Predicting the global minimum variance portfolio

by Laura Reh, Fabian Krüger, Roman Liesenfeld

Predicting the Global Minimum Variance Portfolio[*]

Laura Reh[a,b]   Fabian Krüger[c]   Roman Liesenfeld[b]

July 5, 2020

Abstract

We propose a novel dynamic approach to forecast the weights of the global minimum variance portfolio (GMVP). The GMVP weights are the population coefficients of a linear regression of a benchmark return on a vector of return differences. This representation enables us to derive a consistent loss function from which we can infer the optimal GMVP weights without imposing any distributional assumptions on the returns. In order to capture time variation in the returns' conditional covariance structure, we model the portfolio weights through a recursive least squares (RLS) scheme as well as by generalized autoregressive score (GAS) type dynamics. Sparse parameterizations combined with targeting towards nonlinear shrinkage estimates of the long-run GMVP weights ensure scalability with respect to the number of assets. An empirical analysis of daily and monthly financial returns shows that the proposed models perform well in- and out-of-sample in comparison to existing approaches.

JEL classification: C14, C32, C51, C53, C58, G11, G17

Keywords: Consistent loss function; Elicitability; Forecasting; Generalized autoregressive score; Nonlinear shrinkage; Recursive least squares.

1 Introduction

The global minimum variance portfolio (GMVP) allocates a given budget among n financial assets such that the risk for the rate of expected portfolio return is minimized. In contrast to the classical mean-variance optimal portfolio (Markowitz, 1952), the weights of the GMVP do not depend on the expected returns of the assets. The expected returns have a major impact on the mean-variance optimal strategy (Best and Grauer, 1991) but are notoriously hard to predict (Welch and Goyal, 2008). Studies like Jagannathan and Ma (2003) hence advocate the use of the GMVP, which only depends on the covariance matrix of the asset returns. Conceptually, the GMVP is a special case of the mean-variance optimal portfolio because it results from the latter either if all assets have the same expected return or if the investor is infinitely risk averse.

In the present paper, we develop a dynamic approach for predicting the GMVP weights. Our approach builds upon a consistent loss function (Gneiting, 2011) that arises from a representation of the GMVP weights as population coefficients in an auxiliary linear regression problem (Kempf and Memmel, 2006). We first provide a new theoretical result which further justifies the use of this loss function for portfolio selection problems. In particular, we show that the expected GMVP return entering the loss function as a nuisance parameter is unavoidable since there can be no loss function that uniquely identifies the GMVP weights on their own. We then use the loss function in a dynamic context to develop time series models for the GMVP weights, relying on a simple recursive least squares (RLS) with forgetting factor approach (Ljung and Söderström, 1983) as well as generalized autoregressive score (GAS; Creal et al., 2013) recursions. Our approach combines a direct dynamic parametrization of the weights with a consistent loss function for estimating the parameters and evaluating the resulting predictions. This unified setup is in contrast to much existing work that uses different loss functions for estimating versus evaluating an econometric model. In combining a 'non-standard' but consistent loss function with GAS specifications we follow Patton et al. (2019) who consider dynamic models for Expected Shortfall and Value-at-Risk.

When the number of assets is large a dynamic GMVP model requires a parsimonious parameterization in order to be tractable in practice. While our RLS-based model is parsimonious by construction, the GAS version requires parameter restrictions in order to be applicable in high dimensions. To achieve parsimony, the GMVP-GAS approach can be combined with the restriction that the long-run mean of the GMVP weights be equal to those of a benchmark portfolio or with a targeting approach by which the long-run GMVP weights are estimated from the sample covariance matrix of the asset returns. However, since the standard sample covariance matrix is a poor estimator of the true covariance matrix when the number of assets is large compared to the sample size (Ledoit and Wolf, 2003), we suggest using the nonlinear shrinkage approach of Ledoit and Wolf (2012) for estimating the long-run GMVP targeting weights. In an empirical analysis of daily and monthly financial returns, our dynamic GMVP approaches based on RLS and GAS perform well compared to a wide range of benchmarks from the literature, demonstrating their usefulness for portfolio allocations.

Existing approaches for estimating and predicting GMVP weights can be broadly classified into three types. The first type considers dynamic models for the covariance matrix of the asset returns and constructs a plug-in prediction of the GMVP weights from a forecast

of the covariance matrix (see, e.g., Ledoit et al. 2003, Engle and Kelly 2012, Clements et al. 2015). This approach is designed to account for the well-documented time variation in the assets' covariance structure, which can be expected to be important for GMVP predictions. The covariance models are typically estimated using (quasi) maximum likelihood (ML) or related techniques. Clearly, if the covariance model is correctly specified, ML asymptotically identifies the correct model, which results in optimal forecasts of the GMVP weights. However, this approach can be problematic if the model is misspecified, such that not all properties of the data-generating process (DGP) are correctly captured. Then the estimator resorts to minimizing a measure of discrepancy between the covariance model and the data (such as the Kullback-Leibler divergence in the ML case). However, such measures are not directly related to the economic problem at hand so that there is the risk that the fitted miss-specified model fails to capture the economically critically important properties of the DGP while at the same time capturing irrelevant ones. See Elliott et al. (2016) for a more detailed discussion and further references.

The second type of approaches uses a static framework for the GMVP based on either the sample covariance matrix of the asset returns or a shrinkage version thereof (see, e.g., Ledoit and Wolf, 2003, 2004; DeMiguel et al., 2009; Frahm and Memmel, 2010; Candelon et al., 2012; Frey and Pohlmeier, 2016). As we detail further below, such a static approach is implicitly based on the Kempf and Memmel (2006) loss function for the GMVP weights, and thus relies on an estimation principle which under miss-specification appears to be advantageous compared to likelihood-based estimation that is typically used for the plug-in strategies mentioned above. A further promising strategy is to directly estimate the inverse of the covariance matrix, rather than the covariance matrix itself. In particular, Callot et al. (2019) present a sparse estimation technique that allows for consistent estimation of the GMVP portfolio weights and the Markowitz (1952) portfolio weights. By definition, the static approach aims for the long-run (or unconditional) vector of portfolio weights. Hence it abstracts from any short-run fluctuations in the portfolio weights, which seems restrictive in view of well-documented time variation in the covariance matrix of asset returns (see previous paragraph, as well as Section 2.1 below).

A third group of studies models the portfolio weights as functions of potentially relevant state variables such as firm characteristics; see Brandt (2009, Section 4) for a review. The optimal relation between the portfolio weights and the state variables are found by optimizing a pre-specified utility function. However, in this approach dynamics in the portfolio weights are accounted for only indirectly via time variation in the state variables.

Our proposed approach combines what we consider the strengths of those three groups of studies: The time series dynamics featured by the first group, the loss function perspective taken by the second and third group, and the third group's proposal to model the weights directly. In contrast to the third group of studies, we construct dynamic models for the weights using pure autoregressive specifications (with current weights assumed to depend on past weights). This modeling approach allows us to easily represent key properties of the GMVP weights, such as their persistence or their long-run averages under stationarity. From a conceptual perspective, our use of an economically motivated loss function is in line with studies such as Weiss (1996), Hand and Vinciotti (2003), Christoffersen and Jacobs (2004) and Engle and Colacito (2006). In addition to its economic appeal, we demonstrate that the GMVP loss function is convenient for practical modeling purposes.

The remainder of this paper is organized as follows. Section 2 introduces the Kempf and Memmel (2006) GMVP loss function and provides a theoretical result motivating its use. Section 3 introduces RLS and GAS models for forecasting the GMVP weights. Section 4 presents empirical results, and Section 5 concludes. Proofs and implementation details are deferred to the online Appendix.

2 Consistent loss function for the GMVP

2.1 Setup

Let $R_t = (R_{1t}, ..., R_{nt})'$ denote a vector of returns on n assets at period t. For ease of exposition, we initially assume that R_t is independent across time, and denote the covariance matrix of R_t by Σ. The vector of weights representing the GMVP for the n assets (subject to the constraint that they sum to one) is denoted by $w^* = (w_1^*, \ldots, w_N^*)'$ and obtains as

$$w^* = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}, \qquad \iota'w^* = 1, \tag{1}$$

where ι is an $n \times 1$ vector of ones.

According to Kempf and Memmel (2006) the GMVP weights w^* can be represented using the following auxiliary linear regression:

$$Y_t = X_t'\beta + \varepsilon_t, \qquad \mathbb{E}(\varepsilon_t|X_t) = 0, \tag{2}$$

where $Y_t = R_{nt}$ is the return of an (arbitrarily selected) baseline asset and the vector $X_t = (1, R_{nt} - R_{1t}, \ldots, R_{nt} - R_{n-1t})'$ consists of the return differences between the baseline asset and the remaining $n - 1$ ones (including a one). The corresponding population regression coefficients are $\beta = (\beta_0, \ldots, \beta_{n-1})'$, and are defined by

$$\beta = \arg\min_b \mathbb{E}[L(b, R_t)], \tag{3}$$

where

$$L(b, R_t) = (Y_t - X_t'b)^2. \tag{4}$$

As shown in Kempf and Memmel (2006, Proposition 1), the population coefficients for the slopes in the auxiliary regression (2) coincide with the true GMVP weights while the intercept represents the expected return of the GMVP, in that

$$w_i^* = \begin{cases} \beta_i & i = 1, \ldots, n-1 \\ 1 - \sum_{j=1}^{n-1} \beta_j & i = n \end{cases}, \qquad \mathbb{E}[R_t'\omega^*] = \beta_0. \tag{5}$$

In the terminology of Gneiting (2011), $L(b, R_t)$ as given by (4) and used in (3) is a strictly consistent scoring (or loss) function for β, which implies that the vector β is elicitable[1]. In

---

[1]A functional of a probability distribution is called elicitable if there exists a strictly consistent loss function for this functional. It is hence possible to define a loss function that incentivizes a risk neutral forecaster to state the value of the functional (see Gneiting 2011 for further discussion).

the following we will use $\beta_{1:n-1}$ to denote the sub-vector representing the GMVP weights $(\beta_1, \ldots, \beta_{n-1})'$ so that $\beta = (\beta_0, \beta'_{1:n-1})'$.

The intercept β_0 is the only element of β that is not directly related to the GMVP weights. It therefore appears tempting to formulate an alternative loss function that exclusively elicits the GMVP weights $\beta_{1:n-1}$. However, as we state in the following proposition such a loss function does not exist.

Proposition 1. *The sub-vector $\beta_{1:n-1}$ is not elicitable. That is, there is no loss function whose expected value is uniquely minimized by $\beta_{1:n-1}$.*

Proof. See online Appendix A. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Clearly, a loss function that elicits $\beta_{1:n-1}$ without β_0 would be attractive in principle. By stating that such a loss function does not exist, Proposition 1 motivates the use of Kempf and Memmel's loss function as a feasible and (strictly) consistent choice. We also note that the joint elicitability of the composite vector $\beta = (\beta_0, \beta'_{1:n-1})'$ parallels theoretical results on the joint elicitability of Value-at-Risk and Expected Shortfall (Fissler and Ziegel, 2016): Just like the GMVP weights $\beta_{1:n-1}$ are not elicitable without β_0, Expected Shortfall is not elicitable without Value-at-Risk.

So far, we have assumed that the returns R_t are temporally independent with a constant covariance matrix, such that the vector β with the GMVP weights is time invariant. However, it is well-known that financial returns are conditionally heteroscedastic. To account for this stylized fact we can replace the constant return covariance matrix Σ in Equation (1) by the conditional covariance matrix $\Sigma_t = \mathbb{V}(R_t|\mathcal{F}_{t-1})$ given the information set $\mathcal{F}_{t-1}$ comprising data up until time period $t-1$. Time variation in Σ_t typically implies time variation in the GMVP weights[2]. In this dynamic context, the time-dependent population coefficients of the auxiliary regression (2), based on the loss function (4), are given by

$$\beta_t = \arg \min_b \mathbb{E}_{t-1}[L(b, R_t)],$$

where $\mathbb{E}_{t-1}[\cdot] \equiv \mathbb{E}[\cdot|\mathcal{F}_{t-1}]$. Analogously to Equation (5), it follows that

$$w^*_{it} = \begin{cases} \beta_{it} & i = 1, \ldots, n-1 \\ 1 - \sum_{j=1}^{n-1} \beta_{jt} & i = n \end{cases}, \qquad \mathbb{E}_{t-1}(\omega^{*\prime}_t R_t) = \beta_{0t},$$

where $\omega^*_t = (\omega^*_{1t}, \ldots, \omega^*_{nt})'$ is the vector that minimizes the conditional portfolio variance given $\mathcal{F}_{t-1}$ and β_{0t} represents the conditional expectation of the GMVP return.

