariance is measured using the EWMA method. Two different benchmark strategies are used: one with a 50% weight in the risky asset, and one with a 100% weight in the risky asset. The main portfolio consists of 10-industry component portfolios. Monthly frequency data observed over the period January 1964 to December 2021 are used.

Table 5 – The aggregated success rates of forecasting methods (MA covariance, 6-component portfolios)

	Performance measure							
	OOS-R^2 success rate				∇PaR success rate			
Method	FF1	FF3	FF4	FF5	FF1	FF3	FF4	FF5
Panel A. US data: 1964 to 2022 (max sample)								
$\mathcal{A}$	20.166	22.302	22.229	21.075	45.195	60.253	64.704	65.916
$\mathcal{B}$	33.297	30.455	30.794	32.381	40.657	20.620	16.962	15.065
$\mathcal{C}$	46.537	47.244	46.977	46.544	14.149	19.127	18.333	19.019
Panel B. US data: 1964 to 1990 (max sample)								
$\mathcal{A}$	12.762	17.271	19.002	13.446	45.129	55.878	54.791	50.483
$\mathcal{B}$	20.048	15.177	23.752	35.467	27.415	10.467	20.733	21.377
$\mathcal{C}$	67.190	67.552	57.246	51.087	27.456	33.655	24.477	28.140
Panel C. US data: 1991 to 2022 (max sample)								
$\mathcal{A}$	11.357	12.315	13.273	7.827	32.540	39.956	44.171	40.394
$\mathcal{B}$	29.995	30.843	28.517	44.007	46.853	29.721	24.330	30.679
$\mathcal{C}$	58.648	56.842	58.210	48.166	20.608	30.323	31.500	28.927
Panel D. Asia-Pacific (excluding Japan) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	9.360	12.131	5.480	13.547	39.470	40.579	33.067	38.793
$\mathcal{B}$	8.128	6.773	29.126	14.039	4.988	3.941	20.567	13.424
$\mathcal{C}$	82.512	81.096	65.394	72.414	55.542	55.480	46.367	47.783
Panel E. Developed data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.387	5.049	4.865	3.571	13.177	14.286	14.101	15.764
$\mathcal{B}$	7.759	10.099	36.946	30.727	13.793	11.700	40.825	27.217
$\mathcal{C}$	88.855	84.852	58.190	65.702	73.030	74.015	45.074	57.020
Panel F. Developed (excluding US) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.264	4.187	4.372	4.126	12.131	12.377	11.268	11.700
$\mathcal{B}$	5.111	10.406	44.643	16.379	8.621	13.608	51.416	15.456
$\mathcal{C}$	91.626	85.406	50.985	79.495	79.249	74.015	37.315	72.845
Panel G. Europe data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.325	2.586	2.340	1.416	13.670	20.874	12.315	20.259
$\mathcal{B}$	7.943	12.254	51.108	38.916	16.256	12.007	54.680	33.374
$\mathcal{C}$	88.732	85.160	46.552	59.667	70.074	67.118	33.005	46.367
Panel H. Japan data: 1991 to 2022 (max sample)								
$\mathcal{A}$	14.409	8.682	7.882	9.544	38.362	28.879	20.135	29.741
$\mathcal{B}$	11.576	50.616	66.749	58.498	16.564	51.663	66.441	52.094
$\mathcal{C}$	74.015	40.702	25.369	31.958	45.074	19.458	13.424	18.165
Panel I. North America data: 1991 to 2022 (max sample)								
$\mathcal{A}$	8.990	8.190	8.313	7.389	28.448	27.340	25.616	23.153
$\mathcal{B}$	6.034	15.517	24.384	26.355	19.150	16.995	26.416	21.429
$\mathcal{C}$	84.975	76.293	67.303	66.256	52.401	55.665	47.968	55.419

