ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis: International Econometrics and Operations Research

A Relaxation On Volatility-Managed Portfolios

Trevor Visser
443480

ABSTRACT

We consider the performance results of the volatility-managed portfolio and reaffirm past conclusions on the poor out-of-sample performance of any real-time combination strategy of volatility management. Of the possible reasons to this poor performance, we analyze the effect of estimation risk on the optimal portfolio weighting scheme. From this estimation risk, we propose a relaxation of the term 'volatility-managed' such that we can make use of a multinomial logistic regression (Softmax model) that attempts to incorporate the estimation risk in its formulation as well as a heuristic algorithm that incorporates a model selection process. When compared to three other portfolio strategies including the $1/N$, volatility-managed, and mean-variance combination, the results show that the resulting Softmax model is consistently among the top performing strategies with regards to different risk-associated performance measures such as the Sharpe ratio and certainty equivalent returns.

Supervisor: Dr. X. Xiao
Second assessor: S.H.L.C.G. Vermeulen

Date final version: July 4th 2019

Contents

I. Introduction

Portfolio management is a hot topic in the financial world. Through the allocation of funds into several assets, investors are able to reduce their risk. One measure of this risk is the volatility of a particular asset. Based on this, Moreira and Muir (2017) introduced a particular portfolio strategy where the wealth of an investor is distributed inversely to the expected volatility of an asset—assets with higher (lower) volatility receive a smaller (larger) portion of the available wealth. Under their analysis, these volatility-managed portfolios were determined to outperform their unmanaged counterparts with respect to risk associated performance measures such as the Sharpe ratio and the certainty equivalent.

In response to the findings of Moreira and Muir (2017), Cederburg, O'Doherty, Wang, and Yan (2019) assessed the proposed volatility-managed strategy under 103 different equity strategies. Their findings show that the analysis of Moreira and Muir (2017) is not entirely true as their is evidence of forward looking bias. Besides this, they concluded that the volatility-management strategy is not implementable in practice, specifically in real-time situations and had to thus create their own real-time combination strategy. With regards to these developed strategies, the volatility-managed portfolios were not seen to significantly outperform their unmanaged counterparts.

With these results in mind, we replicate the research of Cederburg et al. (2019) and find that although their are numerical differences, the conclusions are very similar. Because of this, we investigate why the out-of-sample performance of the volatility-managed portfolios is poor. Particularly, we address the estimation risk associated with weight computation through volatility-management. It is found that the weights themselves are subject to extreme fluctuation as a result of poor estimation of portfolio moment conditions and thus a transformation of the mean-variance portfolio optimization into a multinomial logistic model (Softmax model) is proposed. This proposed Softmax model is tested against three different portfolio weighting strategies over seven different aggregate portfolios and is found to consistently be amongst the best-performing strategies in terms of risk-associated performance measures such as the Sharpe ratio and the certainty equivalent.

II. Data

In the replication of Cederburg et al. (2019), we attempt to use a dataset that is as similar as can be to Cederburg et al.'s research. To this end, we make use of both monthly and daily returns (in percentage points) for 8 factors over the time period from July 1963 to March 2019. The first few factors come from the French-Fama three and five-factor models (Fama and French (1993); Fama and French (2015))[1]. These include the market factor (Mkt), the momentum factor (MOM), the size (SMB) and value (HML) factors, and the investment (CMA) and profitability factors (RMW). While these factors align with those used by Cederburg et al. (2019), there was difficulty in finding the factors used in the q-factor model of Hou, Xue, and Zhang (2015), namely their profitability

[1]Courtesy of Cederburg et al. (2019), the data for the French-Fama factors can be found at the following website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

(ROE) and investment (IA) factors. Instead, we make use of a betting-against-beta factor (BAB) and a quality-minus-junk factor (QMJ)[2].

Although there was the potential to use other factors, the decision to omit these factors was based on data availability: while the data existed, they only consisted of monthly returns where both monthly and daily returns were required for portfolio construction.

For a brief comparison between the two datasets, we calculate the annualized mean return and standard deviation of the factors and compare with those found in Cederburg et al. (2019). The summary statistics can be found below in Table I.

Table I Summary Statistics Of Datasets

Dataset	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Mean	10.93	2.88	3.74	7.82	3.08	3.30	9.87	4.62
Standard Deviation	15.17	10.47	9.71	14.48	7.51	6.92	11.29	7.77
Cederburg et al. (2019)	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Mean	7.80	2.57	4.84	7.94	2.92	3.72	8.23	*NAN*
Standard Deviation	18.61	11.12	12.14	16.39	7.71	6.97	10.71	*NAN*

This table consists of both the annualized mean return and standard deviation return (in percentage points) of each factor in both the dataset used for the replication as well as the dataset used by Cederburg et al. (2019). The values are calculated through monthly return percentage points over the periods July 1963 to March 2019. The NAN's signify values that are not available as Cederburg et al. (2019) did not make use of the quality-minus-junk QMJ factor.

As can be seen, there are clear differences between the factors used by Cederburg et al. (2019) and those used in the replication. Because of this, it is to be expected that the numerical results following portfolio construction and performance evaluation will also differ.

III. Replication: Volatility-Managed Portfolio

To replicate the results of Cederburg et al. (2019), we consider both the in-sample and out-of-sample aspects. Beginning with the in-sample replication, the mathematical specifications of both portfolio construction and portfolio performance tests in Sections III.A and III.B. Once this is done, we give the in-sample results of the constructed portfolios' performances in Section III.C. Following this, we move to Cederburg et al.'s out-of-sample replication with Sections III.D and III.E.

A. Portfolio Construction

Suppose we have $n = 1, 2, \dots, N$ factors where r_t^n denotes the return of factor n in period t. For each time period t, the investor must allocate his investment position amongst these N factors and the real-interest rate. Let r_t^p be his excess portfolio return at time period t. Based on this

[2]Both the data for BAB and QMJ can be found on the website of Andrea Frazzini: http://people.stern.nyu.edu/afrazzin/

portfolio return, the volatility managed excess return $r^p_{\sigma,t}$ of r^p_t can be calculated by the following equation:

$$r^p_{\sigma,t} = \frac{c^*}{\hat{\sigma}^2_{t-1}} r^p_t \tag{1}$$

From Equation 1, the $\hat{\sigma}^2_{t-1}$ denotes the realized variance of the previous time period $t-1$ and c^* denotes a constant that renders the full sample variances of $r^p_{\sigma,t}$ and r^p_t to be equal. By the above formulation, the volatility-managed portfolio is a portfolio that bases its investment positions inversely to the lagged variances. It also implicitly has access to the real-interest rate.

B. *Portfolio Performance Measures*

To compare the performances of the volatility-managed portfolio and its original unmanaged counterpart, we consider a few performance measures besides the standard excess mean and standard deviation measurements.

The first is the Sharpe ratio which is defined as a measurement of a portfolio's return per unit of volatility. Our use of the Sharpe ratio considers the excess returns of each portfolio. Mathematically, denote r_t as the portfolio excess return for the months $t = 1, 2, \ldots, T$. The Sharpe ratio will be calculated as such:

$$\frac{\hat{\mu}_r \sqrt{12}}{\sqrt{\frac{\sum_{t=1}^{T}(r_t - \hat{\mu}_r)^2}{T-1}}}$$

where $\hat{\mu}_r$ is the average excess return $\sum_{t=1}^{T} r_t / T$ and $\sqrt{12}$ is the annualization factor.