We use this dynamic GMVP framework based on the consistent loss function $L(b, R_t)$ to develop predictive models for the GMVP weights, to estimate their parameters and to evaluate their predictive performance. The models consist of parametric functions for the weights $\beta_{1:n-1t} = \omega^*_{1:n-1t}$ and the expected GMVP returns β_{0t} assumed to be measurable w.r.t. the information set $\mathcal{F}_{t-1}$,

$$(\beta_{0t}, \beta'_{1:n-1t})' = \beta_t = \beta(Z_{t-1}; \theta), \qquad t = 1, \ldots, T, \qquad (6)$$

---

[2]Certain restrictive forms of time variation in Σ_t are compatible with constant GMVP weights. Examples include the case $\Sigma_t = \tau_t \Sigma$, where $\tau_t \in \mathbb{R}_+$ is a positive-valued scalar and Σ is a constant matrix. By comparison with Equation (1), τ_t does not affect the GMVP weights.

where θ denotes the parameter indexing the model, T is the sample size, and $Z_{t-1} \in \mathcal{F}_{t-1}$. In the terminology of Patton et al. (2019) these GMVP models are semi-parametric as they do not impose a parametric class of conditional distributions on the asset returns. For estimating θ we follow Patton et al. (2019) and use an M-type estimator which minimizes the average GMVP loss, so that

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{T} \sum_{t=1}^{T} L\left(\beta\left(Z_{t-1}; \theta\right), R_t\right). \tag{7}$$

Given our choice of $L\left(\beta\left(Z_{t-1}; \theta\right), R_t\right)$ defined in Equation (4), $\hat{\theta}$ is hence a nonlinear least squares (NLS) estimator. Before we present our proposed dynamic GMVP models in Section 3, the next section compares the expected GMVP loss function and the variance of the portfolio returns as potential performance measures for GMVP models.

2.2 Relation between expected GMVP loss and variance of portfolio returns

In empirical studies on the GMVP, it is common to assess the performance of a prediction rule for the GMVP weights by using the sample variance of the portfolio returns $\{\omega_t' R_t\}$, with small values to be preferred (see, e.g., DeMiguel et al. 2009, Table 3 or Engle et al. 2019, Table 8). Here we detail the relationship between the sample variance and our proposed performance measure, the sample average of the GMVP loss function at (4).

Let $\{b_t\}$ be an arbitrary sequence of parameter values with period-t portfolio weights $\omega_{1:n-1t} = b_{1:n-1t}$ and intercept parameter b_{0t}. Then the conditional period-$(t-1)$ expectation of the period-t GMVP loss can be written as

$$\mathbb{E}_{t-1}[L(b_t, R_t)] = \mathbb{V}_{t-1}(\omega_t' R_t) + \{\mathbb{E}_{t-1}(\omega_t' R_t) - b_{0t}\}^2, \tag{8}$$

so that

$$\begin{aligned}
\mathbb{E}\left[L(b_t, R_t)\right] &= \mathbb{E}\left(\mathbb{E}_{t-1}\left[L(b_t, R_t)\right]\right) \\
&= \mathbb{V}\left(\omega_t' R_t\right) - 2\,\mathbb{C}ov\left[b_{0t}, \mathbb{E}_{t-1}(\omega_t' R_t)\right] + \mathbb{V}(b_{0t}) \\
&\quad + \{\mathbb{E}(\omega_t' R_t) - \mathbb{E}(b_{0t})\}^2.
\end{aligned} \tag{9}$$

Equation (9) states that the expected loss $\mathbb{E}[L(b_t, R_t)]$ differs from the variance of the portfolio returns $\mathbb{V}(\omega_t' R_t)$ that is usually taken as a performance measure in empirical applications. The following special cases are worth mentioning. First, if b_{0t} is constant over time, with $b_0 = \mathbb{E}(\omega_t' R_t)$, the difference between the expected loss and the variance of the portfolio returns collapses. This case is discussed by Kempf and Memmel (2006, Proposition 1, Statement 3) in the framework where returns are temporally independent with time-invariant moments. Second, if b_{0t} is the best predictor for the portfolio returns, i.e. if $b_{0,t} = \mathbb{E}_{t-1}(\omega_t' R_t)$ $\forall t$, we have that $\mathbb{C}ov[b_{0t}, \mathbb{E}_{t-1}(\omega_t' R_t)] = \mathbb{V}(b_{0t})$ and $\mathbb{E}(\omega_t' R_t) = \mathbb{E}(b_{0t})$ so that the expected loss simplifies to $\mathbb{E}[L(b_t, R_t)] = \mathbb{V}(\omega_t' R_t) - \mathbb{V}(b_{0t})$.

This result raises the question of whether $\mathbb{E}[L(b_t, R_t)]$ or $\mathbb{V}(\omega_t' R_t)$ should be used for forecast evaluation. Mirroring the results in Section 2.1, neither of the two alternatives

exclusively evaluates forecasts of the GMVP weights, as would be desirable. That $\mathbb{V}(\omega_t' R_t)$ does not exclusively evaluate forecasts for the GMVP weights becomes evident from the law of total variance which implies that

$$\mathbb{V}(\omega_t' R_t) = \mathbb{E}\left[\mathbb{V}_{t-1}(\omega_t' R_t)\right] + \mathbb{V}\left[\mathbb{E}_{t-1}(\omega_t' R_t)\right]. \tag{10}$$

The first summand in (10) measures the average predicted portfolio variance and achieves its minimum for the sequence of true GMVP weights $\omega_t = \omega_t^*$. This first summand is relevant for assessing the performance of GMVP predictions. By contrast, the second summand in (10) measures the variance of the predicted portfolio returns, which is not relevant for GMVP performance assessment. Since both summands depend on ω_t, $\mathbb{V}(\omega_t' R_t)$ entails an undesirable trade-off between them. In particular, since the second summand does not disappear for ω_t^*, it can generate an 'evaluation bias' and therefore should not enter model forecast comparisons for the GMVP weights. This reasoning against using $\mathbb{V}(\omega_t' R_t)$ is also found in Voev (2009).

For $\mathbb{E}\left[L(b_t, R_t)\right]$, it is the parameter b_{0t} for the predicted portfolio return (i.e., the first element of b_t) which prevents evaluating solely the forecasts for the GMVP weights. Equation (8) implies that

$$\mathbb{E}[L(b_t, R_t)] = \mathbb{E}\left[\mathbb{V}_{t-1}(\omega_t' R_t)\right] + \mathbb{E}\left[\{\mathbb{E}_{t-1}(\omega_t' R_t) - b_{0t}\}^2\right].$$

This representation of $\mathbb{E}[L(b_t, R_t)]$ shows that here the trade-off is between the average predicted portfolio variance and the average squared error of the intercept b_{0t} as an approximation to the best predictor of the portfolio returns. However, in contrast to $\mathbb{V}(\omega_t' R_t)$, the trade-off-generating component disappears for the true GMVP weights $\omega_t = \omega_t^*$ and $b_{0t} = \beta_{0t}$, reflecting the consistency of the loss function L for the GMVP weights together with the expected GMVP return. Moreover, by specifying the portfolio mean parameter b_{0t} one can directly control the trade-off generating factor, which is not feasible when using $\mathbb{V}(\omega_t' R_t)$. A further important advantage of $\mathbb{E}\left[L(b_t, R_t)\right]$ is that it corresponds to a standard loss function that can be consistently estimated by the average loss over time. Therefore, it can directly be used for estimating parameters of prediction models and for comparing forecasts in terms of their average out-of-sample loss. The latter is central to pairwise (Diebold and Mariano, 1995) and multiple (Hansen et al., 2011) comparisons of predictive ability that are commonly used in econometrics.

These advantages of $\mathbb{E}\left[L(b_t, R_t)\right]$ relative to $\mathbb{V}(\omega_t' R_t)$ suggest to use the former for forecast evaluation. This being said, it turns out that both measures yield similar results in our empirical analysis (see, e.g., Table 2 in Section 4).

3   Dynamic GMVP models

In this section we describe our proposed dynamic models that specify a mapping from lagged information to the current values of the GMVP weights and mean return (see Equation 6). The first model we consider is based on an RLS approach with forgetting factor while the second one belongs to the class of GAS models. In both models, the M-estimator from Equation (7), together with the GMVP loss function, allows for convenient parameter

estimation. The asymptotic statistical properties of the M-estimator for such semiparametric models are discussed in Patton et al. (2019). Details on these properties of the M-estimator for our proposed GMVP models as well as a description of its implementation are provided in the online Appendix C.

3.1   Recursive least squares with forgetting factor

RLS with forgetting factor is a popular approach used to track the parameters in linear regression models when they are time-varying (Ljung and Söderström, 1983; Young, 2011). The RLS estimates with forgetting factor for the sequence of parameters $\{\beta_t\}$ in the auxiliary regression (2) defining the GMVP weights $\beta_{1:n-1t}$ and mean returns β_{0t} obtain from the recursion

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \frac{\Omega_{t-1}^{-1}X_t}{\lambda + X_t'\Omega_{t-1}^{-1}X_t}(Y_t - X_t'\hat{\beta}_{t-1}), \tag{11}$$

$$\Omega_t = X_tX_t' + \lambda\Omega_{t-1}. \tag{12}$$

The parameter $\lambda \in (0, 1]$ is the forgetting factor which operates as an exponential weight decreasing for more remote observations: Observations τ periods in the past have weight λ^τ in the estimate of β_t. In empirical applications λ is typically set to value slightly less than one (Raftery et al., 2010). For the special case that $\lambda = 1$, the RLS recursion in Equations (11) and (12) becomes

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \frac{[\sum_{\tau=1}^{t-1}X_\tau X_\tau']^{-1}X_t}{1 + X_t'[\sum_{\tau=1}^{t-1}X_\tau X_\tau']^{-1}X_t}(Y_t - X_t'\hat{\beta}_{t-1}),$$

which is the standard formula for updating the ordinary LS estimate $\hat{\beta}_{t-1}$ for the parameters in regression (2) computed for the observations $\{Y_\tau, X_\tau\}_{\tau=1}^{t-1}$, when a new pair of observation (Y_t, X_t) is added to the sample and the parameters are assumed to be time-invariant (Harvey, 1993, Section 4.5).

Another interpretation of RLS with forgetting factor emerges from the Kalman filter with forgetting factor for a standard linear Gaussian state-space model (Kulhavỳ and Zarrop, 1993; Raftery et al., 2010; Koop and Korobilis, 2013). When taking the auxiliary regression (2) as a measurement equation for the time-varying parameters β_t with Gaussian measurement errors ε_t and assuming that β_t follows a Gaussian random walk, then $\hat{\beta}_t$ as obtained from the RLS recursion in Equations (11) and (12) is equivalent to the expectation of β_t under its filtering distribution resulting from the Kalman filter with forgetting factor (for details see Raftery et al., 2010).

The GMVP model based on RLS with forgetting factor consists of the mapping $\beta_{t+1} = \beta(Z_t; \theta) = \hat{\beta}_t$, such that

$$\beta_{t+1} = \beta_t + \frac{\Omega_{t-1}^{-1}X_t}{\lambda + X_t'\Omega_{t-1}^{-1}X_t}(Y_t - X_t'\beta_t), \tag{13}$$

with some initial conditions (β_1, Ω_0). The forgetting factor λ is treated as an unknown parameter $(\theta = \lambda)$ to be estimated by the M-estimator in Equation (7). Note that under

8

the RLS-GMVP model as given by Equation (13) the period-$(t+1)$ value β_{t+1} is defined to be the period-t RLS estimate $\hat{\beta}_t$ based on the observations $\{Y_\tau, X_\tau\}_{\tau=1}^t$. This is akin to the prediction step of the corresponding Kalman filter with forgetting factor when β_t follows a random walk (see, e.g., Raftery et al., 2010).