Notes: This table contains the proportion of times (as a percentage) that each method is the dominant method (that is, has a higher OOS-R^2 or ∇PaR value than the other methods) under various Fama-French factor model specifications. When using the 1964 to 2022 sample (US dataset only), dominance is aggregated over 1540 sub-samples and 14 different portfolio sorts, when using the 1964 to 1990 sample (US dataset only), it is 171 sub-samples and 14 different portfolio sorts, and when using the 1991 to 2022 sample (all datasets) it is 406 sub-samples and four different portfolio sorts. Conditional covariance is measured using the MA method. The main portfolio consists of 6-component portfolios based on sorts of various firm characteristics. Monthly frequency data observed over the period January 1964 to December 2022 (US data only) and the period January 1991 to December 2022 are used.

Table 6 – The aggregated success rates of forecasting methods (EWMA covariance, 6-component portfolios)

	Performance measure							
	OOS-R^2 success rate				∇PaR success rate			
Method	FF1	FF3	FF4	FF5	FF1	FF3	FF4	FF5
Panel A. US data: 1964 to 2022 (max sample)								
$\mathcal{A}$	7.893	8.001	5.851	6.659	24.076	31.342	35.548	34.473
$\mathcal{B}$	28.925	33.434	34.105	38.968	47.655	35.967	27.107	30.483
$\mathcal{C}$	63.182	58.564	60.043	54.372	28.268	32.691	37.345	35.043
Panel B. US data: 1964 to 1990 (max sample)								
$\mathcal{A}$	5.354	8.132	5.636	5.193	20.531	27.899	28.060	19.646
$\mathcal{B}$	27.939	27.134	36.151	56.200	50.604	35.507	36.071	52.657
$\mathcal{C}$	66.707	64.734	58.213	38.607	28.865	36.594	35.870	27.697
Panel C. US data: 1991 to 2022 (max sample)								
$\mathcal{A}$	6.349	5.337	5.911	3.886	20.881	25.534	29.475	24.001
$\mathcal{B}$	10.947	18.966	14.587	28.188	30.022	24.658	14.833	25.780
$\mathcal{C}$	82.704	75.698	79.502	67.926	49.097	49.808	55.692	50.219
Panel D. Asia-Pacific (excluding Japan) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.355	0.924	0.493	1.355	12.192	12.685	9.852	11.268
$\mathcal{B}$	1.663	5.603	26.786	15.025	7.759	6.404	32.081	15.271
$\mathcal{C}$	96.983	93.473	72.722	83.621	80.049	80.911	58.067	73.461
Panel E. Developed data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.017	2.894	3.079	2.833	9.236	11.946	10.406	9.791
$\mathcal{B}$	5.172	6.835	27.155	52.340	10.653	13.793	33.867	51.539
$\mathcal{C}$	91.810	90.271	69.766	44.828	80.111	74.261	55.727	38.670
Panel F. Developed (excluding US) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.909	2.155	2.278	1.601	9.236	9.667	9.421	8.190
$\mathcal{B}$	5.049	10.899	30.357	26.416	10.837	14.347	36.145	25.369
$\mathcal{C}$	93.042	86.946	67.365	71.983	79.926	75.985	54.433	66.441
Panel G. Europe data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.970	1.847	0.862	0.985	14.717	15.579	8.990	9.052
$\mathcal{B}$	5.234	14.347	43.904	48.707	13.424	19.581	47.106	47.044
$\mathcal{C}$	92.796	83.805	55.234	50.308	71.860	64.840	43.904	43.904
Panel H. Japan data: 1991 to 2022 (max sample)								
$\mathcal{A}$	8.682	8.067	7.266	8.867	27.771	23.830	19.951	25.062
$\mathcal{B}$	7.635	27.709	32.081	30.234	13.054	31.466	33.867	28.571
$\mathcal{C}$	83.682	64.224	60.653	60.899	59.175	44.704	46.182	46.367
Panel I. North America data: 1991 to 2022 (max sample)								
$\mathcal{A}$	2.894	3.264	3.079	3.633	18.473	12.685	16.379	16.441
$\mathcal{B}$	4.926	20.135	20.751	22.906	14.594	26.170	21.059	23.399
$\mathcal{C}$	92.180	76.601	76.170	73.461	66.933	61.145	62.562	60.160