In order to test if the Sharpe ratio of the volatility managed portfolio differs significantly from that of the unmanaged portfolio, we make use of a test as proposed by Jobson and Korkie (1981). Consider two portfolios x and y. We denote $\hat{\mu}_x$, $\hat{\sigma}_x$ and $\hat{\mu}_y$, $\hat{\sigma}_y$ as the excess mean returns and standard deviations for both portfolios respectively. By similar intuition, let $\hat{\sigma}_{x,y}$ denote the covariance between the excess returns of portfolios x and y. The asymptotically standard normal test statistic of Jobson and Korkie (1981) can thus be calculated as follows:

$$\hat{z}_{JK} = \frac{\hat{\sigma}_y \hat{\mu}_x - \hat{\sigma}_x \hat{\mu}_y}{\sqrt{\hat{\theta}}} \tag{2}$$

where

$$\hat{\theta} = \frac{1}{T}\left(2\hat{\sigma}^2_x \hat{\sigma}^2_y - 2\hat{\sigma}_x \hat{\sigma}_y \hat{\sigma}_{x,y} + \frac{1}{2}\hat{\mu}^2_x \hat{\sigma}^2_y + \frac{1}{2}\hat{\mu}^2_y \hat{\sigma}^2_x - \frac{\hat{\mu}_x \hat{\mu}_y}{\hat{\sigma}_x \hat{\sigma}_y}\hat{\sigma}^2_{x,y}\right) \tag{3}$$

The second performance measure is the certainty equivalent return (CER) which is loosely defined as the return an investor would take now instead of taking a chance on a riskier, higher return in the future. Returning to the previous mentioned notation as used in the Sharpe ratio calculation, let $\hat{\mu}_r$ and $\hat{\sigma}_r$ be the average excess return and standard deviation of the portfolio in question. The CER can be calculated as follows:

$$CER = \hat{\mu}_r - \frac{\gamma}{2}\hat{\sigma}^2_r$$

where γ is an investor specific risk aversion parameter. In the case of our replication, we follow the format of Cederburg et al. (2019) and set $\gamma = 5$. Similar to the Sharpe ratio test, we consider an asymptotically standard normal CER test as proposed by DeMiguel, Garlappi, and Uppal (2007).

$$\hat{z}_{DGU} = \frac{(\hat{\mu}_x - \frac{\gamma}{2}\hat{\sigma}_x^2) - (\hat{\mu}_y - \frac{\gamma}{2}\hat{\sigma}_y^2)}{\sqrt{\hat{\theta}}} \tag{4}$$

where

$$\hat{\theta} = \frac{1}{T}\left(\begin{bmatrix} 1 & -1 & -\frac{\gamma}{2} & \frac{\gamma}{2} \end{bmatrix} \begin{bmatrix} \hat{\sigma}_x^2 & \hat{\sigma}_{x,y} & 0 & 0 \\ \hat{\sigma}_{x,y} & \hat{\sigma}_y^2 & 0 & 0 \\ 0 & 0 & 2\hat{\sigma}_x^4 & 2\hat{\sigma}_{x,y}^2 \\ 0 & 0 & 2\hat{\sigma}_{x,y}^2 & 2\hat{\sigma}_y^4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -\frac{\gamma}{2} \\ \frac{\gamma}{2} \end{bmatrix} \right)$$

The final performance measure in consideration is termed as a spanning regression by Moreira and Muir (2017). Consider the terminology as used in Equation 1. Namely, the volatility-managed and unmanaged returns $r_{\sigma,t}^p$ and r_t^p. The spanning regression consists of regressing the volatility managed returns on the unmanaged returns.

$$r_{\sigma,t}^p = \alpha + \beta r_t^p + \epsilon_t$$

In the event that the volatility managed returns, on average, outperform those of their unmanaged counterparts, the value of the estimated α will be positive and significantly different from zero.

C. *In-Sample Results*

The construction of the volatility-managed portfolio of Equation 1 was applied to each factor in our dataset. By this construction, the portfolios consist of investment positions allocated between the factor itself and the real-interest rate. Table II shows the results of the performance measures on each portfolio. The excess means, standard deviations, Sharpe ratios, and α coefficients have all been annualized. It is important to note the distinction between Tables I and II: the results of Table I are based on *actual* returns whereas the results of Table II are based on *excess* returns.

As can be seen, there is mixed evidence on the effectiveness of the volatility-managed portfolios. From the Sharpe ratio perspective, only the MOM and BAB factors exhibit increased performance at the 5% significance level; from the CER point of view, there is no clear improvement for any of the factors; and lastly, from the spanning regression perspective, it is clear that the α exhibit a clear positive tendency, hinting at the effectiveness of volatility-management.

In comparison to the results of Cederburg et al. (2019), we see that although the numerical values are quite different—perhaps as a result of a differing dataset—, similar conclusions are drawn for the Sharpe ratios. In particular, Cederburg et al. (2019) conclude that only the factors of MOM and BAB exhibit significant increases in performance[3]. The α coefficients themselves

[3] Cederburg et al. (2019) also conclude that their IA factor has a significant performance increase. However, as this replication does not have the data for the IA factor, we chose to omit this particular detail.

exhibit strong similarities to that of Cederburg et al. (2019), despite their numerical differences. Excluding the difference in the HML results, both sides conclude significance of the α coefficient except for the factors of SMB and CMA. The difference in conclusions drawn for the HML factor may once again be a result of the different in datasets.

Table II In-Sample Performance of Volatility-Managed Portfolios

Unmanaged (Excess)	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Mean	6.36	-1.70	-0.83	3.25	-1.49	-1.28	5.29	0.04
Standard Deviation	15.22	10.54	9.67	14.45	7.56	6.91	11.32	7.79
Sharpe Ratio	0.42	-0.16	-0.09	0.23	-0.32	-0.19	0.47	0.01
CER	-47.69	-23.29	-19.56	-43.20	-12.01	-10.05	-26.27	-12.64
Managed (Excess)	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Mean	7.16	-2.69	-0.75	10.21	-0.52	-2.13	9.47	0.00
Standard Deviation	15.23	10.54	9.73	14.41	7.49	6.91	11.35	7.73
Sharpe Ratio	0.47	-0.26	-0.08	0.71	-0.07	-0.31	0.83	0.00
CER	-47.66	-23.34	-19.75	-42.36	-11.72	-10.12	-26.02	-12.44
Comparison	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Sharpe Ratio Difference	0.05	-0.10	0.01	0.48	0.13	-0.12	0.37	-0.00
p-value (JK)	0.63	0.35	0.95	0.00	0.28	0.23	0.00	0.98
CER Difference	0.03	-0.05	-0.20	0.84	0.29	-0.07	0.25	0.20
p-value (DGU)	0.49	0.47	0.37	0.28	0.21	0.40	0.38	0.30
Spanning Regression α	4.34	-0.16	1.33	10.51	2.22	0.16	8.14	1.70
p-value (α)	0.00	0.87	0.18	0.00	0.01	0.82	0.00	0.04

This table consists of direct performance comparisons between the volatility-managed portfolios and their unmanaged counterparts (over the time period July 1963 to March 2019) with regards to their excess returns. The mean, standard deviation, Sharpe ratio, and α coefficient have been annualized while the p-values were calculated from the test-statistics of Equations 2 and 4.

D. *Real-Time Combination*

With regards to the previous sections, the comparisons between volatility-managed and unmanaged portfolios were done purely on in-sample data. As this is not realistic for an investor in real-time, Cederburg et al. (2019) proposed a real-time combination strategy between two kinds of portfolios: a portfolio that has access to both the volatility-managed and unmanaged strategies; and a portfolio that only has access to the unmanaged strategy[4].

To briefly summarize their simulation approach, consider a time period of T months. Cederburg et al. (2019) choose the beginning K (such that $K < T$) months of their full dataset to estimate

[4]By Cederburg et al.'s formulation, these portfolios also have access to the real-interest rate.

sample moments of the given portfolio assets, $\hat{\mu}_{K+1}$ and $\hat{\Sigma}_{K+1}$, as well as c^* of Equation 1. Once these are estimated, the portfolio weights for the next time period $K+1$ are determined by the following equation:

$$\begin{bmatrix} \omega_{\sigma,K+1} \\ \omega_{K+1} \end{bmatrix} = \frac{1}{\gamma}\hat{\Sigma}^{-1}_{K+1}\hat{\mu}_{K+1} \tag{5}$$

In relative weight terms, Equation 5 can be written as

$$\begin{bmatrix} \lambda_{\sigma,K+1} \\ \lambda_{K+1} \end{bmatrix} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{|1^\intercal\hat{\Sigma}^{-1}\hat{\mu}|} \tag{6}$$

where 1 denotes a vector of ones. With regards to Equation 5, the investment position in the same period $K+1$ thus amounts to

$$y_{K+1} = \omega_{\sigma,K+1}\left(\frac{c^*}{\hat{\sigma}^2_K}\right) + \omega_{K+1}$$

where $\hat{\sigma}^2_K$ is a vector of the realized variances of the assets for the previous month. From the investment position, the excess return of period $K+1$ can be calculated by simply multiplying y_{K+1} with the unmanaged excess returns r_{K+1}. This process repeats with an expanding window to estimate the sample moments and c^* until a vector of length $T-K$ is created containing real-time excess returns. To sum up of the procedure, we consider the psuedocode presented below to generate an excess return series for both the investor that has access to the volatility-managed part and the investor that does not.