In our empirical applications below, the GMVP weights have very persistent time series behavior. Hence the selection of the initial conditions β_1 and Ω_0 in the predictive recursion (13) can become critical for the out-of-sample forecast performance when the length of the estimation window T is small. A natural choice of β_1 is the OLS estimate for β in the static auxiliary regression (2) based on the data in the estimation period; for Ω_0, one can choose the corresponding sample average of $X_t X_t'$. We refer to these simple choices as RLS-ols. However, if the number of assets n is large in relation to T, the OLS estimate is known to be inaccurate, so that RLS-ols provides poor starting values. For high-dimensional applications, we therefore consider another set of initial conditions (RLS-shr) that obtain from the nonlinear shrinkage approach of Ledoit and Wolf (2012, 2015) described in Section 3.2.2.

The RLS-GMVP model as described above is specified in terms of the variables (Y_t, X_t) which, in turn, are obtained using asset n as baseline asset (see below Equation 2). As the choice of the baseline asset is arbitrary, it is desirable for a GMVP model to be invariant w.r.t. this choice, such that the results (estimation and tests) obtained under different baseline assets are mutually compatible and lead to the same predictions for the GMVP weights. In the online Appendix B we show that the RLS-GMVP model exhibits this invariance property.

3.2 Generalized Autoregressive Score models

3.2.1 Model specification

The RLS model for the GMVP in Section 3.1 includes only a single parameter (the forgetting factor λ). This makes it scalable in terms of the number of assets, but could be too restrictive. As an alternative to model the GMVP, we thus consider a GAS approach. The GAS model as introduced by Creal et al. (2013) assumes that the random variable to be modeled has a parametric conditional distribution with parameters that follow an autoregression driven by the scaled score of the log-likelihood. Following Patton et al. (2019), we adapt this parametric GAS approach to our semi-parametric framework by replacing the log-likelihood score with the score of the GMVP loss function in Equation (4).

The particular GAS recursion we consider for the mapping $\beta_{t+1} = \beta(Z_t; \theta)$ is

$$\beta_{t+1} = c + B\beta_t + AH_t^{-1}\nabla_t, \tag{14}$$

with initial conditions β_1, where $c = (c_i)$ denotes an $(n \times 1)$ vector and $A = (a_{ij})$ and $B = (b_{ij})$ are two $(n \times n)$ matrices of parameters such that $\theta = (c, B, A)$. The vector ∇_t is the score of the GMVP loss function given by

$$\nabla_t = \frac{\partial L(\beta_t, R_t)}{\partial \beta_t} = -2X_t(Y_t - X_t'\beta_t),$$

and H_t is an $(n \times n)$ matrix scaling the score. A common choice of the scaling matrix H_t in parametric GAS models is to use a measure for the predicted local curvature of the observation density (Creal et al., 2013). Accordingly, we select H_t to be related to the curvature of the loss function with respect to β_t as measured by its predicted Hessian given by

$$H_t = \mathbb{E}_{t-1}\left[\frac{\partial^2 L(\beta_t, R_t)}{\partial \beta_t \partial \beta_t'}\right] = 2\,\mathbb{E}_{t-1}\left[X_t X_t'\right].$$

For $\mathbb{E}_{t-1}\left[X_t X_t'\right]$ we use the predictions obtained from a simple exponential weighted moving average (EWMA) computed as[3]

$$\mathbb{E}_t\left[X_{t+1} X_{t+1}'\right] = \kappa \mathbb{E}_{t-1}\left[X_t X_t'\right] + (1-\kappa)X_t X_t', \tag{15}$$

with some initial value for $\mathbb{E}_0\left[X_1 X_1'\right]$. In our empirical implementation of the GAS model the smoothing parameter κ is set equal to its typical value 0.94 (Callot et al., 2017) and the initial conditions $(\beta_1, \mathbb{E}_0\left[X_1 X_1'\right])$ are selected in the same way as those for the RLS model (see Section 3.1).

A necessary condition for covariance stationarity of β_t under the GAS recursion (14) is that the roots of B lie inside the unit circle, in which case the stationary mean of β_t obtains as (Creal et al., 2013)

$$m \equiv \mathbb{E}(\beta_t) = (I_n - B)^{-1}c. \tag{16}$$

A reparametrization of the GAS recursion (14) in terms of this stationary mean will be instrumental for imposing parameter restrictions and for parameter estimation subject to targeting constraints, which we discuss in the next section.

As shown in the online Appendix B, the GAS-GMVP model as given by Equations (14)-(15), when used without any restrictions on the parameters θ, is invariant w.r.t. the choice of the baseline asset. However, certain types of restrictions on θ can compromise this invariance (see Section 3.2.2 as well as online Appendix B). Therefore, a key challenge in the GAS model is to derive restrictions on θ that are empirically useful (that is, achieving a good trade-off between parsimony and flexibility) without compromising the model's invariance properties.

3.2.2 Restricted GAS models

The vector of parameters in the unrestricted GAS-GMVP specification consists of $n + 2n^2$ parameters, making it difficult to estimate and prone to in-sample overfitting when the number of assets (n) is large. At the same time, the optimal degree of model complexity (simplicity) is primarily an empirical question. We therefore consider several restricted versions of the GAS model, taking into account only restrictions that do not affect the

---

[3]In an initial explorative analysis we also implemented the GAS model using for H_t the Fisher information $\mathbb{E}_{t-1}\left[\nabla_t \nabla_t'\right]$ (Creal et al., 2013) and the outer product $\beta_t \beta_t'$ (similar to Opschoor et al. (2018)). However, these choices did not improve upon the trade-off between numerical stability and computational speed achieved by using the Hessian predicted by the EWMA. For other common alternatives discussed in Creal et al. (2013), including $H_t = I_n$ and $H_t = \mathbb{E}_{t-1}\left[\nabla_t \nabla_t'\right]^{\nu}$, $\nu \in (0,1)$ the resulting GAS model fails to be invariant w.r.t. the choice of the baseline asset (see online Appendix B).

invariance property. For describing these restrictions we partition the vector of the scaled score $s_t = H_t^{-1}\nabla_t$ conformably with the vector of the GMVP mean return and weights $\beta_t = (\beta_{0t}, \beta'_{1:n-1t})'$ into $s_t = (s_{0t}, s'_{1:n-1t})'$, so that the system of GAS Equation (14) can be written as

$$
\begin{pmatrix} \beta_{0t+1} \\ \beta_{1:n-1t+1} \end{pmatrix} = \begin{pmatrix} c_0 \\ c_{1:n-1} \end{pmatrix} + \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & B_{11} \end{pmatrix} \begin{pmatrix} \beta_{0t} \\ \beta_{1:n-1t} \end{pmatrix} \tag{17}
$$
$$
+ \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & A_{11} \end{pmatrix} \begin{pmatrix} s_{0t} \\ s_{1:n-1t} \end{pmatrix}.
$$

First, we consider diagonal matrices A and B by imposing the following restrictions in Equation (17):

$$
b_{01} = 0, \quad b_{10} = 0, \quad B_{11} = b_{11}I_{n-1}, \quad a_{01} = 0, \quad a_{10} = 0, \quad A_{11} = a_{11}I_{n-1},
$$

where b_{11} and a_{11} are scalar parameters so that the number of parameters in A and B is reduced from $2n^2$ to 4. Under this diagonal restriction the dynamic structure of the GMVP mean return $\beta_{0,t+1}$ (directed by the parameters b_{00} and a_{00}) is allowed to differ from that of the GMVP weights $\beta_{i,t+1}, i = 1, \ldots, n-1$ (directed by the parameters b_{11} and a_{11}), but the dynamic structure for the weights is restricted to be the same across all assets. The latter is needed to ensure that the restricted diagonal model remains to be invariant w.r.t. the choice of baseline asset, which does not allow the elements in the diagonal matrices B_{11} and A_{11} to differ (see online Appendix B).

Second, we consider the restriction that the expected GMVP return $\beta_{0,t+1}$ is constant over time which obtains by using in Equation (17)

$$
b_{00} = 0, \quad b_{01} = 0, \quad a_{00} = 0, \quad a_{01} = 0,
$$

so that the number of parameters is reduced by $2n$. This restriction appears to be reasonable since portfolio returns are typically difficult to predict based on past information, at least at the daily frequency (Cochrane, 2005, Chapter 20).

Third, we consider the restriction that the long-run mean of the portfolio weights correspond to the equally weighted portfolio, so that $\mathbb{E}(\beta_{it}) = 1/n$ for $i = 1, \ldots, n-1$. Using Equation (16) and partitioning $m = \mathbb{E}(\beta_t)$ conformably with the vector of intercepts $c = (c_0, c'_{1:n-1})'$, this restriction can be represented as

$$
\begin{pmatrix} c_0 \\ c_{1:n-1} \end{pmatrix} = \left[I_n - \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & B_{11} \end{pmatrix} \right] \begin{pmatrix} m_0 \\ m_{1:n-1} \end{pmatrix}, \quad \text{with} \quad m_{1:n-1} = \iota_{n-1}/n, \tag{18}
$$

where ι_{n-1} is a $(n-1)$-dimensional vector full of ones. By this restriction, fixing $n-1$ parameters, all GMVP weights are forced to fluctuate around the benchmark value $1/n$ defined by the equally weighted portfolio. Our use of this benchmark follows DeMiguel et al. (2009); Candelon et al. (2012); Frey and Pohlmeier (2016) who consider shrinkage of the GMVP weights towards equality in a static framework. More broadly, the popularity of the equally weighted portfolio can perhaps be explained by its simplicity and the fact that it avoids estimation errors (DeMiguel et al., 2007).

Finally, we consider a reduction of the number of parameters by using a targeting estimation approach that replaces the long-run mean vector of the GMVP weights m in Equation (18) by a sample estimate. A simple estimate obtains by running the static auxiliary regression of Kempf and Memmel (2006) using OLS or, equivalently, by replacing in Equation (1) the population covariance Σ by the sample covariance matrix. However, when the number of assets n is large compared to the sample size T, the estimation of Σ by the sample covariance matrix is prone to in-sample overfitting due to the excessive number of free parameters in Σ. Hence for an increasing ratio n/T, the accuracy of the sample covariance matrix deteriorates, leading to an increasing systematic bias in the estimates for the long-run mean of the GMVP weights (Basak et al., 2009). To robustify the targeting approach against large dimensions, we thus propose to use the nonlinear shrinkage approach of Ledoit and Wolf (2012, 2015) that is suitable for large covariance matrices. This shrinkage approach operates on the eigenvalues of the covariance matrix and relies on the inversion of the Marčenko and Pastur (1967) equation for the nonlinear relationship between the population and sample eigenvalues under large-dimensional asymptotics. It pulls up the smallest sample eigenvalues which are systematically downward biased and pulls down the largest eigenvalues which are upward biased. For implementing this nonlinear shrinkage approach we use Ledoit and Wolf's (2017) numerical inversion technique for the Marčenko-Pastur equation, which is available as a Matlab function on Michael Wolf's webpage (`https://www.econ.uzh.ch/en/people/faculty/wolf.html`).

The GAS models we consider in our empirical work include the unrestricted model and models with the four restrictions described above, as well as combinations thereof. Table 1 lists the resulting model specifications: The (non-diagonal) model without any restriction (GAS), the diagonal model (d-GAS), the diagonal model with constant mean return (d-GAS-cβ_0), equal long-run weights (d-GAS-ew), constant mean return combined with equal long-run weights (d-GAS-cβ_0-ew), constant mean return combined with targeting towards the OLS estimates of the long-run weights (d-GAS-cβ_0-ols-ta) and combined with targeting towards the nonlinear shrinkage estimates of the long-run weights (d-GAS-cβ_0-shr-ta). In accordance with how its target is constructed, we use the nonlinear shrinkage estimate for β and the second moments $\mathbb{E}(X_t X_t')$ as initial conditions for the d-GAS-cβ_0-shr-ta model. For all other GAS specifications, the OLS estimate and the sample second moments are used for initialization.

4 Empirical Results

In this section we apply our proposed GMVP prediction models to historical return data, and compare them to a set of benchmark methods that are described in Section (4.1). Like our proposed models, these benchmarks aim to predict the GMVP weights in period t based on data up until period $t-1$, i.e., based on the information set $\mathcal{F}_{t-1}$. We consider both daily data (Sections 4.2 and 4.3) and monthly data (Section 4.4). In doing so, we intend to cover various practically relevant scenarios regarding the ratio n/T (number of assets divided by length of the time series).