Notes: This table contains the proportion of times (as a percentage) that each method is the dominant method (that is, has a higher OOS-R^2 or ∇PaR value than the other methods) under various Fama-French factor model specifications. When using the 1964 to 2022 sample (US dataset only), dominance is aggregated over 1540 sub-samples and 14 different portfolio sorts, when using the 1964 to 1990 sample (US dataset only), it is 171 sub-samples and 14 different portfolio sorts, and when using the 1991 to 2022 sample (all datasets) it is 406 sub-samples and four different portfolio sorts. Conditional covariance is measured using the EWMA method. The main portfolio consists of 6-component portfolios based on sorts of various firm characteristics. Monthly frequency data observed over the period January 1964 to December 2022 (US data only) and the period January 1991 to December 2022 are used.

Table 7 – The aggregated success rates of forecasting methods (MA covariance, 25-component portfolios)

	Performance measure							
	OOS-R^2 success rate				∇PaR success rate			
Method	FF1	FF3	FF4	FF5	FF1	FF3	FF4	FF5
Panel A. US data: 1964 to 2022 (max sample)								
$\mathcal{A}$	3.618	3.664	3.669	3.131	25.877	31.331	32.259	31.011
$\mathcal{B}$	3.780	3.961	5.079	6.749	19.244	10.649	9.740	11.039
$\mathcal{C}$	92.602	92.375	91.252	90.121	54.879	58.019	58.001	57.950
Panel B. US data: 1964 to 1990 (max sample)								
$\mathcal{A}$	3.623	3.546	3.856	2.562	31.082	43.375	39.855	31.910
$\mathcal{B}$	9.627	9.420	15.114	20.109	33.049	17.417	23.810	31.004
$\mathcal{C}$	86.749	87.034	81.030	77.329	35.870	39.208	36.335	37.086
Panel C. US data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.413	3.835	4.293	2.797	17.365	18.526	19.775	14.743
$\mathcal{B}$	2.709	2.604	2.041	6.756	13.406	9.096	7.196	14.409
$\mathcal{C}$	93.878	93.561	93.666	90.447	69.229	72.379	73.030	70.848
Panel D. Asia-Pacific (excluding Japan) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.416	1.047	0.431	1.416	7.266	7.204	6.527	5.172
$\mathcal{B}$	0.246	0.800	2.032	2.032	1.786	2.032	2.833	5.603
$\mathcal{C}$	98.337	98.153	97.537	96.552	90.948	90.764	90.640	89.224
Panel E. Developed data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.909	2.586	2.709	1.970	11.453	11.207	11.823	10.714
$\mathcal{B}$	2.586	2.278	1.909	4.187	5.850	5.911	7.882	10.406
$\mathcal{C}$	95.505	95.135	95.382	93.842	82.697	82.882	80.296	78.879
Panel F. Developed (excluding US) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	2.833	2.894	3.633	2.709	10.037	9.175	9.113	8.559
$\mathcal{B}$	1.909	2.463	3.264	3.325	2.956	5.296	14.532	6.650
$\mathcal{C}$	95.259	94.643	93.103	93.966	87.007	85.530	76.355	84.791
Panel G. Europe data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.847	1.355	1.293	0.862	11.884	10.961	9.729	9.544
$\mathcal{B}$	2.463	4.926	7.266	11.884	7.143	9.483	21.675	25.924
$\mathcal{C}$	95.690	93.719	91.441	87.254	80.973	79.557	68.596	64.532
Panel H. Japan data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.633	5.049	4.618	5.480	25.554	20.874	17.796	20.074
$\mathcal{B}$	6.404	13.054	15.333	18.842	14.963	35.099	40.025	37.808
$\mathcal{C}$	89.963	81.897	80.049	75.677	59.483	44.027	42.180	42.118
Panel I. North America data: 1991 to 2022 (max sample)								
$\mathcal{A}$	3.695	4.310	4.803	4.372	20.012	20.012	20.628	17.919
$\mathcal{B}$	2.894	3.264	2.894	6.158	16.995	14.470	14.778	15.825
$\mathcal{C}$	93.411	92.426	92.303	89.470	62.993	65.517	64.594	66.256