Algorithm 1: Real-Time Combination Simulation

Result: Excess Return Series ER

Set initial training period K;

Set risk aversion parameter γ;

Create original excess return series R of each asset over full time period;

Initialize empty excess return series ER;

Initialize $t = K+1$;

while $t \leq T$ **do**

> Calculate $\hat{\mu}_t$, $\hat{\Sigma}_t$, and c^* from historical data up to and including $t-1$;
>
> Calculate portfolio weights $\begin{bmatrix} \omega_{\sigma,t} & \omega_t \end{bmatrix}^\intercal = (1/\gamma)\hat{\Sigma}^{-1}_t\hat{\mu}_t$;
>
> Calculate investment position $y_t = \omega_{\sigma,t}\left(\frac{c^*}{\hat{\sigma}^2_{t-1}}\right) + \omega_t$;
>
> Calculate excess return $er_t = y_t R_t$;
>
> Add er_t to excess return series ER;
>
> Set $t = t+1$;

E. Real-Time Results

After implementing the real-time combination for each factor, we applied the performance measures and received the following results in Table III for both portfolios that had access to volatility management and portfolios that did not. In our replication, we maintained the same parameter values as Cederburg et al. (2019): namely, an initial training period of $K = 120$ months and a risk aversion parameter of $\gamma = 5$.

From Table III, there is no clear indication that the volatility-managed portfolio outperforms its unmanaged counterpart in a real-time situation. In fact, we see that the volatility-managed portfolio has a tendency to perform worse despite its in-sample results. This conclusion, despite numerical differences, is similar to that of Cederburg et al. (2019) as their results are mixed in the sense that some factors perform well in a volatility managed scenario whereas others perform well in a unmanaged scenario.

Table III Out-of-Sample Performance of Volatility-Managed Portfolios

Unmanaged (Excess/Real-Time)	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Mean	0.024	0.004	-0.004	0.011	0.059	0.004	0.041	0.007
Standard Deviation	0.062	0.048	0.025	0.081	0.174	0.062	0.098	0.076
Sharpe Ratio	0.384	0.088	-0.164	0.141	0.337	0.070	0.423	0.094
CER	0.001	-0.000	-0.000	-0.000	-0.001	-0.000	0.002	-0.001
Managed (Excess/Real-Time)	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Mean	0.022	-0.006	-0.008	0.059	0.054	0.006	0.061	0.007
Standard Deviation	0.067	0.037	0.037	0.132	0.179	0.074	0.100	0.076
Sharpe Ratio	0.334	-0.162	-0.208	0.451	0.302	0.080	0.603	0.079
CER	0.001	-0.001	-0.001	0.001	-0.002	-0.001	0.003	-0.001
Comparison	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
Sharpe Ratio Difference	-0.050	-0.250	-0.045	0.310	-0.035	0.011	0.180	-0.015
p-value (JK)	0.001	0.000	0.320	0.000	0.000	0.421	0.000	0.233
CER Difference	-0.000	-0.001	-0.001	0.002	-0.001	-0.000	0.002	-0.001
p-value (DGU)	0.003	0.003	0.033	0.016	0.000	0.259	0.000	0.001

A table consisting of results from the implementation of Algorithm 1 over the time period of August 1973 to March 2019. The implementation was done assuming a portfolio that consisted of the factor in question and the real interest rate. The mean, standard deviation, and Sharpe ratio have all been annualized and were calculated based on excess returns.

IV. The Proposal

A. The Problem

Following their results, Cederburg et al. (2019) posit that the poor out-of-sample performance of their proposed volatility-managed portfolios is a result of a few things. In particular, they acknowledged the estimation risk on the optimal portfolio weights—portfolios that have been constructed based on estimations of sample moments are often seen to dramatically change weights over time. To better visualize this, consider the mean-variance portfolio optimization below that Cederburg et al. (2019) made use of to compute the vector of optimal ex-post[5] weights, $\mathbf{y}$.

$$\max_{\mathbf{y}} \quad \mathbf{y}^\intercal \hat{\mu} - \frac{\gamma}{2}\mathbf{y}^\intercal \hat{\Sigma}\mathbf{y} \tag{7}$$

Where $\hat{\mu}$ is a vector of the mean excess returns per asset, γ is the particular risk-aversion parameter, and $\hat{\Sigma}$ is the variance-covariance matrix of excess asset returns. From Equation 7, the optimal portfolio weights can be found by

$$\mathbf{y} = \frac{1}{\gamma}\hat{\Sigma}^{-1}\hat{\mu}$$

which is simply Equation 5 rewritten. As can be seen, the optimal portfolio weights are dependent on the particular sample moment estimations of $\hat{\mu}$ and $\hat{\Sigma}$. In particular, the optimality of the portfolio weights are dependent on the accuracy of each sample moment estimation. To understand the weights fluctuations as a result of the estimation risk, we consider the derivative of the optimal weights $\mathbf{y}$ with respect to the one of the sample moments: the variance-covariance matrix $\hat{\Sigma}$.

$$\frac{\partial \mathbf{y}}{\partial \hat{\Sigma}} = -\frac{1}{\gamma}(\hat{\Sigma}^{-1}\hat{\mu})^\intercal \otimes \hat{\Sigma}^{-1} \tag{8}$$

Where the $\otimes$ signifies a tensor product[6]. From Equation 8, it is clear that the fluctuation of the optimal weights with respect to the variance-covariance matrix is highly dependent on the sample moment estimations[7]. In order for the weights to achieve optimality, the estimations must be highly precise. However, in general, estimations of the mean excess returns $\hat{\mu}$ and the variance-covariance matrix $\hat{\Sigma}$ is an extremely difficult topic.

As seen in their real-time combination strategy, Cederburg et al. (2019) estimate the sample moments by means of an expanding window over time, which provided extreme fluctuation in the optimal weights and thus poor out-of-sample performance.

B. Relaxation of Volatility-Managemenet

Because of the estimation risk associated with the optimal portfolio weights and the resulting poor out-of-sample performance, we propose a relaxation of the volatility-managed concept and a resulting model from this relaxation. As mentioned before, in Moreira and Muir (2017) and Cederburg et al. (2019), volatility-managed refers to the concept of minimizing a portfolio's overall volatility—an asset with higher (lower) volatility will receive lower (higher) weights. This decrease in volatility lowers an investor's overall risk while providing a lower return. Instead of this concept

[5]Ex-post meaning for actual results rather than for predictions.
[6]Given two vectors $\mathbf{v}$ and $\mathbf{w}$, a tensor product is defined as the outer product $\mathbf{v}\mathbf{w}^\intercal$.
[7]By the same logic, the derivative of Equation 6 is also found to depend on the sample moment estimation themselves. For a visualization of this, please see Appendix B.

of volatility management, we relax the idea of weights determined by inverse volatility levels and instead, propose the construction of weights based on linear combinations of the volatility levels.

To motivate this, consider the discrepancy between 'true' volatility and its predicted value, σ_t and $\hat{\sigma}_t$ respectively. Under the mean-variance optimization, the predicted volatility $\hat{\sigma}_t$ is taken to be the 'true' value and the weights are thus computed based on this prediction. However, this decision constitutes for a strong assumption on the chosen volatility estimator's predictive capability: with regards to the relationship between the chosen consistent and unbiased volatility estimator and the true volatility,

$$\sigma_t = \gamma \hat{\sigma}_t \tag{9}$$

the γ is assumed to be approximately one. This same intuition holds for the estimation of the true return, μ_t. However, the notion shown in Equation 9 is essentially a linear relationship. Based on this, we can incorporate the possible linear combination if the following assumption holds:

ASSUMPTION 1: *For any optimal portfolio weight combination at arbitrary time t, $\mathbf{y}_t$, their exists a direct linear relationship with the vector of true returns μ_t and true volatilities σ_t. In other words, μ_t and σ_t are statistically significant indicators:*

$$\begin{aligned} \mathbf{y}_t &= \alpha + \beta_1 \mu_t + \beta_2 \sigma_t \\ s.t. \quad \beta &= \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix}^\intercal \neq 0 \end{aligned}$$

If Assumption 1 holds, the inclusion of both the volatility and expected return estimators in a linear combination with their own β parameters allows for a methodology to account for the discrepancy between the 'true' and predicted values—a relaxation of the predictive capabilities assumption. To visualize this, consider the linear relationship presented in Assumption 1 and assume that instead of their 'true' values, we have predictions of the return and volatility: $\hat{\mu}_t$ and $\hat{\sigma}_t$.

$$\mathbf{y}_t = \alpha + \beta_1 \hat{\mu}_t + \beta_2 \hat{\sigma}_t \tag{10}$$

Now assume that their is discrepancy between the predictions and their true values such that the γ of Equation 9 is significantly different than one. Substituting the relationship of Equation 9 into Equation 10,

$$\begin{aligned} \mathbf{y}_t &= \alpha + \beta_1 \hat{\mu}_t + \beta_2 \hat{\sigma}_t \\ &= \alpha + \beta_1 (\gamma_1 \mu_t) + \beta_2 (\gamma_2 \sigma_t) \\ &= \alpha + \lambda_1 \mu_t + \lambda_2 \mu_t \end{aligned} \tag{11}$$

where $\lambda_1 = \beta_1 \gamma_1$ and $\lambda_2 = \beta_2 \gamma_2$. Under Equation 11 and Assumption 1, the linear combination consists of weights λ that inherently take into account both the significance and estimation discrepancy. In the case that their is no discrepancy between the true and predicted values, Equation 11 reverts back to Equation 10.