Model	Restriction	# Params
GAS	none	$2n^2 + n$
d-GAS	$B_{11} = b_{11}I_{n-1}, \quad A_{11} = a_{11}I_{n-1},$ $b_{01} = a_{01} = 0, \quad b_{10} = a_{10} = 0$	$n + 4$
d-GAS-cβ_0	& $b_{00} = a_{00} = 0$	$n + 2$
d-GAS-ew	& $m_{1:n-1} = \iota_{n-1}/n$	5
d-GAS-cβ_0-ew	& $b_{00} = a_{00} = 0, \ m_{1:n-1} = \iota_{n-1}/n$	3
d-GAS-cβ_0-ols-ta	& $b_{00} = a_{00} = 0, \ m$: OLS estimate	2
d-GAS-cβ_0-shr-ta	& $b_{00} = a_{00} = 0, \ m$: nonlinear shrinkage estimate	2

Table 1: List of GAS model specifications according to Equations (17) and (18); n is the number of assets in the portfolio and $m = \mathbb{E}(\beta_t)$. '# Params' indicates the number of model parameters to be estimated by the M-estimator given in Equation (7).

4.1 Competing models

As alternatives to our proposed GMVP models we use several static and dynamic approaches that are motivated by their popularity in the literature:

(i) The OLS estimator constructs the period-t prediction of the GMVP weights by replacing the return covariance matrix Σ in the GMVP formula (1) with the sample covariance matrix of the returns observed up to period $t - 1$. This approach is equivalent to running Kempf and Memmel's (2006) static auxiliary regression using OLS for each period.

(ii) The linear shrinkage (SHR-l) estimator modifies the OLS estimator by estimating Σ in (1) via the linear shrinkage estimator of Ledoit and Wolf (2004). This estimator shrinks the sample covariance matrix towards the identity matrix. It minimizes the expected Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix by pushing the eigenvalues of the sample covariance matrix towards the average of all eigenvalues.

(iii) The nonlinear shrinkage (SHR-nl) estimator estimates Σ in (1) via the nonlinear shrinkage procedure of Ledoit and Wolf (2012, 2015), which we also adopt for the targeting approach in the GMVP-GAS model (see Section 3.2.2). While the linear shrinkage estimator shrinks all sample eigenvalues with the same intensity so as to minimize the expected Frobenius norm, the nonlinear shrinkage approach uses an individualized intensity for each eigenvalue.

(iv) The naïve estimator sets the prediction of the GMVP weights equal to the weights of the equally weighted portfolio.

(v) The standard dynamic conditional correlation (DCC) model proposed by Engle (2002). In this model, the conditional covariance matrix Σ_t of a return vector R_t is decomposed into

$$\Sigma_t = D_t^{1/2} C_t D_t^{1/2},$$

where C_t is the conditional correlation matrix and $D_t = \text{diag}(h_{1t}, \ldots, h_{nt})$ is a diagonal matrix with the conditional return variances, each of which follows a univariate GARCH(1,1) process. The correlation matrix C_t is given by

$$C_t = (Q_t^*)^{-1/2} Q_t (Q_t^*)^{-1/2},$$

with $Q_t = (1-\alpha-\beta)S + \alpha e_{t-1} e'_{t-1} + \beta Q_{t-1}$, where $e_t = D_t^{-1/2} R_t$ is the vector of standardized returns. The diagonal matrix Q_t^* is composed of the diagonal elements of Q_t, and S is the unconditional covariance matrix of the standardized returns. In order to implement correlation targeting, S is estimated by the sample covariance of the standardized returns e_t that are based on univariate GARCH models. We construct the period-t prediction for the GMVP weights by plugging the covariance prediction $\hat{\Sigma}_t$ (computed in period $t-1$) into the GMVP formula given in Equation (1).

(vi) The DCC model with correlation targeting based on nonlinear shrinkage estimates (DCC-nl) is developed and recommended by Engle et al. (2019) who consider it the 'new DCC standard in large dimensions'. It modifies the original DCC model by using the nonlinear shrinkage estimator of Ledoit and Wolf (2012, 2015) – instead of the sample covariance matrix – for correlation targeting.

Both DCC models are implemented using the assumption of normally distributed errors and are estimated by a composite likelihood approach as recommended by Engle et al. (2019).

4.2 In-sample results for daily data

For the experiments with daily returns we use a data set considered by Moura et al. (2019). It consists of the daily stock prices of all NYSE, AMEX and NASDAQ stocks of which we select the 200 stocks with the largest market capitalization at the last trading day of the sample period. The sample covers the period from January 2, 2002 to December 6, 2016 for a total of $T = 3,759$ trading days. Following Engle et al. (2019), for asset pairs with sample correlation larger than 0.95 we remove the respective stock with the lower market volume at the last day of the sample period in order to avoid highly similar asset pairs. The daily prices P_{it} are transformed into continuously compounded returns: $R_{it} = 100 \ln(P_{it}/P_{it-1})$.

4.2.1 Low-dimensional application

Estimating the heavily parameterized non-diagonal unrestricted GAS specification is prohibitively difficult for a large number of stocks. In an initial experiment, we thus consider a low-dimensional application of our proposed GMVP models to five of the 200 stocks. This experiment allows us to compare the seven GAS specifications in Table 1 and the RLS model and to investigate the impact of different degrees of sparsity of the GAS parameter matrices c, A and B. The five stocks we use for this experiment are those with the largest market capitalization at the last trading day of the estimation period: Apple, Microsoft, ExxonMobil, Amazon and Johnson & Johnson.

Table 2 reports the average in-sample GMVP loss for the seven GAS specifications and the two versions of the RLS model (initialized by the OLS and nonlinear shrinkage approaches).

	GAS	d-GAS	d-GAS $-c\beta_0$	d-GAS $-c\beta_0$ -ols-ta	d-GAS $-c\beta_0$ -shr-ta	d-GAS -ew	d-GAS $-c\beta_0$ -ew	RLS -ols	RLS -shr
Avg. loss	0.928	1.029	1.055	1.055	1.055	1.050	1.071	1.067	1.067
Portf. var.	0.928	1.046	1.055	1.055	1.055	1.066	1.071	1.061	1.061

Table 2: Average in-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) of the GAS and RLS models for $n = 5$ stocks. The sample period ranges from January 2, 2002 to December 6, 2016 ($T = 3,759$).

As expected, the lowest average loss is attained by the unrestricted (non-diagonal) GAS specification with 55 parameters, which allows for the largest degree of flexibility in approximating the dynamic behavior of the GMVP weights. The models with the largest average losses are the d-GAS-cβ_0-ew (constant expected portfolio returns and equal long-run means of the portfolio weights) with three parameters and the RLS with only one parameter. However, even for a dimension as low as in the present experiment ($n = 5$), NLS based parameter estimation in the unrestricted GAS model turned out to be numerically challenging[4]. Moreover, the parameter estimates for the unrestricted GAS model (not presented here) reveal that none of the model's 55 parameters is significantly different from zero at the 10% level, indicating that the model is clearly over-parameterized. With regard to out-of-sample forecast accuracy (analyzed below), these results suggest using restrictions to eliminate many of the unnecessary parameters. By contrast, the estimates of the d-GAS parameter b_{11} are highly significant and exceed 0.98 for all diagonal GAS specifications, indicating high temporal persistence of the GMVP weights. The estimates for a_{00} and b_{00} are for all specifications insignificant at conventional levels, so that there is no evidence for predictable dynamics in the conditional mean of the daily GMVP return. The estimate for the forgetting factor of the RLS model λ is statistically highly significant and its estimated value is 0.988, which implies that observations from a year ago still have a weight of 5% in the current GMVP prediction, whereas observations from two years ago have a weight of 0.2%.

In addition to these results we find that the predictions of the portfolio weights obtained under the sparsely parameterized diagonal GAS specifications and the RLS model are less noisy than those for the unrestricted GAS model, so that the former ask for less portfolio regrouping. This is illustrated in Figure 1, which shows the time series plots of the portfolio weights for the Johnson & Johnson stock predicted under the unrestricted GAS and the d-GAS-cβ_0. While the trends of the two series are very similar, the weight predictions of d-GAS-cβ_0 are less volatile than those of the unrestricted GAS. In particular, d-GAS-cβ_0 does not feature the economically questionable jumps predicted by the unrestricted GAS.

In addition to the average GMVP loss, we report in Table 2 the sample variance of the predicted GMVP returns. The comparison of the values for the two performance measures shows that they are typically close to each other. As discussed in Section 2.2, this suggests that the estimated GMVP mean returns β_{0t} are fairly stable over time, so that the observed small differences in the two measures are in line with our previous finding on the lack of

---

[4]One reason of the numerical problems appears to be the large number of local minima of the average GMVP loss function.
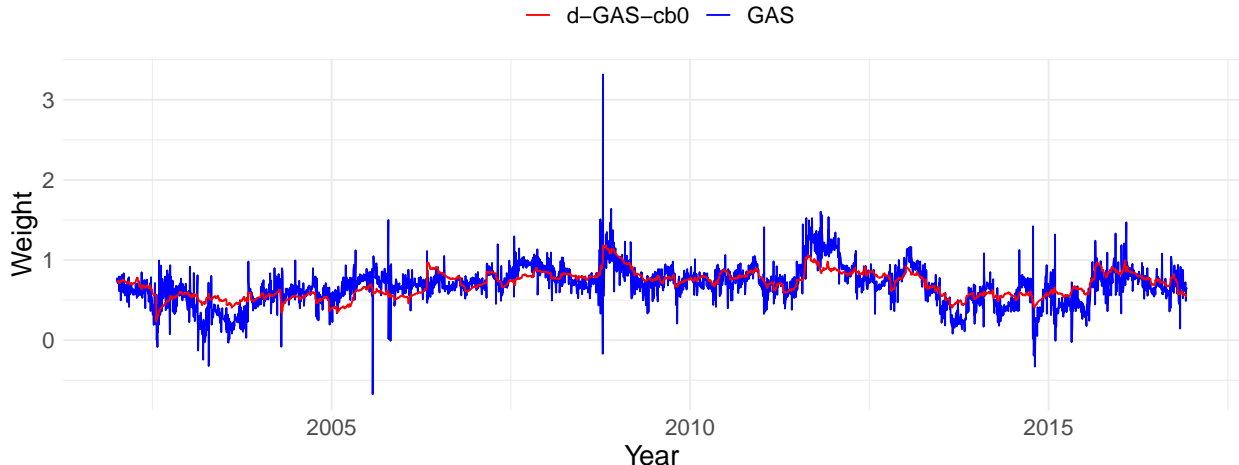
Figure 1: Estimated GMVP weights for the Johnson & Johnson stock in a portfolio of $n = 5$ assets, for the (fully unrestricted) GAS (blue line) and the d-GAS-cβ_0 model (red line).

predictability of GMVP returns.

4.2.2 High-dimensional application

Based on the results of our low-dimensional application, it seems empirically and practically reasonable to focus on the sparsely parameterized versions of our proposed GMVP models, i.e. the diagonal GAS models and the RLS model. The average in-sample loss and sample variance of the portfolio returns for those models applied to all $n = 200$ stocks in our daily data set are reported in Table 3 which also provides results for the benchmark models described in Section 4.1.

	d-GAS	d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -ew	d-GAS -cβ_0 -ew	RLS -ols	RLS -shr
Avg. loss	0.350	0.351	0.354	0.355	0.354	0.355	0.352	0.352
Portf. var.	0.350	0.351	0.354	0.355	0.355	0.355	0.352	0.352
	DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve		
Avg. loss	0.425	0.425	0.355	0.356	0.357	1.603		
Portf. var.	0.425	0.425	0.355	0.355	0.357	1.601		

Table 3: Average in-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) of the GAS and RLS models and the benchmark models for $n = 200$ stocks. The sample period ranges from January 2, 2002 to December 6, 2016 ($T = 3,759$).

Table 3 shows that the RLS model and nearly all d-GAS models attain smaller in-sample loss than the competing benchmark models. For example, the two-parameter d-GAS-cβ_0-ols-ta and d-GAS-cβ_0-shr-ta model, and even the single parameter RLS approach, result in portfolio predictions with much lower in-sample losses than the portfolios predicted by the

16

highly parameterized DCC with 602 parameters. It is only the static OLS estimator which is on par with the d-GAS-cβ_0-shr-ta and the d-GAS-cβ_0-eω model. Next, we find that the increase in the average loss resulting from imposing the cβ_0-restriction of constant expected portfolio returns in the d-GAS model is negligible and is only about 0.3%. The loss increase for the ew-restriction of equal long-run means of the portfolio weights is somewhat larger at 1%. This suggests that the stationary equilibrium differs from the naïve equally weighted portfolio which is consistent with the result that the naïve portfolio model produces by far the largest average loss among all models. It is also in line with our finding that the optimal degree of shrinkage towards the naïve portfolio for the shrinkage estimators is close to zero, so that the weight predictions and the resulting average losses for the static OLS, SHR-l and SHR-nl estimators are very close to each other.