Notes: This table contains the proportion of times (as a percentage) that each method is the dominant method (that is, has a higher OOS-R^2 or ∇PaR value than the other methods) under various Fama-French factor model specifications. When using the 1964 to 2022 sample (US dataset only), dominance is aggregated over 1540 sub-samples and 14 different portfolio sorts, when using the 1964 to 1990 sample (US dataset only), it is 171 sub-samples and 14 different portfolio sorts, and when using the 1991 to 2022 sample (all datasets) it is 406 sub-samples and four different portfolio sorts. Conditional covariance is measured using the MA method. The main portfolio consists of 25-component portfolios based on sorts of various firm characteristics. Monthly frequency data observed over the period January 1964 to December 2022 (US data only) and the period January 1991 to December 2022 are used.

Table 8 – The aggregated success rates of forecasting methods (EWMA covariance, 25-component portfolios)

	Performance measure							
	OOS-R^2 success rate				∇PaR success rate			
Method	FF1	FF3	FF4	FF5	FF1	FF3	FF4	FF5
Panel A. US data: 1964 to 2022 (max sample)								
$\mathcal{A}$	1.438	1.433	0.779	0.965	12.175	13.868	14.374	13.182
$\mathcal{B}$	2.890	3.353	4.443	5.441	13.646	10.645	9.272	12.166
$\mathcal{C}$	95.673	95.213	94.777	93.595	74.179	75.487	76.354	74.652
Panel B. US data: 1964 to 1990 (max sample)								
$\mathcal{A}$	1.449	2.019	1.449	1.087	12.189	17.107	17.314	12.759
$\mathcal{B}$	5.564	5.978	7.790	13.742	32.764	24.534	23.473	35.637
$\mathcal{C}$	92.987	92.003	90.761	85.171	55.047	58.359	59.213	51.605
Panel C. US data: 1991 to 2022 (max sample)								
$\mathcal{A}$	0.827	0.686	0.757	0.616	9.289	8.023	8.955	6.175
$\mathcal{B}$	0.352	1.108	0.968	2.762	2.270	4.082	2.604	7.477
$\mathcal{C}$	98.821	98.205	98.276	96.622	88.441	87.896	88.441	86.348
Panel D. Asia-Pacific (excluding Japan) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	0.185	0.123	0.123	0.000	2.032	1.724	1.478	0.862
$\mathcal{B}$	0.062	0.431	0.924	1.416	1.355	2.217	2.771	4.187
$\mathcal{C}$	99.754	99.446	98.953	98.584	96.613	96.059	95.751	94.951
Panel E. Developed data: 1991 to 2022 (max sample)								
$\mathcal{A}$	0.677	0.800	0.616	0.616	5.788	5.542	5.234	4.803
$\mathcal{B}$	0.369	0.431	0.800	2.648	2.340	4.926	4.803	7.020
$\mathcal{C}$	98.953	98.768	98.584	96.736	91.872	89.532	89.963	88.177
Panel F. Developed (excluding US) data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.970	1.478	1.909	1.601	10.222	7.512	7.943	7.389
$\mathcal{B}$	1.108	4.187	6.342	5.542	5.480	9.914	13.300	11.207
$\mathcal{C}$	96.921	94.335	91.749	92.857	84.298	82.574	78.756	81.404
Panel G. Europe data: 1991 to 2022 (max sample)								
$\mathcal{A}$	0.739	0.616	0.308	0.308	5.788	4.495	3.818	2.833
$\mathcal{B}$	0.554	2.155	3.510	5.788	3.510	7.574	11.761	17.118
$\mathcal{C}$	98.707	97.229	96.182	93.904	90.702	87.931	84.421	80.049
Panel H. Japan data: 1991 to 2022 (max sample)								
$\mathcal{A}$	1.909	2.094	1.293	1.663	13.793	16.441	12.069	16.318
$\mathcal{B}$	3.941	4.741	6.835	5.172	9.360	8.682	13.362	8.498
$\mathcal{C}$	94.150	93.165	91.872	93.165	76.847	74.877	74.569	75.185
Panel I. North America data: 1991 to 2022 (max sample)								
$\mathcal{A}$	0.677	0.308	0.369	0.185	8.190	7.143	8.498	7.820
$\mathcal{B}$	0.246	1.539	1.047	2.401	7.697	8.374	6.342	7.943
$\mathcal{C}$	99.076	98.153	98.584	97.414	84.113	84.483	85.160	84.236