With this intuition, the relaxation of the volatility-managed structure intuitively leads to the transformation of the mean-variance portfolio optimization into a regression model that consists of a linear combination of each portfolio's asset's estimated volatility and returns to output asset weights. In particular, this model will be a multinomial logistic regression (or in simpler terms, a Softmax regression).

V. Methodology

A. The Softmax Regression Model

The multinomial logistic model, or Softmax model, is no stranger to financial market analysis. In particular, the works Upadhyay, Bandyopadhyay, and Dutta (2012) and Abdullah, Halim, Ahmad, and Rus (2008) provide considerable evidence of a logistic model's ability to predict stock performance and the likelihood of corporate failure. Specifically, Upadhyay et al. (2012) make use of previous stock returns and volatility as inputs to predict stock performance.

To better justify the choice of a Softmax regression, we must examine its particular mathematical properties. In essence, the Softmax regression is a combination of N linear equations with a logistic transformation on the subsequent N outputs. The resulting output is a probability distribution where each of the N outputs is transformed to possess a weight relative to its counterparts—a key difference between simply estimating N separate logistic models. By performing the logistic transformation on all N together, the Softmax model takes into account the interrelationships of both inputs and outputs.

In the context of portfolio optimization, the Softmax model is estimated over a time series effectively becoming a panel regression—the implications of this will be discussed after we introduce the relevant mathematical notation. Consider an investor at time t who has access to N assets and must divide his/her wealth amongst these assets. Besides these N assets, he/she possesses a vector of M explanatory variables, $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \ldots, x_{M,t}]^\intercal$, that are assumed to be significant indicators of the optimal weighting scheme for the following period $t+1$, $\mathbf{y}_{t+1} = [y_{1,t+1}, y_{2,t+1}, \ldots, y_{N,t+1}]^\intercal$. To maintain the relaxed volatility-managed format, the inputs $\mathbf{x}_t$ will at least contain a measure of the volatility for each asset. The Softmax regression begins with N linear combinations (one for each asset) to produce a "score" for each asset based on the inputs. This vector of scores will be denoted as $\mathbf{z}_{t+1}$ and can be seen below.

$$\mathbf{z}_{t+1} = \mathbf{W}\mathbf{x}_t + \mathbf{b} \tag{12}$$

where

$$\mathbf{z}_{t+1} = \begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \\ \vdots \\ z_{N,t} \end{bmatrix}, \qquad \mathbf{W} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,M} \\ \vdots & \ddots & \vdots \\ w_{N,1} & \cdots & w_{N,M} \end{bmatrix}, \qquad \mathbf{x}_t = \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{M,t} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$$

The matrix $\mathbf{W}$ and vector $\mathbf{b}$ correspond to parameters that must be optimized. Under Assumption 1, Equation 12 will be the linear combination that takes into account both variable significance and estimation discrepancy. Following the calculation of the score vector $\mathbf{z}_{t+1}$, the Softmax regression performs a logistic transformation on $\mathbf{z}_{t+1}$ to predict portfolio weights for the next period $\hat{\mathbf{y}}_{t+1}$.

$$\hat{\mathbf{y}}_{t+1} = \begin{bmatrix} y_{1,t+1} \\ y_{2,t+1} \\ \vdots \\ y_{N,t+1} \end{bmatrix} = \begin{bmatrix} \frac{e^{z_{1,t+1}}}{\sum_{k=1}^{N} e^{z_{k,t+1}}} & \frac{e^{z_{2,t+1}}}{\sum_{k=1}^{N} e^{z_{k,t+1}}} & \cdots & \frac{e^{z_{N,t+1}}}{\sum_{k=1}^{N} e^{z_{k,t+1}}} \end{bmatrix}^{\intercal} = \phi(\mathbf{z}_{t+1}) \tag{13}$$

where ϕ denotes a logistic function. However, Equation 13 has a clear numerical instability as $\lim_{\mathbf{z}_{t+1}\to\infty} \hat{\mathbf{y}}_t = NAN$. As a result of this, a solution arrives by providing a standardization in the form a λ as seen below in Equation 14.

$$\hat{\mathbf{y}}_{t+1} = \begin{bmatrix} y_{1,t+1} \\ y_{2,t+1} \\ \vdots \\ y_{N,t+1} \end{bmatrix} = \begin{bmatrix} \frac{e^{(z_{1,t+1}+\lambda)}}{\sum_{k=1}^{N} e^{(z_{k,t+1}+\lambda)}} & \frac{e^{(z_{2,t+1}+\lambda)}}{\sum_{k=1}^{N} e^{(z_{k,t+1}+\lambda)}} & \cdots & \frac{e^{(z_{N,t+1}+\lambda)}}{\sum_{k=1}^{N} e^{(z_{k,t+1}+\lambda)}} \end{bmatrix}^{\intercal} = \phi(\mathbf{z}_{t+1}) \tag{14}$$

Such that $\lambda = -max\{z_{1,t+1}, z_{2,t+1}, \ldots, z_{N,t+1}\}$. As a generalization of a logistic model, the Softmax regression possesses similar properties. In particular, it makes no assumptions on the distributions of both the independent variables $\mathbf{x}_t$ and dependent variables $\mathbf{y}_{t+1}$—an attractive feature that allows for increased robustness. Besides this, the construction of the Softmax regression allows for non-linear modeling and does not require inputs to be within a certain interval as it provides its own scaling (the logistic transformation). One of the assumptions made with the choice of the Softmax model on a time series (aka panel regression) is dynamic completeness. In other words, we assume that at each time period, our inputs are all encompassing:

ASSUMPTION 2: *The Softmax model specification is dynamically complete:*

$$Pr(\mathbf{\hat{y}_{t+1}}|\mathbf{x_t}) = Pr(\mathbf{\hat{y}_{t+1}}|\mathbf{x_t}, \mathbf{x_{t-1}}, \ldots)$$

This, in combination with Assumption 1, assures that the Softmax model parameter estimates account for variable significance, estimation discrepancy, and consistency. As such, all further analysis of the Softmax model is based on the fact that Assumptions 1 and 2 hold.

With regards to the mean-variance optimization mentioned earlier, one of the issues was its high dependency on the estimation accuracy—the optimal weights fluctuate dramatically with changes in estimations. Besides the inclusion of a linear combination as motivated in Section IV, the Softmax model possesses a desirable feature in its derivative. Similar to the derivative analysis of the mean-variance, consider the derivative of the Softmax output $\hat{\mathbf{y}}_{t+1}$ with respect to its input vector $\mathbf{x}_t$. As mentioned earlier, the input vector $\mathbf{x}_t$ will at least contain a measure of the volatility of each asset to maintain the relaxed volatility-management aspect. Thus the following derivative will implicitly take into account the volatility input.

$$\frac{\partial \hat{\mathbf{y}}_{t+1}}{\partial \mathbf{x}_t} = \frac{\partial \hat{\mathbf{y}}_{t+1}}{\partial \mathbf{z}_{t+1}} \frac{\partial \mathbf{z}_{t+1}}{\partial \mathbf{x}_t} = \hat{\mathbf{y}}_{t+1}\mathbf{1}^{\intercal} \circ (\mathbf{I} - \mathbf{1}\hat{\mathbf{y}}_{t+1}^{\intercal})\mathbf{W}$$

Where $\mathbf{1}$ denotes a vector of ones, $\mathbf{I}$ is the identity matrix, and $\circ$ denotes the Hadamard product. As can be seen, a change in the output weighting vector $\hat{\mathbf{y}}_{t+1}$ is a result of itself and the weighting matrix $\mathbf{W}$. If the log-likelihood function is specified correctly—in our case, a maximum likelihood function of the logistic distribution—then the estimate of $\mathbf{W}$ is consistent and unbiased. As a result, the change in weights possess robust features and is less dependent on the volatility estimates. This, in comparison to the mean-variance portfolio, shows that the Softmax model can possess both the ability to 'parameterize' input variable estimation discrepancies as well as output weights that are more robust to changes in the inputs $\mathbf{x}_t$.