As in the low-dimensional application, the parameter estimates of the GAS models (not presented here) indicate high persistence of the GMVP portfolio weights with estimates for b_{11} larger than 0.9 for all d-GAS specifications. The estimate for the forgetting factor λ in the RLS model is 0.999, which is also close to the value found in the low-dimensional experiment.

All in all, the results of the in-sample experiment show that all loss function based dynamic GMVP models perform broadly similar and outperform the DCC model fitted by a likelihood based estimation technique and then used within the common plug-in approach.

4.3   Out-of-sample results for daily data

We now analyze the out-of-sample forecasting performance of our GMVP models and compare it to that of the competing benchmark approaches. For this purpose, we use the same data set as for the in-sample experiments. In our out-of-sample experiments we focus on one-day-ahead forecasts obtained by re-estimating the model parameters every month on a rolling window scheme, where we follow the convention that 21 consecutive days constitute one month. The out-of-sample period starts on January 3, 2007 and ends on December 6, 2016 which results in a total of 2,501 daily point forecasts. We consider estimation window lengths of $T = 250$ (one year) and $T = 1,250$ (five years) and portfolio sizes of $n = 50$ and $n = 200$. This allows us to compare the performance of the competing models for different ratios n/T.

Following Engle et al. (2019), the n stocks included in the respective portfolio are redetermined before re-estimating the parameters each (virtual) month. The stocks are selected as follows: First, we identify the stocks that have a complete series of reported returns over the most recent T trading days and over the next 21 trading days. Then, we identify all pairs of stocks with a sample correlation larger than 0.95 over the past T days and remove the respective stock with lower trading volume observed at the time of re-estimation. Finally, we select the largest n stocks in terms of market capitalization at the re-estimation period.

Since the in-sample analysis yields virtually no evidence for temporal variation in the daily expected portfolio returns, we focus on the d-GAS specifications with the cβ_0-restriction in our out-of-sample analysis. Tables 4 and 5 provide the average out-of-sample loss and portfolio variance of these d-GAS-cβ_0 models, the RLS and the benchmark models. Table 4 reports the results for the portfolio size $n = 50$ based on the 1-year and the 5-year estimation window with a ratio n/T of 0.2 and 0.04, respectively, and Table 5 the corresponding results

$T = 1,250$ $n/T = 0.04$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS -ols	RLS -shr
	Avg. loss	0.671	0.672	0.665	0.675	0.639	**0.630**
	Portf. var.	0.671	0.672	0.665	0.675	0.638	0.630
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.654	0.647	0.669	0.667	0.665	1.574
	Portf. var.	0.654	0.647	0.669	0.667	0.665	1.574
$T = 250$ $n/T = 0.20$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS -ols	RLS -shr
	Avg. loss	1.778	0.832	0.696	0.833	0.751	**0.654**
	Portf. var.	1.764	0.829	0.693	0.830	0.748	0.651
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.737	0.673	0.755	0.729	0.677	1.584
	Portf. var.	0.736	0.672	0.752	0.729	0.677	1.585

Table 4: Average out-of-sample GMVP loss (Avg. loss) and variance of the predicted GMVP portfolio (Portf. var.) of the GAS and RLS models and the benchmark models for $n = 50$ stocks. Parameter estimation is based on a sample of length T. The out-of-sample period ranges from January 3, 2007 to December 6, 2016 ($2,501$ observations). Bold numbers indicate the smallest average GMVP loss and grey cells indicate that the model belongs to the 90% Model Confidence Set.

for $n = 200$ with value for n/T of 0.8 and 0.16. For assessing the statistical significance of differences in the average out-of-sample loss across models, we use the model confidence set (MCS) approach of Hansen et al. (2011). Based on the maximal t-statistic for the pairwise loss differentials of all models under consideration, the MCS is constructed to contain the best-performing models at a given confidence level, which we set equal to 90%. In the bootstrap implementation of the MCS, we use a block bootstrap with block length $\lfloor T_{eval}^{1/3} \rfloor$, where $T_{eval} = 2,501$ is the size of the evaluation sample, and a bootstrap sample size of $10,000$[5]. Results of a forecast comparison based on the test for Superior Predictive Ability (Hansen, 2005, SPA) are qualitatively very similar and are available in the online Appendix D.

Tables 4 and 5 show that, for all four n/T-scenarios, the two best performing specifications within our proposed GMVP approach are the d-GAS-cβ_0 model with targeting based on the nonlinear shrinkage approach (d-GAS-cβ_0-shr-ta) and RLS initialized by the nonlinear shrinkage approach (RLS-shr). We also see that their performance gains relative to RLS-ols and the other d-GAS-cβ_0 models consistently increase with the ratio n/T. This finding indicates that a precise estimator of the long-run-weights (provided by nonlinear shrinkage)

---

[5]We use the implementation of the MCS procedure in the Oxford MFE toolbox that is available on https://www.kevinsheppard.com/code/matlab/mfe-toolbox.

$T = 1,250$ $n/T = 0.16$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS -ols	RLS -shr
	Avg. loss	0.625	0.526	0.506	0.525	0.510	0.492
	Portf. var.	0.622	0.526	0.506	0.525	0.510	0.491
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.493	**0.474**	0.524	0.521	0.504	1.808
	Portf. var.	0.493	0.474	0.524	0.521	0.504	1.808
$T = 250$ $n/T = 0.80$		d-GAS -cβ_0	d-GAS -cβ_0 -ols-ta	d-GAS -cβ_0 -shr-ta	d-GAS -cβ_0 -ew	RLS -ols	RLS -shr
	Avg. loss	1.767	2.666	0.513	4.387	1.543	**0.486**
	Portf. var.	1.767	2.661	0.511	4.383	1.538	0.484
		DCC	DCC-nl	OLS	SHR-l	SHR-nl	naïve
	Avg. loss	0.984	0.510	1.575	0.864	0.495	1.813
	Portf. var.	0.984	0.510	1.571	0.864	0.495	1.813

Table 5: Same as Table 4, but for $n = 200$ stocks.

is particularly valuable when the number of assets is large relative to the sample size. The finding is also fully in line with the forecast improvements we find when moving from standard DCC with OLS targeting (DCC) to DCC with nonlinear shrinkage targeting (DCC-nl) and from static OLS to the nonlinear shrinkage estimator (SHR-nl). The comparison of all competing models shows that RLS-shr belongs to the 90% MCS for all n/T ratios and also has the smallest average loss, except in the scenario with $(n, T) = (200, 1250)$, where DCC-nl performs best. Finally, we find that the performance of the static SHR-nl estimator improves compared to the dynamic models when the sample size T decreases, which is to be expected, since the shorter the (rolling) estimation window, the easier it is for a static approach to adapt to local parameter change.

Figure 2 shows time series plots of the period-wise accumulated out-of-sample GMVP losses for the RLS-shr, d-GAS-cβ_0-shr-ta and DCC-nl in terms of their differences relative to the accumulated losses of the SHR-nl model for $n = 200$ and $T = 250$. They reveal that RLS-shr achieves its largest gains in out-of-sample fit relative to the static SHR-nl benchmark and the two dynamic alternative models during the 2007-2008 financial crises and also performance quite well during the 2011 stock market fall. Beyond these periods, however, the out-of-sample performance of all dynamic models generally appears to be worse than that of the static SHR-nl as indicated by the mostly negative trends in the differences of the accumulated losses.

4.4 Out-of-sample results for monthly data

In this section, we analyze the robustness of our previous results to changes in the sampling frequency by considering the out-of-sample forecasting performance of the GMVP models
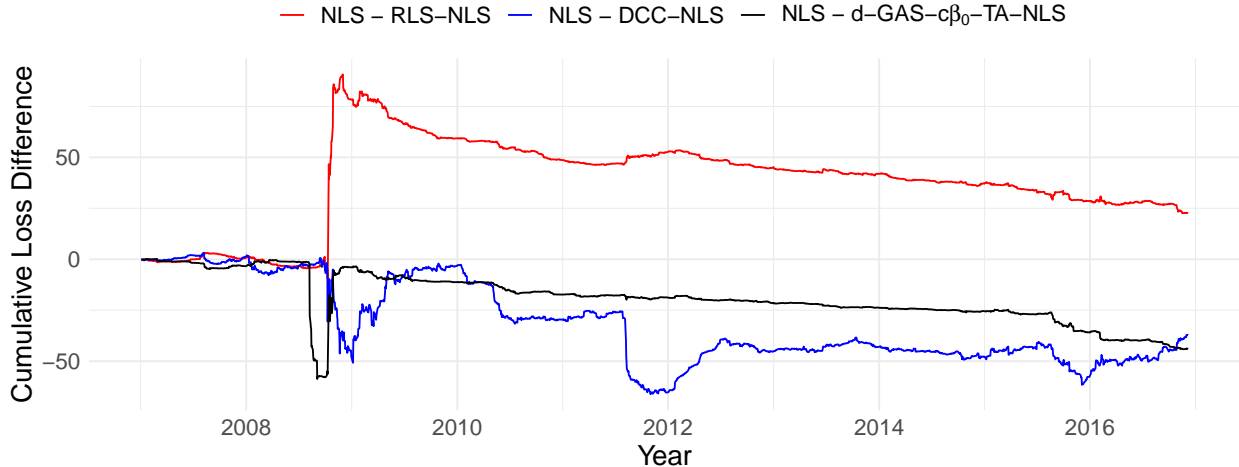
Figure 2: Difference in accumulated out-of-sample GMVP loss for SHR-nl versus RLS-shr (red line), SHR-nl versus DCC-nl (blue line) and SHR-nl versus d-GAS-cβ_0-shr-ta (black line) for $n = 200$ and an estimation window of length $T = 250$. A positive loss difference indicates that the other model outperforms SHR-nl.

when applied to monthly asset returns. For this experiment we rely upon four data sets used by DeMiguel et al. (2009), which we have extended to cover the period from July 1963 to February 2019. They consist of returns for Fama-French portfolios sorted by size and book-to-market ratio and industry portfolios representing the U.S. stock market: The first data set contains six Fama-French portfolios (6-FF), the second one 25 Fama-French portfolios (25-FF), the third one 10 industry portfolios (10-Ind), and the last one 48 industry portfolios (48-Ind)[6]. As in our application to daily returns, we consider one-period-ahead forecasts and use a rolling window for parameter estimation. The window length is set equal to $T = 120$ (10 years) and the out-of-sample period ranges from July 1973 to February 2019, which yields a total of 548 monthly forecasts for each data set.

Table 6 contains the resulting average out-of-sample GMVP loss of our proposed models and the competing alternatives. They show that for all four portfolios, the best performing model in the class of RLS and d-GAS models attains smaller average loss than both versions of the DCC approach. For the 6-FF, 25-FF and 10-Ind portfolios, for which the n/T-ratio is at most 0.21, the static SHR-l estimator attains the smallest loss. However, the RLS-shr model is always contained in the 90% MCS, indicating that the loss difference between RLS-shr and SHR-l is not statistically significant. As explained above, one possible explanation for the relatively good performance of this static SHR estimator is that the estimation window in this experiment is rather short ($T = 120$), much shorter than in our previous experiments with daily data in Section 4.3. For the 48-Ind portfolio for which we have the most challenging scenario in terms of the n/T-ratio, the best model is the RLS-shr with a performance which is significantly better than that for all competing models. Furthermore, we observe that RLS-shr belongs to the 90% MCS for all four portfolios, which is fully in

---

[6]The data has been obtained from `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

	6-FF	25-FF	10-Ind	48-Ind
n/T	0.05	0.21	0.08	0.49
d-GAS-cβ_0	17.14	19.81	14.03	21.35
d-GAS-cβ_0-ols-ta	16.41	15.90	13.69	19.79
d-GAS-cβ_0-shr-ta	16.54	13.80	13.22	13.96
d-GAS-cβ_0-ew	16.06	15.75	13.66	20.49
RLS-ols	17.60	13.90	14.04	10.10
RLS-shr	17.60	13.67	14.04	**10.07**
DCC	17.22	16.99	13.77	16.35
DCC-nl	17.16	16.16	13.45	13.22
OLS	15.72	14.83	12.86	17.30
SHR-l	**15.54**	**13.27**	**12.63**	14.84
SHR-nl	15.72	13.63	12.79	13.11
naïve	24.49	26.48	18.35	23.79

Table 6: Average out-of-sample GMVP loss of the diagonal GAS models, the RLS and the benchmark models for monthly portfolio returns. The length of the estimation window is 10 years ($T = 120$), the out-of-sample forecasting period ranges from July 1973 to February 2019 (548 observations). Bold numbers indicate the smallest average GMVP loss and grey shaded cells indicate that the model belongs to the 90% Model Confidence Set.

line with the results for the experiments based on daily data. These results regarding the forecast performance are corroborated by the results of the SPA test, which are provided in the online Appendix D.