Notes: This table contains the proportion of times (as a percentage) that each method is the dominant method (that is, has a higher OOS-R^2 or ∇PaR value than the other methods) under various Fama-French factor model specifications. When using the 1964 to 2022 sample (US dataset only), dominance is aggregated over 1540 sub-samples and 14 different portfolio sorts, when using the 1964 to 1990 sample (US dataset only), it is 171 sub-samples and 14 different portfolio sorts, and when using the 1991 to 2022 sample (all datasets) it is 406 sub-samples and four different portfolio sorts. Conditional covariance is measured using the EWMA method. The main portfolio consists of 25-component portfolios based on sorts of various firm characteristics. Monthly frequency data observed over the period January 1964 to December 2022 (US data only) and the period January 1991 to December 2022 are used.

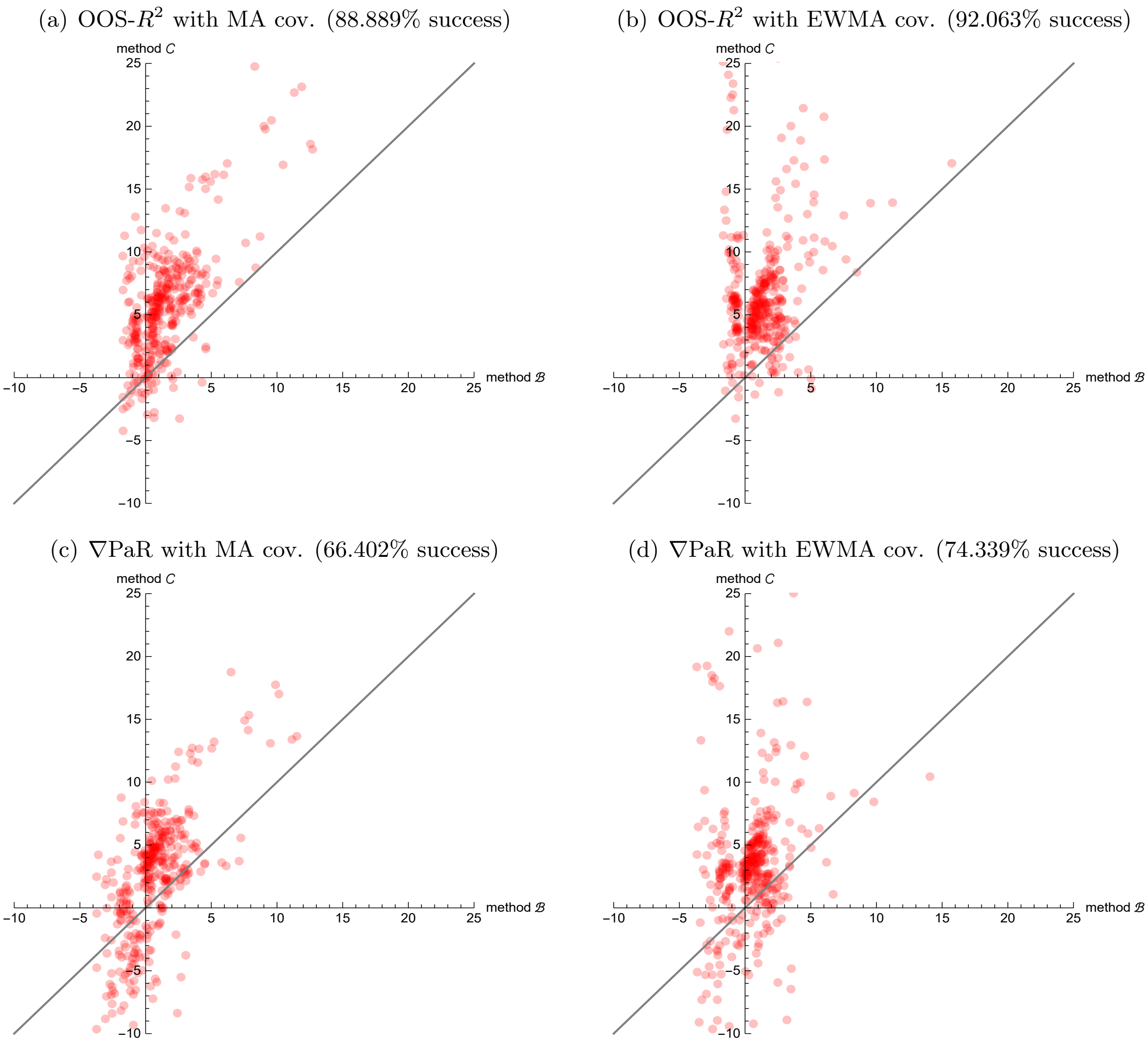
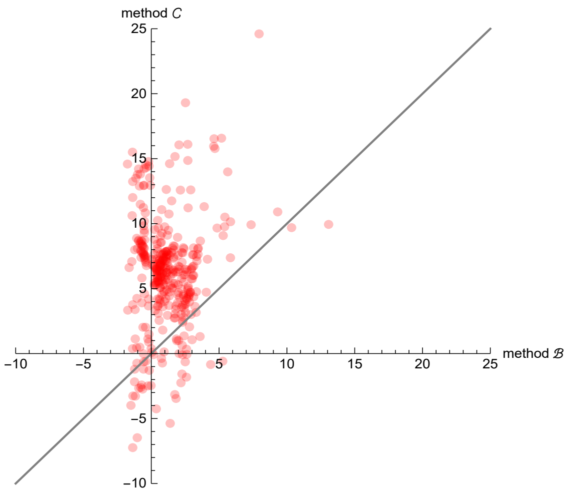


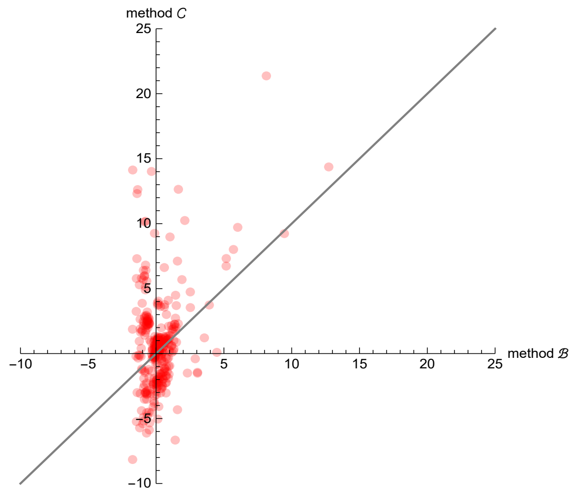
Figure 1 – Relative performance (global stock market prediction)

This figure contains the scatter plots of the OOS-R^2 (∇PaR) values associated with method $\mathcal{C}$ against the OOS-R^2 (∇PaR) values associated with method $\mathcal{B}$. The main portfolio consists of the Asia-Pacific (excluding Japan), Japan, North America, US, and developed (excluding US) market portfolios. The MK return is used as the state variable. Each point in the plot corresponds to a sub-sample result (a total of 406 sub-samples are considered). Conditional covariance is measured using the MA and EWMA methods. Monthly frequency data observed over the period January 1991 to December 2022 are used.