B. *Model Estimation/Selection*

Besides the proposal of the Softmax model, we also propose an algorithmic procedure to define the Softmax model parameters $\mathbf{W}$ and $\mathbf{b}$. Instead of simply estimating parameters through some optimization process, model estimation will be broken down in two steps: parameter optimization and model selection.

B.1. Parameter Optimization

To estimate the parameters of the Softmax regression, we employ the use of gradient descent to minimize the negative log likelihood of the Softmax model. Consider the vector outputs of the Softmax regression $\hat{\mathbf{y}}_t = [\hat{y}_{1,t}, \hat{y}_{2,t}, \ldots, \hat{y}_{N,t}]^{\intercal}$ in comparison to the 'true' vector output $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{N,t}]^{\intercal}$. For this single time period t, intuition suggests that we maximize the following:

$$\prod_{n=1}^{N} \hat{y}_{n,t}^{y_{n,t}}$$

From this, it is trivial to see that for a complete time span of T time periods, we seek to maximize

$$\prod_{t=1}^{T} \prod_{n=1}^{N} \hat{y}_{n,t}^{y_{n,t}}.$$

Converting this to the log likelihood function and taking the negative, we get our loss function, L, for the parameter optimization process. This function is otherwise known as the Cross Entropy loss function.

$$L = -\sum_{t=1}^{T} \sum_{n=1}^{N} y_{n,t} \log\left(\hat{y}_{n,t}\right)$$

The gradient descent method employs a line-search based method that propagates the derivative of the loss function to each respective parameter in the Softmax model. Once this derivative is found, it is multiplied by a particular step size to determine how much the parameter should be adjusted. As an example, consider the derivative of the loss function L with respect to the parameter $w_{N,M}$ (as can be found in the definition of the weighting matrix $\mathbf{W}$). By means of the chain rule,

$$\frac{\partial L}{\partial w_{N,M}} = \sum_{i=1}^{N} \frac{\partial L}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial z_N} \frac{\partial z_N}{\partial w_{N,M}}$$

we get the portion of the derivative associated with the particular parameter $w_{N,M}$. Once this value is known, the parameter is updated with the following rule

$$w_{N,M} = w_{N,M} - \alpha \frac{\partial L}{\partial w_{N,M}} \tag{15}$$

where α is some predefined step size. In matrix format, the parameter matrix $\mathbf{W}$ and parameter vector $\mathbf{b}$ are updated as follows:

$$\mathbf{W} = \mathbf{W} - \alpha \frac{\partial L}{\partial \mathbf{W}}, \qquad \mathbf{b} = \mathbf{W} - \alpha \frac{\partial L}{\partial \mathbf{b}} \tag{16}$$

Each iteration through the training dataset performs small steps for each parameter as shown by Equation 15. For the actual implementation of the model estimation, we made use of Chollet et al.'s Keras module in python.

B.2. Model Selection

In model estimation, the parameters will be optimized over a categorical dataset that has been created to mark the highest earning asset in each particular time period. In this manner, the parameters will be optimized to create weights where the highest (lowest) weight is given to the asset which is predicted to have the highest (lowest) return. This, in respect to the probability distribution output of the Softmax model, provides a return ranking prediction.

However, this particular estimation scheme is counter-intuitive to the idea of decreasing overall portfolio risk. As such, an extension is added in the form of model comparison and selection. The model selection process consists of an iterative procedure between two different Softmax models: one to be denoted as the best model (BM), and the other to be denoted as the training model (TM). After undergoing a certain number of gradient descent iterations, the TM is put through a simulation on a 'test' dataset that has not been optimized over. The un-optimized BM is also put through a simulation on the same 'test' dataset. Both simulations output each models' Sharpe ratio and the model with the best Sharpe ratio value relative to the benchmark 1/N model Sharpe ratio becomes the new BM. To better visualize this process, we provide pseudocode in Algorithm 2.

By repeating this procedure several times, we ensure that the end best model (BM) is tuned to maximize profit (through parameter estimation) as well as its return-risk trade-off and robustness (through the model selection process). As a result of this methodology for model estimation and selection, we inherently converge to a model that is optimized over both the train and test dataset. As a result of this, any subsequent out-of-sample tests will be done on an untouched validation dataset. Thus, for an arbitrary dataset, there will be three partitions: one for model estimation (the training dataset), one for model selection (the testing dataset), and one for out-of-sample performance evaluation (the validation dataset).

Algorithm 2: Softmax parameter estimation and model selection

---

Result: Softmax Model

Create a randomly initialized instance of the Softmax model, denoted as the training model TM;

Create a randomly initialized instance of the Softmax model, denoted as the best model BM;

Set the parameters of the BM to be the same as the TM;

Initialize best model version = 0;

Set number of iterations for gradient descent E;

while *true* **do**

- **for** *E iterations over training dataset* **do**
  - Calculate derivative of loss function $\frac{\partial L}{\partial \mathbf{W}}$ and $\frac{\partial L}{\partial \mathbf{b}}$;
  - Update parameter values $\mathbf{W} = \mathbf{W} - \alpha \frac{\partial L}{\partial \mathbf{W}}$;
  - Update parameter values $\mathbf{b} = \mathbf{b} - \alpha \frac{\partial L}{\partial \mathbf{b}}$;
- Use TM to make predictions over test dataset to compute returns and Sharpe ratio SR_{TM};
- Use BM to make predictions over test dataset to compute returns and Sharpe ratio SR_{BM};
- Compute absolute value Sharpe ratio of $1/N$ model over test dataset $SR_{1N} = |SR_{1N}|$;
- **if** $\frac{SR_{TM}}{SR_{1N}} > \frac{SR_{BM}}{SR_{1N}}$ **then**
  - Set BM version +1;
  - Set parameters of the BM to be the same as TM;

C. Performance Evaluation

The evaluation of the Softmax model performance will be based on two things: in-sample evaluation of Assumption 1 and the model selection process; and a direct out-of-sample performance comparison with the volatility-managed portfolio strategy of Moreira and Muir (2017) (inverse weighting to expected volatility) and the mean-variance combination strategy of Cederburg et al. (2019) (Equation 6). In addition to this, we will also make use of the naive $1/N$ portfolio to set a benchmark in performance standards. This is motivated by the $1/N$ performance strength as seen by DeMiguel et al. (2007). The actual performance measures used will be the same as those in Section III.B.

In the simulations of Cederburg et al. (2019), the performances are evaluated between portfolio combinations that primarily focus on one factor and the real-interest rate. Instead of this approach, we consider aggregate portfolios that contain multiple factors as assets. By default, the real-interest rate will always be included. As the Softmax model parameters will be estimated over the training dataset, the performance evaluation of the four strategies will be simulated and reported over both the test and validation dataset. Besides this, to ensure clear comparisons, we restrict all strategies, if applicable, to make use of the same inputs: the expected return and expected volatility of each asset. For simplicity, these two moment conditions are calculated based on the daily values of the previous month in question. While we acknowledge that this restriction may result in the violation of the Softmax model assumption of dynamic completeness, we recall that one of the main arguments to the choice of a Softmax model is in its attempt to model the discrepancy in estimation risk. As such, for the sake of comparison, we retain our argument of keeping similar inputs. In real-life situations, however, the choice of inputs is open-ended.

VI. Results

The results will be broken into three subsections, each corresponding to the relevant results for the created datasets: train, test, validation. As the training dataset is used to estimate parameters for the Softmax model, it will essentially comprise of in-sample analysis with respect to variable significance and the effectiveness of the model selection process. The test and validation dataset sections will consist of their respective out-of-sample analysis in the form of performance comparisons between the four portfolio management strategies on seven different aggregate portfolios. As explained before, these four portfolio management strategies include the $1/N$ strategy, the original volatility-managed strategy of Moreira and Muir (2017), and the real-time combination strategy of Cederburg et al. (2019). With regards to the seven aggregate portfolios, the first (and largest) consisted of all available factors in addition to the real interest rate (nine assets) while the six other aggregate portfolios were created by successfully removing one asset at a time. To visualize this Table X in Appendix A lists the assets contained within each aggregate portfolio.

A. Training Dataset

Parameter estimation of the Softmax model occurred on the first 334 months of the total dataset from July 1963 to March 2005. Although there were several Softmax models estimated (one for each aggregate portfolio), we will primarily focus our analysis on the largest, nine asset model[8]. As mentioned in Section V.C, we restricted the inputs to the sample moments of the previous month for each asset. With the model estimated, Table IV below shows a brief summary of the p-values of each sample moments' significance for their respective asset.