A visual inspection of the period-wise accumulated out-of-sample GMVP losses for the 48-Ind portfolio (see Figure 3) again indicates gains in the out-of-sample fit of RLS-shr relative to the static SHR-nl benchmark and the two other dynamic models in the 2007-2008 financial crisis. In contrast to the daily results, however, the out-of-sample performance of RLS-shr appears to be uniformly better than that of the static SHR-nl, as shown by the overall positive trend in the accumulated loss differences also in tranquil periods.

5   Conclusion

In this paper, we propose to use the Kempf and Memmel (2006) loss function in order to model the weights of the global minimum variance portfolio (GMVP). We provide a new theoretical result justifying the use of the loss function, and propose tractable time series models for the portfolio weights. While all of these models make extensive use of the Kempf and Memmel (2006) loss function, we consider two types of dynamic specifications: Recursive least squares (RLS) with forgetting factor, as well as generalized autoregressive score (GAS). All of our proposed specifications are invariant with respect to the choice of baseline asset, which is important both conceptually and practically.
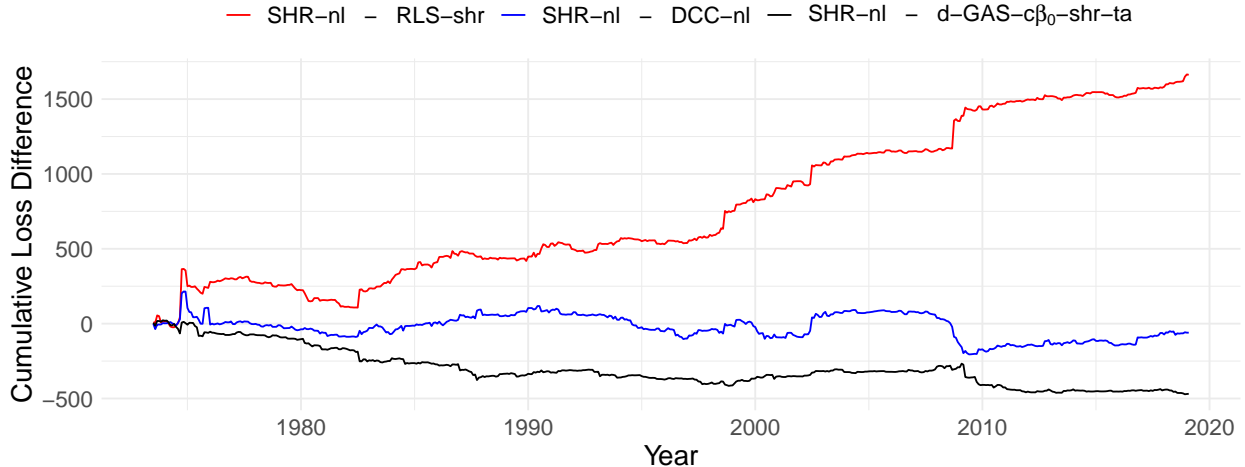
Figure 3: Difference in accumulated out-of-sample GMVP loss for SHR-nl versus RLS-shr (red line), SHR-nl versus DCC-nl (blue line) and SHR-nl versus d-GAS-cβ_0-shr-ta (black line) for the 48-Ind portfolio with $n = 48$ and an estimation window of length $T = 120$. A positive loss difference indicates that the other model outperforms SHR-nl.

In order to use our GMVP models for out-of-sample portfolio weight forecasting in large dimensions we propose to regularize them using nonlinear shrinkage. We use the latter technique for long-run mean targeting (when applied to GAS) and for initializing the autoregressive model recursions (when applied to RLS). In out-of-sample forecast experiments, we find that RLS combined with nonlinear shrinkage (RLS-shr) performs particularly well across a broad range of empirical settings, covering daily and monthly data as well as various ratios of n (number of assets) over T (length of estimation sample). Unlike any other model we consider, the RLS-shr approach is contained in the 90% Model Confidence Set (Hansen et al., 2011) in all of the experiments. This finding indicates that the approach manages to adjust the predicted portfolio weights (in particular, the amount of shrinkage and the forgetting factor governing the weights' dynamic properties) in a way that matches the data at hand.

Acknowledgements

References

Basak, G. K., Jagannathan, R., and Ma, T. (2009). Jackknife estimator for tracking error variance of optimal portfolios. *Management Science*, 55:990–1002.

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Review of Financial Studies*, 4:315–342.

Brandt, M. W. (2009). Portfolio choice problems. In Ait-Sahalia, Y. and Hansen, L., editors, *Handbook of Financial Econometrics*, pages 269–336. Elsevier.

Callot, L., Caner, M., Önder, A. Ö., and Ulaşan, E. (2019). A nodewise regression approach to estimating large portfolios. *Journal of Business & Economic Statistics*. Forthcoming.

Callot, L. A., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32:140–158.

Candelon, B., Hurlin, C., and Tokpavi, S. (2012). Sampling error and double shrinkage estimation of minimum variance portfolios. *Journal of Empirical Finance*, 19:511–527.

Christoffersen, P. and Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72:291–318.

Clements, A., Scott, A., and Silvennoinen, A. (2015). On the benefits of equicorrelation for portfolio allocation. *Journal of Forecasting*, 34:507–522.

Cochrane, J. H. (2005). *Asset Pricing: Revised Edition*. Princeton University Press.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28:777–795.

DeMiguel, V., Garlappi, L., Nogales, F. J., and Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 5:798–812.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22:1915–1953.

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

Elliott, G., Ghanem, D., and Krüger, F. (2016). Forecasting conditional probabilities of binary outcomes under misspecification. *Review of Economics and Statistics*, 98:742–755.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20:339–350.

Engle, R. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, 24:238–253.

Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30:212–228.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37:363–375.

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

Frahm, G. and Memmel, C. (2010). Dominating estimators for minimum-variance portfolios. *Journal of Econometrics*, 159:289–302.

Frey, C. and Pohlmeier, W. (2016). Bayesian shrinkage of portfolio weights. Working paper, University of Konstanz.

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

Hand, D. J. and Vinciotti, V. (2003). Local versus global models for classification problems: Fitting models where it matters. *The American Statistician*, 57:124–131.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23:365–380.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79:453–497.

Harvey, A. C. (1993). *Time Series Models*. Pearson.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58:1651–1683.

Kempf, A. and Memmel, C. (2006). Estimating the global minimum variance portfolio. *Schmalenbach Business Review*, 58:332–348.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177:185–198.

Kulhavỳ, R. and Zarrop, M. B. (1993). On a general concept of forgetting. *International Journal of Control*, 58:905–924.

Ledoit, O., Santa-Clara, P., and Wolf, M. (2003). Flexible multivariate garch modeling with an application to international stock markets. *Review of Economics and Statistics*, 85:735–747.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10:603–621.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40:1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.

Ledoit, O. and Wolf, M. (2017). Numerical implementation of the quest function. *Computational Statistics & Data Analysis*, 115:199–223.

Ljung, L. and Söderström, T. (1983). *Theory and Practice of Recursive Identification*. MIT Press.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457–483.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7:77–91.

Moura, G. V., Santos, A. A., and Ruiz, E. (2019). Comparing high dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking and Finance*. Forthcoming.

Opschoor, A., Janus, P., Lucas, A., and Van Dijk, D. (2018). New heavy models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 36:643–657.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics*, 211:388–413.

Raftery, A. E., Kárný, M., and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52:52–66.

Voev, V. (2009). On the economic evaluation of volatility forecasts. Working Paper, CREATES (Research Paper 2009-56).

Weiss, A. A. (1996). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics*, 11:539–560.

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

Young, P. C. (2011). *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Springer.

Predicting the Global Minimum Variance Portfolio[*]

Online Appendix

Laura Reh[a,b]   Fabian Krüger[c]   Roman Liesenfeld[b]

July 5, 2020

A Elicitability of GMVP weights

We begin with stating an auxiliary result. While the result is straightforward, we are not aware of a reference and thus provide a proof.

Lemma 1. *A mixture of two n-variate distributions with mean vectors μ_a, μ_b and covariance matrices Σ_a, Σ_b has precision matrix*

$$\Sigma_\pi^{-1} = \frac{\pi(1-\pi)}{1+\pi(1-\pi)d'V_\pi^{-1}d}\, dd'V_\pi^{-1},$$

where $\pi \in [0,1]$ is the mixture probability for the first component, and

$$\begin{aligned} d &= (\mu_a - \mu_b),\\ V_\pi &= \pi\Sigma_a + (1-\pi)\Sigma_b. \end{aligned}$$

Proof. The covariance matrix of the mixture is given by

$$\Sigma_\pi = V_\pi + (\mu_a, \mu_b)\begin{pmatrix} \pi(1-\pi) & -\pi(1-\pi) \\ -\pi(1-\pi) & \pi(1-\pi) \end{pmatrix}\begin{pmatrix} \mu_a' \\ \mu_b' \end{pmatrix}.$$

The result then follows from a variant of the Woodbury matrix identity (Petersen and Petersen, 2012, Equation 159). $\square$

To discuss elicitability, we next introduce some notation. Let $\mathcal{G}$ denote the family of n-variate continuous distributions with finite mean vector and covariance matrix. For a typical member G of this family, we denote the covariance matrix associated with G by $\Sigma(G)$. The $(n-1) \times 1$ vector $\beta_{1:n-1}(G)$ contains the associated GMVP weights for the first $n-1$ assets (the remaining weight is implied by the constraint that the weights sum to unity). Moreover, the $n \times 1$ vector $\beta(G)$ is given by $(\beta_0(G), \beta_{1:n-1}(G)')'$, where the first element is the expected GMVP return implied by G.

Proposition 1. *$\beta_{1:n-1}$ does not have convex level sets. That is, a convex combination G_π of two distributions $G_a, G_b \in \mathcal{G}$ such that $\beta_{1:n-1}(G_a) = \beta_{1:n-1}(G_b) = b_{1:n-1}$ generally has $\beta_{1:n-1}(G_\pi) \neq b_{1:n-1}$.*

Proof. Follows from Lemma 1. As a simple example, the violation can be checked for the case $n = 2$, $\Sigma(G_a) = \Sigma(G_b) = I_2$, $\mu(G_a) = (1,1)'$ and $\mu(G_b) = (2,2)'$. $\square$

Since $\beta_{1:n-1}$ does not have convex level sets, it can not be elicitable. The necessity of convex level sets for elicitablity is formally stated for the case of a univariate predictand in Gneiting (2011, Theorem 6). However, the condition is also necessary in the present case of an $n-1$ variate predictand, c.f. the proof of Lemma 1 in Lambert et al. (2008) as well as the discussion by Fissler and Ziegel (2019, p. 1170).

Since β is elicitable, it must have convex level sets. For example, consider two distributions $G_a, G_b \in \mathcal{G}$ such that $\Sigma(G_a) \propto \Sigma(G_b)$ and $\beta_0(G_a) = \beta_0(G_b)$. This setup implies that $\beta(G_a) = \beta(G_b)$. Using Lemma 1 and the fact that

$$\big(\mu(G_a) - \mu(G_b)\big)'\Sigma(G_a)^{-1}\iota = 0,$$

it can be shown that $\Sigma(G_\pi)^{-1}\iota \propto \Sigma(G_a)^{-1}\iota$ and hence $\beta(G_\pi) = \beta(G_a) = \beta(G_b)$, which illustrates that β has convex level sets.

B Invariance of the GMVP models

B.1 Preliminaries

In order discuss the invariance of the dynamic GMVP models based on the RLS recursion (in Section 3.1) and the GAS specifications (in Section 3.2), we begin with stating the relationship between the auxiliary regressions identifying the GMVP weights for different baseline assets.