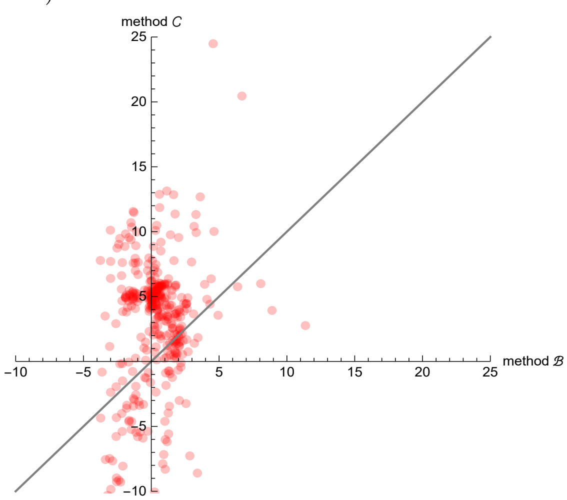
(a) OOS-R^2 with EWMA cov. 3 month (87.302% success)



(b) OOS-R^2 with EWMA cov. 12 month (56.878% success)



(c) ∇PaR with EWMA cov. 3 month (69.841% success)



(d) ∇PaR with EWMA cov. 12 month (8.466% success)

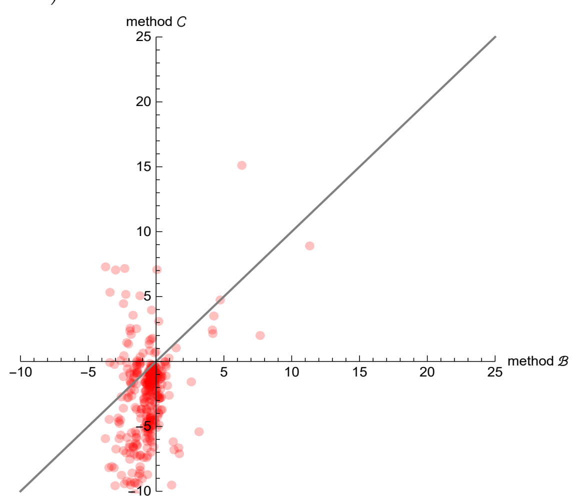


Figure 2 – Relative performance (global stock market prediction and alternative horizons)
This figure contains the scatter plots of the OOS-R^2 (∇PaR) values associated with method $\mathcal{C}$ against the OOS-R^2 (∇PaR) values associated with method $\mathcal{B}$. The forecasting horizons considered are 3-month and 12-month step ahead. The main portfolio consists of the Asia-Pacific (excluding Japan), Japan, Europe, North America, US, and developed (excluding US) market portfolios. The MK return is used as the state variable. Each point in the plot corresponds to a sub-sample result (a total of 406 sub-samples are considered). Conditional covariance is measured using the EWMA method. Monthly frequency data observed over the period January 1991 to December 2022 are used.