[8]In order of occurrence, the nine asset aggregate portfolio consisted of the real interest rate RF, Mkt, SMB, HML, MOM, RMW, CMA, BAB, QMJ.

Table IV ***p-Values of Input Significance For Softmax Model***

	Mkt	*SMB*	*HML*	*MOM*	*RMW*	*CMA*	*BAB*	*QMJ*
$\hat{\mu}$	0.000	0.691	0.007	0.014	0.460	0.310	0.000	0.634
$\hat{\sigma}^2$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

A table consisting of the p-values for the significance of the sample moment inputs for the Softmax model estimation of the nine asset aggregate portfolio. The $\hat{\mu}$ and $\hat{\sigma}$ denote the first and second sample moments respectively and were calculated using the daily values of the previous month.

From the table, it is clear to see that the expected variance $\hat{\sigma}$ is significant for each asset. The expected return $\hat{\mu}$ is significant at the 5% significance level for half the assets. From this we can conclude that Assumption 1 holds for half the assets contained in the aggregate portfolio. The implications of this is that the Softmax model may no longer account for estimation risk when the particular factors are included in its aggregate portfolio.

A key point in the methodology section V was the inclusion of a model selection process besides the usual parameter estimation. In order to see the results of this inclusion, we compare two Softmax models estimated over our nine asset aggregate portfolio: one with the model selection (MS) process, one without. Table V below shows the model specific performance over the training dataset.

Table V ***In-Sample Softmax Model Comparison***

Performance	Softmax MS	Softmax
Mean (Excess)	6.211	3.259
Standard Deviation (Excess)	6.599	6.344
Sharpe Ratio	0.943	0.515
CER	-8.527	-8.088
Comparison	Softmax MS	Softmax
p-value ($\hat{z}_{JK}$)	0.000	0.000
p-value ($\hat{z}_{DGU}$)	0.195	0.195

This table consists of a direct performance comparison between a Softmax model estimated with model selection (Softmax MS) and a Softmax model estimated without model selection (Softmax) the training dataset for the nine asset aggregate portfolio (in reference to Algorithm 2). The means, standard deviations and Sharpe ratio have all been annualized and all performance indicators have been calculated with respect to excess returns.

From Table V, it is clear that the Softmax model that underwent the model selection significantly outperforms its counterpart. These results are similar when extended to the test and validation dataset. If interested, these results can be found in Appendix A in Table IX.

B. Test Dataset

The test dataset consists of 167 months over the time period from June 1991 to April 2005. Although the Softmax MS model was technically optimized over our test dataset (by means of the model selection procedure), we evaluate its performance in an out-of-sample perspective against the three other portfolio management strategies. This is motivated by the fact that although the model was chosen to perform best Sharpe ratio-wise over the test dataset, it was still subject to parameter estimations made solely through the training dataset. In the hypothetical situation where the Softmax MS parameters were estimated poorly, the performance in the test dataset would justifiably reach a lower, upper bound.

The performance of the four strategies for the seven different aggregate portfolios are shown below in Table VI.

Table VI Test Dataset Strategy Performance

1/N	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	3.830	4.010	2.804	2.998	3.334	2.268	2.105
Standard Deviation	4.790	4.848	3.764	3.868	4.692	4.270	6.665
Sharpe Ratio	0.802	0.830	0.747	0.778	0.713	0.533	0.317
CER	-4.431	-4.533	-2.701	-2.848	-4.282	-3.586	-9.024
Volatility-Managed	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	0.033	0.024	0.008	0.014	0.013	-0.006	-0.002
Standard Deviation	0.055	0.042	0.037	0.030	0.023	0.019	0.017
Sharpe Ratio	0.607	0.567	0.219	0.452	0.575	-0.303	-0.103
CER	0.002	0.002	0.000	0.001	0.001	-0.001	0.000
Mean-Variance	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	0.002	0.002	0.004	0.003	0.001	-0.001	-0.003
Standard Deviation	0.021	0.021	0.019	0.021	0.018	0.011	0.011
Sharpe Ratio	0.105	0.089	0.188	0.161	0.072	-0.057	-0.309
CER	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Softmax MS	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	10.021	10.789	7.646	7.409	9.658	5.461	3.597
Standard Deviation	6.767	7.161	8.110	8.350	8.478	6.766	9.092
Sharpe Ratio	1.485	1.511	0.946	0.890	1.143	0.810	0.397
CER	-8.649	-9.719	-12.983	-13.823	-14.079	-9.025	-16.820

A table consisting of the performances of four different weighting strategies (naive $1/N$, volatility-management, mean-variance combination, and Softmax MS outputs) for 9 different aggregate portfolio combinations. The four strategies were simulated over a test dataset that ranged June 1991 to April 2005. The means, standard deviations, and Sharpe ratios have been annualized.

Upon an initial glance, we begin to see that with respect to the Sharpe ratio, the results suggest that the Softmax MS model consistently performs best relative to the benchmark $1/N$ strategy. In particular, we acknowledge that on average, it outperforms the volatility-managed and mean-variance strategies. However, the certainty equivalent of the Softmax MS model is clearly the worst.

Instead of eyeballing these numerical results, we consider the hypothesis tests of significant difference. Table VII below shows the results of these tests with respect to the $1/N$ strategy. It shows that, relative to the $1/N$, the Softmax MS model only significantly outperforms the $1/N$ once out of all seven aggregate portfolios. In all other aggregate portfolios, it statistically performs on par with the $1/N$. Despite these results, we can still conclude that the Softmax MS model outperforms the mean-variance and volatility-managed strategies by seeing that they are outperformed by the $1/N$ strategy. Besides this, we see that the Softmax MS model performs best on aggregate portfolios with a higher number of assets—an expected characteristic as more assets provides a greater possibility of diversity.

Table VII Test Dataset Performance Comparison

Volatility-Managed	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Sharpe Ratio Difference	-0.195	-0.263	-0.528	-0.326	-0.138	-0.836	-0.420
$\hat{z}_{JK}$	-0.662	-0.957	-1.768	-1.085	-0.531	-3.055	-1.811
p-value	0.508	0.339	0.077	0.278	0.595	0.002	0.070
Mean-Variance	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Sharpe Ratio Difference	-0.697	-0.741	-0.559	-0.616	-0.641	-0.590	-0.626
$\hat{z}_{JK}$	-1.927	-2.054	-1.497	-1.662	-1.638	-1.592	-1.615
p-value	0.054	0.040	0.134	0.097	0.101	0.111	0.106
Softmax MS	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Sharpe Ratio Difference	0.683	0.682	0.199	0.112	0.430	0.277	0.080
$\hat{z}_{JK}$	1.939	1.999	0.842	0.409	1.848	1.461	0.855
p-value	0.053	0.046	0.400	0.683	0.065	0.144	0.932

A table consisting of direct performance calculation of the weighting strategies of volatility-managed, mean-variance and the Softmax with model selection (Softmax MS) against the $1/N$ strategy. The calculations are done based on the test dataset simulation with $\hat{z}_{JK}$ denoting the Sharpe ratio difference test statistic calculated through Equation 2.

C. Validation Dataset

Similar to the analysis of the test dataset Section VI.B, we compare performances of the four strategies on seven aggregate portfolios, but on the validation dataset. The validation dataset also consists of 167 months but ranges from May 2005 to March 2019—the time period after the test dataset. To avoid redundancy, we choose to omit the numerical results themselves and instead look immediately at the comparison Table VIII. If interested, however, the table of validation numerical results can be found in Appendix A in Table XI.

Returning our attention to the validation comparison Table VIII, we see very different results from those of the test dataset. In particular, the volatility-managed strategy now performs on par with the Softmax MS model with regards to their respective best performing aggregate portfolios. Both the volatility-managed and Softmax MS strategy significantly outperform the $1/N$ model in one aggregate portfolio and perform on par with the $1/N$ for all other aggregate portfolios. It is also particularly interesting to see the degradation in the performance of the Softmax MS model in comparison to the test dataset.

While one may argue that this is because the Softmax MS model was optimized over the test dataset and thus the validation dataset is a true indicator of its performance, we again refer back to our argument specifying that although the Softmax MS model was selected to perform best Sharpe ratio-wise on the test dataset, this selection process is subject to the optimal parameter estimation found over the training dataset—poor parameter estimation of the training dataset will provide a low, upper bound on the model selection process.