When asset n is used as baseline asset, then the GMVP auxiliary regression (as reproduced from Section 2.1) is

$$Y_t = X_t'\beta_t + \varepsilon_t, \tag{A-1}$$
$$Y_t = R_{nt}, \qquad X_t' = (1, R_{nt} - R_{1t}, \ldots, R_{nt} - R_{n-1t}),$$

with expected GMVP return β_{0t} and GMVP weights

$$w_{it}^* = \begin{cases} \beta_{it} & i = 1, \ldots, n-1, \\ 1 - \sum_{j=1}^{n-1} \beta_{jt} & i = n. \end{cases}$$

Suppose we select a different baseline asset, say asset k instead of asset n. Then the corresponding variables in the GMVP regression of the form (A-1) are given by

$$\tilde{Y}_t = R_{kt}, \qquad \tilde{X}_t' = (1, R_{kt} - R_{1t}, \ldots, R_{kt} - R_{k-1t}, R_{kt} - R_{nt},$$
$$R_{kt} - R_{k+1t}, \ldots, R_{kt} - R_{n-1t}).$$

These regression variables associated with baseline asset k obtain from those for baseline asset n according to the one-to-one transformation

$$\tilde{Y}_t = d_k' X_t + Y_t, \qquad \tilde{X}_t = R_k X_t, \tag{A-2}$$

where

$$d_k' = (0, -e_k'), \qquad R_k = \begin{pmatrix} 1 & 0_{n-1}' \\ 0_{n-1} & S_k \end{pmatrix},$$

$$S_k = \begin{pmatrix} I_{k-1} & -\iota_{k-1} & 0_{k-1 \times n-k-1} \\ 0_{k-1}' & -1 & 0_{n-k-1}' \\ 0_{n-k-1 \times k-1} & -\iota_{n-k-1} & I_{n-k-1} \end{pmatrix}.$$

Here we have used ι_ℓ to denote the ℓ-dimensional vector full of ones, 0_ℓ to denote the ℓ-dimensional Null vector, $0_{\ell_1 \times \ell_2}$ to denote the $(\ell_1 \times \ell_2)$-dimensional Null matrix and e_k to denote the k'th column of the $(n-1)$-dimensional identity matrix I_{n-1}. For the permutation matrix R_k it holds that

$$R_k = R_k^{-1}, \qquad R_k' d_k = -d_k. \tag{A-3}$$

Using the equalities $Y_t = \tilde{Y}_t + d_k' \tilde{X}_t$ and $X_t = R_k \tilde{X}_t$ resulting from Equations (A-2) and (A-3) in the GMVP regression (A-1) for baseline asset n, yields the following equivalent GMVP regression for baseline asset k:

$$\tilde{Y}_t = \tilde{X}_t' \tilde{\beta}_t + \varepsilon_t, \tag{A-4}$$

3

with

$$\tilde{\beta}_t = R'_k \beta_t - d_k, \tag{A-5}$$

$$w^*_{it} = \begin{cases} \tilde{\beta}_{it} & i = 1, \dots, k-1, k+1, \dots, n \\ 1 - \sum_{j=1}^{n-1} \tilde{\beta}_{jt} & i = k, \end{cases}$$

and expected portfolio return $\tilde{\beta}_{0t}$. An immediate implication of Equation (A-5) is that the coefficients β_t in the GMVP regression used to identify the GMVP weights under baseline asset n are in one-to-one correspondence to the coefficients $\tilde{\beta}_t$ in the GMVP regression associated with baseline asset k equivalently identifying the GMVP weights. Moreover, Equation (A-4) shows that the error terms of the GMVP regressions for both baseline assets are the same. Whence the value of the loss function (4) computed for (Y_t, X_t, β_t) is the same as that computed for $(\tilde{Y}_t, \tilde{X}_t, \tilde{\beta}_t)$.

With the one-to-one mapping between (Y_t, X_t, β_t) and $(\tilde{Y}_t, \tilde{X}_t, \tilde{\beta}_t)$ as given by Equations (A-2) and (A-5) we can now discuss the invariance of our proposed dynamic GMVP models with respect to the selection of the baseline asset. For the invariance of a dynamic GMVP model, it is necessary, first, that there exists a parametrization (including initial conditions) for the model associated with baseline asset k which leads to the same predictions for the GMVP weights as the model associated with baseline asset n which requires according to Equation (A-5) that the predictions satisfy $\tilde{\beta}_{t+1} = R'_k \beta_{t+1} - d_k \ \forall t$ and, second, that this parametrization for baseline asset k is in one-to-one correspondence with that for baseline asset n. Since for $\tilde{\beta}_t = R'_k \beta_t - d_k$ the loss function for (Y_t, X_t, β_t) is the same as for $(\tilde{Y}_t, \tilde{X}_t, \tilde{\beta}_t)$ it follows that under those two conditions the M-estimator as defined in Equation (7) is invariant w.r.t. the choice of the baseline asset and leads to the same estimates for the predicted GMVP weights.

B.2 Invariance of the RLS model

In Lemma 2 we provide the parametrization for the RLS model of Section 3.1 for baseline asset k which leads to the same GMVP predictions as the RLS model for baseline asset n.

Lemma 2. *Consider the RLS model for baseline asset n given by*

$$\beta_{t+1} = \beta_t + \frac{\Omega_{t-1}^{-1}}{\lambda + X'_t \Omega_{t-1}^{-1} X_t} X_t (Y_t - X'_t \beta_t), \tag{A-6}$$

$$\Omega_t = X_t X'_t + \lambda \Omega_{t-1}, \tag{A-7}$$

with initial condition β_1 and Ω_0. This RLS model is equivalent to the following RLS model for baseline asset k:

$$\tilde{\beta}_{t+1} = \tilde{\beta}_t + \frac{\tilde{\Omega}_{t-1}^{-1}}{\tilde{\lambda} + \tilde{X}'_t \tilde{\Omega}_{t-1}^{-1} \tilde{X}_t} \tilde{X}_t (\tilde{Y}_t - \tilde{X}'_t \tilde{\beta}_t), \tag{A-8}$$

$$\tilde{\Omega}_t = \tilde{X}_t \tilde{X}'_t + \tilde{\lambda} \tilde{\Omega}_{t-1}, \tag{A-9}$$

4

with

$$
\begin{align}
\tilde{\lambda} &= \lambda, &\text{(A-10)}\\
\tilde{\beta}_1 &= R_k'\beta_1 - d_k, &\text{(A-11)}\\
\tilde{\Omega}_0 &= R_k\Omega_0 R_k'. &\text{(A-12)}
\end{align}
$$

Specifically, for the β_t's obtained according to Equations (A-6)-(A-7) and the $\tilde{\beta}_t$'s obtained according to Equations (A-8)-(A-12) it holds that $\tilde{\beta}_t = R_k'\beta_t - d_k$ for all $t = 2, 3, \ldots$, as desired.

Proof. From (A-9) with (A-10) and (A-12) it follows that

$$
R_k\tilde{\Omega}_t R_k' = \Omega_t, \quad t = 0, 1, 2, \ldots. \tag{A-13}
$$

Using (A-2), (A-10) and (A-13) together with $\tilde{\beta}_t = R_k'\beta_t - d_k$ on the r.h.s. of Equation (A-8) shows that if $\tilde{\beta}_t = R_k'\beta_t - d_k$ and β_{t+1} and $\tilde{\beta}_{t+1}$ are generated according to Equations (A-6) and (A-8), then it holds that $\tilde{\beta}_{t+1} = R_k'\beta_{t+1} - d_k$. This combined with the initial condition $\tilde{\beta}_1 = R_k'\beta_1 - d_k$ in (A-21) completes the proof. □

The implication of Equations (A-10)-(A-12) is that the parameter and initial conditions $(\lambda, \beta_1, \Omega_0)$ associated with baseline asset n are in one-to-one correspondence with the parameter and initial conditions $(\tilde{\lambda}, \tilde{\beta}_1, \tilde{\Omega}_0)$ associated with baseline asset k. Whence the RLS model is invariant w.r.t. the choice of the baseline asset.

B.3 Invariance of the GAS model

Lemma 3 provides the parametrization for the GAS model of Section 3.2 for baseline asset k which leads to the same GVMP predictions as the GAS model for baseline asset n.

Lemma 3. *Consider the GAS model in Equations (14)-(15) for baseline asset n written as*

$$
\begin{align}
\beta_{t+1} &= c + B\beta_t - A(\mathbb{E}_{t-1}[X_t X_t'])^{-1} X_t (Y_t - X_t'\beta_t), &\text{(A-14)}\\
\mathbb{E}_t[X_{t+1} X_{t+1}'] &= \kappa\mathbb{E}_{t-1}[X_t X_t'] + (1 - \kappa) X_t X_t', &\text{(A-15)}
\end{align}
$$

with initial condition β_1 and $\mathbb{E}_0[X_1 X_1']$. This GAS model is equivalent to the following GAS model for baseline asset k:

$$
\begin{align}
\tilde{\beta}_{t+1} &= \tilde{c} + \tilde{B}\tilde{\beta}_t - \tilde{A}(\mathbb{E}_{t-1}[\tilde{X}_t \tilde{X}_t'])^{-1} \tilde{X}_t (\tilde{Y}_t - \tilde{X}_t'\tilde{\beta}_t), &\text{(A-16)}\\
\mathbb{E}_t[\tilde{X}_{t+1} \tilde{X}_{t+1}'] &= \kappa\mathbb{E}_{t-1}[\tilde{X}_t \tilde{X}_t'] + (1 - \kappa) \tilde{X}_t \tilde{X}_t', &\text{(A-17)}
\end{align}
$$

with

$$
\begin{align}
\tilde{c} &= R_k'(c + d_k - Bd_k), &\text{(A-18)}\\
\tilde{B} &= R_k'BR_k', &\text{(A-19)}\\
\tilde{A} &= R_k'AR_k', &\text{(A-20)}\\
\tilde{\beta}_1 &= R_k'\beta_1 - d_k, &\text{(A-21)}\\
\mathbb{E}_0[\tilde{X}_1 \tilde{X}_1'] &= R_k\mathbb{E}_0[X_1 X_1']R_k'. &\text{(A-22)}
\end{align}
$$

Specifically, for the β_t's obtained according to Equations (A-14)-(A-15) and the $\tilde{\beta}_t$'s obtained according to Equations (A-16)-(A-22) it holds that $\tilde{\beta}_t = R'_k \beta_t - d_k$ for all $t = 2, 3, \ldots$, as desired.

Proof. From Equations (A-17) and (A-22) it follows that

$$R_k \mathbb{E}_t[\tilde{X}_{t+1}\tilde{X}'_{t+1}]R'_k = \mathbb{E}_t[X_{t+1}X'_{t+1}], \quad t = 0, 1, 2, \ldots. \tag{A-23}$$

Using (A-2), (A-18)-(A-19) and (A-23) together with $\tilde{\beta}_t = R'_k \beta_t - d_k$ on the r.h.s. of Equation (A-16) shows that if $\tilde{\beta}_t = R'_k \beta_t - d_k$ and β_{t+1} and $\tilde{\beta}_{t+1}$ are generated according to Equations (A-14) and (A-16), then it holds that $\tilde{\beta}_{t+1} = R'_k \beta_{t+1} - d_k$. This combined with the initial condition $\tilde{\beta}_1 = R'_k \beta_1 - d_k$ in (A-21) completes the proof. $\qquad\square$

Equations (A-18)-(A-22) show that the parameters $(c, B, A, \beta_1, \mathbb{E}_0[X_1 X'_1])$ associated with baseline asset n are in one-to-one correspondence to the parameters $(\tilde{c}, \tilde{B}, \tilde{A}, \tilde{\beta}_1, \mathbb{E}_0[\tilde{X}_1 \tilde{X}'_1])$ associated with baseline asset k. This implies that the GAS model without any restrictions on its parameters (c, B, A) is invariant w.r.t. the choice of the baseline asset.