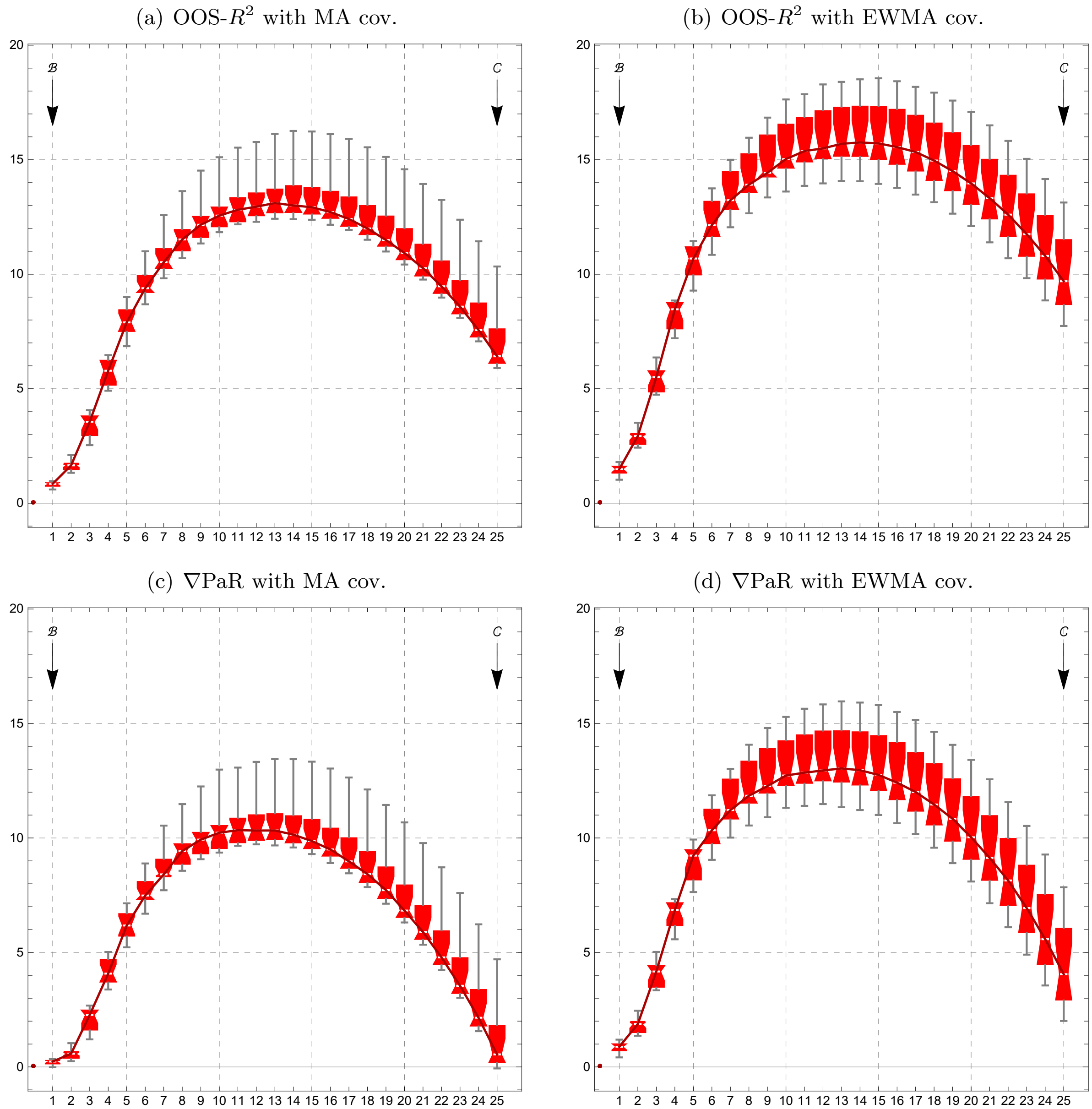


Figure 3 – Relative performance with clustered predictors

This figure contains box-and-whisker plots of the OOS-R^2 (∇PaR) values associated with methods $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{C}$ (with clustering), plotted against the number of clusters used to construct the conditional covariances. Each box contains results based on using 14 US main portfolios each consisting of 25-component portfolios, and 1540 sub-samples. The MK return is used as the state variable. Component conditional covariances are constructed using the MA and EWMA methods. Monthly frequency data observed over the period January 1991 to December 2022 are used.