Table VIII Validation Dataset Performance Comparison

Volatility-Managed	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Sharpe Ratio Difference	0.505	0.386	0.142	0.097	0.078	-0.140	-0.179
$\hat{z}_{JK}$	2.265	1.779	0.697	0.465	0.384	-0.618	-0.760
p-value	0.024	0.075	0.486	0.642	0.701	0.537	0.447
Mean-Variance	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Sharpe Ratio Difference	-0.752	0.060	-0.002	-0.458	-0.721	-0.396	-0.352
$\hat{z}_{JK}$	-1.908	0.162	-0.006	-1.164	-1.801	-0.999	-0.857
p-value	0.056	0.871	0.995	0.244	0.072	0.318	0.391
Softmax MS	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Sharpe Ratio Difference	0.374	0.255	0.029	-0.347	-0.236	0.139	0.090
$\hat{z}_{JK}$	1.295	0.950	0.109	-1.258	-0.869	1.675	2.145
p-value	0.195	0.342	0.913	0.208	0.385	0.094	0.032

A table consisting of direct performance calculation of the weighting strategies of volatility-managed, mean-variance and the Softmax with model selection (Softmax MS) against the $1/N$ strategy. The calculations are done based on the validation dataset simulation with $\hat{z}_{JK}$ denoting the Sharpe ratio difference test statistic calculated through Equation 2.

Instead, we hypothesize that the degradation in performance of the Softmax MS model is a result of the time-varying aspect of the dataset. Due to the input restriction placed on all four strategies, it is likely that the Softmax MS model didn't contain enough input information to account for all time-variability thus violating Assumption 2 on dynamic completeness.

Despite this, however, the Softmax MS model is still not significantly outperformed by any one strategy and we can conclude that in general, the Softmax MS model performs on-par, if not better, than all other portfolio strategies.

VII. Conclusion

Beginning with a replication of the research done by Cederburg et al. (2019) in Section III, we reached similar conclusions despite clear numerical differences on both the in-sample and out-of-sample analysis. The real-time combination strategy proposed by Cederburg et al. (2019) (as a result of the forward-looking bias found in the original volatility-managed research of Moreira and Muir (2017)) suffered poor out-of-sample performance with a possible explanation being in the estimation risk associated with their strategy. As a result, we investigated the mathematical considerations of this estimation risk in Section IV and concluded that highly precise estimators were needed to avoid dramatic portfolio weight fluctuations over time.

To this end, we proposed in Section V a relaxation of the 'volatility-managed' term as proposed by Moreira and Muir (2017) along with the use of a multinomial logistic model—its effectiveness being built on the assumptions of dynamic completeness and a linear relationship between optimal weights and the inputs. Besides this, we also proposed an algorithmic estimation procedure that made use of both parameter estimation and model selection.

It was seen in the results Section VI that the extra model selection portion provided a significant performance increase in comparison to a model that was estimated without the model selection. The results also show, that in comparison to the three other portfolio strategies: the $1/N$ strategy, the pure volatility-managed strategy of Moreira and Muir (2017), and the real-time combination strategy of Cederburg et al. (2019), the Softmax MS model performed relatively well. However, the Softmax MS model showed clear time-related performance degradation—an indicator of model mis-specification. This was due mostly to the input restrictions set on the Softmax MS model in order to ensure absolute performance comparison between each strategy. Because of this, a possible extension to this research may be into the performance of the Softmax MS model given inputs that provide dynamic completeness. Another drawback of the Softmax MS model is that due to its construction, the weights are restricted to being positive such that short-selling is not possible. This leads to another possible extension in the form of a model that allows for short-selling[9].

[9]One possible example of such a model is a tangent hyperbolic regression. By its construction, the outputs are on the interval [-1,1]. However, there is no restriction such that the outputs sum to one and thus this constraint must be manually added.

REFERENCES

Abdullah, Nur Adiana Hiau, Abd Halim, Hamilton Ahmad, and Rohani Md Rus, 2008, Predicting corporate failure of malaysias listed companies: Comparing multiple discriminant analysis, logistic regression and the hazard model, *International Research Journal of Finance and Economics* 15, 201–217.

Cederburg, Scott, Michael S O'Doherty, Feifei Wang, and Xuemin Sterling Yan, 2019, On the performance of volatility-managed portfolios, *Journal of Financial Economics (JFE), Forthcoming* .

Chollet, François, et al., 2015, Keras, `https://keras.io`.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2007, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The review of Financial studies* 22, 1915–1953.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of financial economics* 116, 1–22.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Jobson, J Dave, and Bob M Korkie, 1981, Performance hypothesis testing with the sharpe and treynor measures, *The Journal of Finance* 36, 889–908.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *The Journal of Finance* 72, 1611–1644.

Upadhyay, Arun, Gautam Bandyopadhyay, and Avijan Dutta, 2012, Forecasting stock performance in indian market using multinomial logistic regression, *Journal of Business Studies Quarterly* 3, 16.

Appendix A. Supplementary Results

Table IX Softmax Model Comparison

	Train		**Test**		**Valid**	
	Softmax MS	Softmax	Softmax MS	Softmax	Softmax MS	Softmax
Mean (Excess)	6.211	3.259	10.021	8.326	5.171	4.336
Standard Deviation (Excess)	6.599	6.344	6.767	8.739	5.630	6.590
Sharpe Ratio	0.943	0.515	1.485	0.956	0.921	0.660
CER	-8.527	-8.088	-8.649	-15.122	-6.132	-8.632
p-value ($\hat{z}_{JK}$)	0.000	0.000	0.017	0.017	0.209	0.209
p-value ($\hat{z}_{DGU}$)	0.195	0.195	0.000	0.000	0.005	0.005

A complete table consisting of a direct performance comparison between the Softmax model estimated with (Softmax MS) and without (Softmax) model selection over the training, test, and validation dataset for the nine asset aggregate portfolio. The means, standard deviations and Sharpe ratio have all been annualized and all performance indicators have been calculated with respect to excess returns.

Table X Aggregate Portfolio Construction

9 Assets	RF	Mkt	SMB	HML	MOM	RMW	CMA	BAB	QMJ
8 Assets	RF	Mkt	SMB	HML	MOM	RMW	CMA	BAB	
7 Assets	RF	Mkt	SMB	HML	MOM	RMW	CMA		
6 Assets	RF	Mkt	SMB	HML	MOM	RMW			
5 Assets	RF	Mkt	SMB	HML	MOM				
4 Assets	RF	Mkt	SMB	HML					
3 Assets	RF	Mkt	SMB						

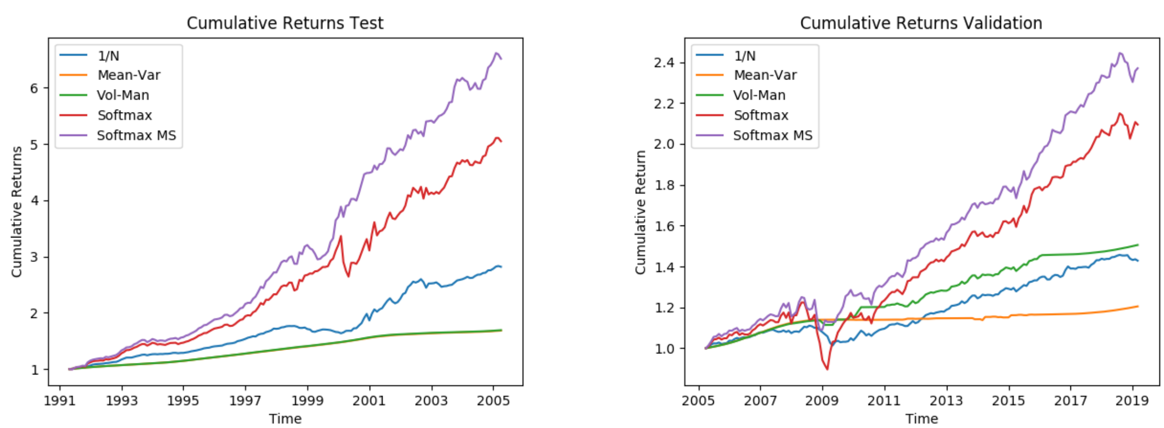
A table consisting of the assets including in each of the aggregate portfolios. As can be seen, each portfolio is created by successfully removing one asset at a time.