However, restrictions on the GAS parameters can, depending on the form of the restriction, violate the necessary one-to-one correspondence of the parameterizations. This is the case for the natural restriction to assume that the matrices $A = (a_{ij})$ and/or $B = (b_{ij})$ are diagonal, say

$$A = \text{diag}(a_{00}, a_{11}, \ldots, a_{n-1n-1}), \quad B = \text{diag}(b_{00}, b_{11}, \ldots, b_{n-1n-1}).$$

Then it is easy to verify by using (A-19) and (A-20) that $\tilde{A}$ and $\tilde{B}$ are in contrast to A and B not diagonal, unless we restrict the diagonal elements such that

$$A = \text{diag}(a_{00}, a_{11}, \ldots, a_{11}), \quad B = \text{diag}(b_{00}, b_{11}, \ldots, b_{11}), \tag{A-24}$$

or

$$A = \text{diag}(a_{00}, a_{00}, \ldots, a_{00}), \quad B = \text{diag}(b_{00}, b_{00}, \ldots, b_{00}).$$

The diagonal GAS models we use in our empirical application are based on the restriction given in Equation (A-24) and are therefore invariant w.r.t. choice of the baseline asset (see Section 3.2.2).

C   Statistical inference based on the M-estimator

For estimating the parameters θ of our proposed dynamic GMVP models specifying a function for $\beta_t = \beta(\theta, Z_{t-1}) = \beta_t(\theta)$ we use the NLS estimator as given in Equation (7), repeated here for convenience:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} L\left(\beta_t(\theta), R_t\right), \quad L\left(\beta_t(\theta), R_t\right) = [Y_t - X'_t \beta_t(\theta)]^2. \tag{A-25}$$

The nonlinear function $\beta_t(\theta)$ for the RLS-GMVP model is given in Equation (13) and that for the GAS-GMVP model in Equation (14). For solving the minimization problem in Equation (A-25) we use the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm based on analytical gradients which are given in Section C.2 below.

C.1 Asymptotic distribution

Under standard regularity conditions for M-estimators such as those of Amemiya (1985, Theorems 4.1.1 and 4.1.3), the NLS estimator $\hat{\theta}$ in Equation (A-25) is consistent and asymptotically normal with

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, D^{-1}WD^{-1}), \tag{A-26}$$

where

$$D = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E} \left[\frac{\partial g_t(\theta)}{\partial \theta'} \right], \quad W = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathbb{E} \left[g_t(\theta) g_t(\theta)' \right], \tag{A-27}$$

and $g_t(\theta) = \partial L(\beta_t(\theta), R_t)/\partial\theta$ is the gradient of the loss function[1].

The asymptotic covariance matrix of the M-estimator as given by Equations (A-26)-(A-27) comprises the expected outer-product of the gradient of the loss function $\mathbb{E}[g_t(\theta)g_t(\theta)']$ and the expected Hessian $\mathbb{E}[\partial g_t(\theta)/\partial\theta']$.

The gradient of the GMVP loss function is given by

$$g_t(\theta) = \frac{\partial L(\beta_t(\theta), R_t)}{\partial \theta} = \frac{\partial \beta_t(\theta)'}{\partial \theta} \frac{\partial L(\beta_t(\theta), R_t)}{\partial \beta_t},$$

with

$$\frac{\partial L(\beta_t(\theta), R_t)}{\partial \beta_t} = \nabla_t = -2X_t[Y_t - X_t'\beta_t(\theta)],$$

so that its expected outer product obtains as

$$\mathbb{E}\left[g_t(\theta)g_t(\theta)'\right] = 4\mathbb{E}\left([Y_t - X_t'\beta_t(\theta)]^2 \ \frac{\partial \beta_t(\theta)'}{\partial \theta} X_t X_t' \frac{\partial \beta_t(\theta)}{\partial \theta'} \right).$$

The Hessian of the GMVP loss function is given by

$$\frac{\partial g_t(\theta)}{\partial \theta'} = \frac{\partial \beta_t(\theta)'}{\partial \theta} \frac{\partial \nabla_t}{\partial \theta'} + (\nabla_t' \otimes I_m) \frac{\partial \text{vec}\left[\partial \beta_t(\theta)'/\partial\theta\right]}{\partial \theta'},$$

where $\text{vec}(\cdot)$ denotes the operator which stacks the columns of a matrix into a vector, and $\otimes$ is the Kronecker matrix product. It follows that

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial g_t(\theta)}{\partial \theta'}\right] &= \mathbb{E}\left(\mathbb{E}_{t-1} \left[\frac{\partial \beta_t(\theta)'}{\partial \theta} \frac{\partial \nabla_t}{\partial \theta'} + (\nabla_t' \otimes I_m) \frac{\partial \text{vec}\left[\partial \beta_t(\theta)'/\partial\theta\right]}{\partial \theta'} \right] \right) \\
&= \mathbb{E}\left(\frac{\partial \beta_t(\theta)'}{\partial \theta} \mathbb{E}_{t-1} \left[\frac{\partial \nabla_t}{\partial \theta'} \right] \right) \\
&= 2\mathbb{E}\left(\frac{\partial \beta_t(\theta)'}{\partial \theta} X_t X_t' \frac{\partial \beta_t(\theta)}{\partial \theta'} \right).
\end{aligned}
$$

---

[1]Patton et al. (2019) provide specific sufficient conditions for the consistency and asymptotic normality (of the form as in Equation A-26) for an M-estimator similar to ours. As in their work, our estimator minimizes a consistent loss function for dynamically evolving parameters (β_t), which are given as functions that are measurable in $\mathcal{F}_{t-1}$ and continuous in the static parameters (θ).

Assuming that

$$T^{-1} \sum_{t=1}^{T} g_t(\theta) g_t(\theta)' - T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[g_t(\theta) g_t(\theta)'\right] \xrightarrow{p} 0,$$

$$T^{-1} \sum_{t=1}^{T} \frac{\partial g_t(\theta)}{\partial \theta'} - T^{-1} \sum_{t=1}^{T} \mathbb{E}\left[\frac{\partial g_t(\theta)}{\partial \theta'}\right] \xrightarrow{p} 0,$$

the components of the asymptotic covariance matrix of the M-estimator W and D in Equations (A-26) and (A-27) can be consistently estimated by

$$\hat{W} = T^{-1} \sum_{t=1}^{T} g_t(\hat{\theta}) g_t(\hat{\theta})'$$

$$= T^{-1} \sum_{t=1}^{T} 4 \left[Y_t - X_t' \beta_t(\hat{\theta})\right]^2 \frac{\partial \beta_t(\hat{\theta})'}{\partial \theta} X_t X_t' \frac{\partial \beta_t(\hat{\theta})}{\partial \theta'},$$

and

$$\hat{D} = T^{-1} \sum_{t=1}^{T} \frac{\partial g_t(\theta)}{\partial \theta'}$$

$$= T^{-1} \sum_{t=1}^{T} 2 \frac{\partial \beta_t(\hat{\theta})'}{\partial \theta} X_t X_t' \frac{\partial \beta_t(\hat{\theta})}{\partial \theta'}.$$

These estimates require to compute the gradients $\partial \beta_t(\hat{\theta})/\partial \theta$ for which we use their analytical form.

C.2 The gradients $\partial \beta_t(\theta)/\partial \theta$ for the RLS and GAS model

For the RLS model

$$\beta_t = \beta_{t-1} + \frac{\Omega_{t-2}^{-1} X_{t-1}}{\lambda + X_{t-1}' \Omega_{t-2}^{-1} X_{t-1}} (Y_{t-1} - X_{t-1}' \beta_{t-1}),$$

with $\theta = \lambda$ the first derivative of $\beta_t(\lambda)$ obtains recursively as

$$
\begin{aligned}
\frac{\partial \beta_t(\lambda)'}{\partial \lambda} &= \frac{\partial \beta_{t-1}(\lambda)'}{\partial \lambda} \\
&\quad + \left\{ (\Psi_{t-2} X_{t-1})'(\lambda + X_{t-1}' \Omega_{t-2}^{-1} X_{t-1}) \right. \\
&\qquad \left. - (\Omega_{t-2}^{-1} X_{t-1})'(1 + X_{t-1}' \Psi_{t-2} X_{t-1}) \right\} \\
&\quad \times \frac{Y_{t-1} - X_{t-1}' \beta_{t-1}(\lambda)}{(\lambda + X_{t-1}' \Omega_{t-2}^{-1} X_{t-1})^2} \\
&\quad - \frac{X_{t-1}' \Omega_{t-2}^{-1}}{\lambda + X_{t-1}' \Omega_{t-2}^{-1} X_{t-1}} \frac{\partial \beta_{t-1}(\lambda)'}{\partial \lambda} X_{t-1},
\end{aligned}
$$

8

where

$$\Psi_t = \frac{\partial \Omega_t^{-1}}{\partial \lambda} = -\Omega_t^{-1} \frac{\partial \Omega_t}{\partial \lambda} \Omega_t^{-1}.$$

For the GAS model

$$\beta_t = c + B\beta_{t-1} + AH_{t-1}^{-1}\nabla_{t-1},$$

with $\theta = (c', \text{vec}(B)', \text{vec}(A)')'$, the corresponding recursion for the first derivative of $\beta_t(\theta)$ is

$$
\begin{aligned}
\frac{\partial \beta_t(\theta)'}{\partial \theta} \;=\; & \frac{\partial c'}{\partial \theta} + \frac{\partial \text{vec}(B)'}{\partial \theta}(\beta_{t-1}(\theta) \otimes I_n) \\
& + \frac{\partial \text{vec}(A)'}{\partial \theta}\left[(H_{t-1}^{-1}\nabla_{t-1}) \otimes I_n\right] \\
& + \frac{\partial \beta_{t-1}(\theta)'}{\partial \theta}\left(2X_{t-1}X_{t-1}'H_{t-1}^{-1}A' + B'\right).
\end{aligned}
$$

D   Superior Predictive Ability Test Results

As a robustness check for the MCS results presented in Section 4, Table A1 reports p-values of the test for Superior Predictive Ability (SPA). The test's hypothesis is that a given benchmark model is at least as good as all of its competitors. Hence a small p-value yields evidence against the benchmark model. Our implementation is based on the stationary bootstrap with a mean block length of $\lfloor T_{eval}^{1/3} \rfloor$, and we use the conservative estimator $\hat{\mu}^u$ for the mean loss differential (Hansen, 2005, p. 372).

The results in Table A1 are qualitatively very similar to the results based on the MCS approach. In particular, successful models that are included in the MCS typically achieve high SPA p-values when used as a benchmark model. Conversely, for poor performing models that are not included in the MCS, the SPA p-value is typically very small. For example, the RLS-shr model, which is contained in the MCS for all eight experiments, generally achieves an SPA p-value exceeding five percent. In this regard, RLS-shr outperforms all other models, each of which in at least one of the empirical experiments has a SPA p-value that is less than one percent. By contrast, the naïve model, which is not part of the MCS for any of the eight forecast experiments, attains an SPA p-value of zero in each case.

	Daily data				Monthly data			
					6-FF	10-Ind	25-FF	48-Ind
n	50	50	200	200	6	10	25	48
T	250	1250	250	1250	120	120	120	120
d-GAS-cβ_0	0.03	0.00	0.00	0.00	0.06	0.02	0.00	0.00
d-GAS-cβ_0-ols-ta	0.00	0.00	0.00	0.00	0.19	0.02	0.00	0.00
d-GAS-cβ_0-shr-ta	0.03	0.01	0.30	0.19	0.14	0.21	0.55	0.00
d-GAS-$c\beta_0$-ew	0.00	0.01	0.01	0.00	0.64	0.05	0.00	0.00
RLS-ols	0.00	0.64	0.00	0.11	0.07	0.14	0.00	0.01
RLS-shr	0.99	0.99	0.97	0.75	0.08	0.14	0.82	1.00
DCC	0.00	0.00	0.00	0.00	0.11	0.01	0.00	0.00
DCC-nl	0.75	0.70	0.80	0.98	0.11	0.35	0.02	0.00
OLS	0.00	0.00	0.00	0.00	0.89	0.54	0.00	0.00
SHR-l	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00
SHR-nl	0.62	0.01	0.96	0.40	0.79	0.11	0.24	0.00
naïve	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table A1: p-values of the Hansen (2005) test for Superior Predictive Ability. Rows correspond to benchmark models, columns correspond to forecast experiments. The one-sided Null hypothesis is that the benchmark is at least as good as all competitors.

References

Amemiya, T. (1985). *Advanced Econometrics*. Havard University Press.

Fissler, T. and Ziegel, J. F. (2019). Order sensitivity and equivariance of scoring functions. *Electronic Journal of Statistics*, 13:1166–1211.

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

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23:365–380.

Lambert, N. S., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138. ACM.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics*, 211:388–413.

Petersen, K. B. and Petersen, M. S. (2012). The matrix cookbook. Available at `http://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`.

Working Paper Series in Economics

recent issues