Table XI Validation Dataset Strategy Performance

1/N	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	1.394	1.182	0.810	1.142	1.076	1.460	2.930
Standard Deviation	2.555	2.884	3.109	3.436	4.388	5.904	6.455
Sharpe Ratio	0.547	0.411	0.261	0.333	0.246	0.248	0.455
CER	-1.236	-1.624	-1.934	-2.350	-3.897	-7.096	-8.385
Volatility-Managed	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	1.753	1.530	0.884	1.023	1.003	0.408	1.102
Standard Deviation	1.671	1.926	2.198	2.382	3.103	3.785	4.003
Sharpe Ratio	1.052	0.797	0.403	0.431	0.324	0.108	0.276
CER	-0.432	-0.641	-0.927	-1.090	-1.910	-2.933	-3.226
Mean-Variance	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	-0.499	0.282	1.933	-0.035	-0.129	-0.035	0.099
Standard Deviation	2.449	0.601	7.479	0.280	0.272	0.238	0.957
Sharpe Ratio	-0.205	0.471	0.259	-0.125	-0.475	-0.148	0.103
CER	-1.283	-0.051	-11.424	-0.019	-0.026	-0.015	-0.182
Softmax MS	*9 Assets*	*8 Assets*	*7 Assets*	*6 Assets*	*5 Assets*	*4 Assets*	*3 Assets*
Mean	5.171	3.797	1.970	-0.164	0.123	2.995	4.408
Standard Deviation	5.630	5.713	6.812	12.134	11.897	7.771	8.110
Sharpe Ratio	0.921	0.667	0.290	-0.014	0.010	0.387	0.545
CER	-6.132	-6.443	-9.447	-30.506	-29.300	-12.257	-13.253

This table consists of the performances of four different weighting strategies (naive $1/N$, volatility-management, mean-variance combination, and Softmax MS outputs) for 9 different aggregate portfolio combinations. The four strategies were simulated over a validation dataset that ranged from May 2005 to March 2019. The means, standard deviations, and Sharpe ratios have been annualized.

Figure 1. Cumulative returns for portfolio weighting strategies



***(a)** Test Dataset* ***(b)** Validation Dataset*

Graphical representation of the cumulative returns for all four portfolio weighting strategies over both the test and validation dataset on the nine asset aggregate portfolio. The returns are calculated on the assumption of no transaction costs and an initial wealth of 1. With regards to the legend, $1/N$ denotes the $1/N$ strategy, Mean-Var denotes the mean-variance combination strategy of Cederburg et al. (2019), Vol-Man denotes the pure volatility-managed strategy, and Softmax and Softmax MS denote the two Softmax models, one without model selection, and one with (MS).

Table XII Correlation: Asset Weights vs Expected Variance

Correlation	*RF*	*Mkt*	*SMB*	*HML*	*MOM*	*RMW*	*CMA*	*BAB*	*QMJ*
1/N	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Volatility-Managed	0.447	0.001	-0.054	-0.052	-0.011	-0.154	-0.197	-0.046	0.055
Mean-Variance	-0.065	0.001	-0.014	0.025	0.010	0.015	0.006	0.003	-0.008
Softmax MS	0.020	-0.036	-0.102	-0.267	0.076	0.004	0.428	-0.236	-0.162

Table consisting of the correlation coefficients between the asset weighting scheme and the respective expected variance input of the nine asset aggregate portfolio. The correlations coefficients have been calculated over the aggregate test+validation dataset.

Appendix B. Supplementary Mathematics

We denote the vector of optimal relative weights in the mean-variance optimization as the following.

$$\hat{\mathbf{y}} = \frac{\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}}{|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|} \tag{B1}$$

To visualize how the optimal weights change with respect to the inputs, we evaluate the derivative with respect to both inputs $\hat{\mathbf{\Sigma}}$ and $\hat{\mu}$. Beginning with the $\hat{\mathbf{\Sigma}}$ input, the derivative can be rewritten as such:

$$\frac{\partial \hat{\mathbf{y}}}{\partial \hat{\mathbf{\Sigma}}} = \frac{|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|\left(\frac{\partial(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})}{\partial\hat{\mathbf{\Sigma}}}\right) - \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}\left(\frac{\partial|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|}{\partial\hat{\mathbf{\Sigma}}}\right)}{(\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})^2} \tag{B2}$$

To evaluate this derivative, we solve the individual derivatives $\frac{\partial(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})}{\partial\hat{\mathbf{\Sigma}}}$ and $\frac{\partial|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|}{\partial\hat{\mathbf{\Sigma}}}$ separately before bringing them back together in Equation B2. This can be seen below in Equation B3 and B4.

$$\frac{\partial(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})}{\partial\hat{\mathbf{\Sigma}}} = -(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}) \otimes \hat{\mathbf{\Sigma}}^{-1} \tag{B3}$$

and

$$\frac{\partial|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|}{\partial\hat{\mathbf{\Sigma}}} = -sign(\hat{\mu}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\mathbf{1})(\hat{\mathbf{\Sigma}}^{-1}\mathbf{1})(\hat{\mu}\hat{\mathbf{\Sigma}}^{-1}) \tag{B4}$$

Where $-sign$ signifies an element-wise signum function. For example, $sign(-3) = -1$, $sign(1e^{99}) = 1$, $sign(\begin{bmatrix}0.55 & -1.48\end{bmatrix}^{\intercal}) = \begin{bmatrix}1 & -1\end{bmatrix}^{\intercal}$, etc. Plugging Equations B3 and B4 into Equation B2, we get the following solution.

$$\frac{\partial \hat{\mathbf{y}}}{\partial \hat{\mathbf{\Sigma}}} = \frac{|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|\Big(-(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}) \otimes \hat{\mathbf{\Sigma}}^{-1}\Big) - \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}\Big(-sign(\hat{\mu}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\mathbf{1})(\hat{\mathbf{\Sigma}}^{-1}\mathbf{1})(\hat{\mu}\hat{\mathbf{\Sigma}}^{-1})\Big)}{(\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})^2} \tag{B5}$$

Just as in Equation 8, the derivative of optimal relative weights depends on the estimations of each sample moment. We now turn to look at the derivative with respect to the input $\hat{\mu}$. Similar to that of the $\hat{\mathbf{\Sigma}}$ input, we rewrite the derivative into its quotient rule form and evaluate the individual derivatives separately.

$$\frac{\partial \hat{\mathbf{y}}}{\partial \hat{\mu}} = \frac{|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|\left(\frac{\partial(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})}{\partial \hat{\mu}}\right) - \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}\left(\frac{\partial|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|}{\partial \hat{\mu}}\right)}{(\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})^2} \tag{B6}$$

such that

$$\frac{\partial(\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})}{\partial \hat{\mu}} = \hat{\mu} \tag{B7}$$

$$\frac{\partial|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|}{\partial \hat{\mu}} = sign\Big(\hat{\mu}^{\intercal}(\hat{\mathbf{\Sigma}}^{-1})^{\intercal}\mathbf{1}\Big)\Big((\hat{\mathbf{\Sigma}}^{-1})^{\intercal}\mathbf{1}\Big) \tag{B8}$$

Taking Equations B7 and B8 and plugging them back into Equation B6, we get the following.

$$\frac{\partial \hat{\mathbf{y}}}{\partial \hat{\mu}} = \frac{|\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu}|(\hat{\mu}) - \hat{\mathbf{\Sigma}}^{-1}\hat{\mu}\Big(sign\Big(\hat{\mu}^{\intercal}(\hat{\mathbf{\Sigma}}^{-1})^{\intercal}\mathbf{1}\Big)\Big((\hat{\mathbf{\Sigma}}^{-1})^{\intercal}\mathbf{1}\Big)\Big)}{(\mathbf{1}^{\intercal}\hat{\mathbf{\Sigma}}^{-1}\hat{\mu})^2} \tag{B9}$$

Appendix C. Programming Code

In this section, we address the contents of the code and folders found in the attached zip file. Beginning with the folders, three are included: *data*, *models*, and *saved_models*.

The *data* folder consists of two csv files that contain the daily and monthly returns for each factor. These csv files are referenced throughout all the code and thus their placement is of importance.

The *models* folder is the folder that each successive 'best model' from Algorithm 2 is saved into. They are organized into sub-folders that are based on the number of factors contained within each aggregate portfolio. The contents of each aggregate portfolio can be determined through the *config.py* file (to be explained below).

Lastly, the *saved_models* folder consists of the final models used in this paper's results. In particular, the model with the highest version number in each sub-folder is used to generate the results.

Moving onto the code itself, there are several python (.py) files used in conjunction to generate the results found in this paper. Below, we list the names of each .py file and their purpose.

config.py: This python file consists of the configuration of the aggregate portfolios. By changing what asset is in the list, the aggregate portfolio used in all other .py files change.

environment.py: This python file consists of an Environment class that generates the train, test, and validation datasets from the data available (from the *data* folder) for the purpose of model estimation and selection.

measures.py: This python file consists of a PerformanceMeasures class that consists of all performance measures used in this paper.

model_estimation_selection.py: This python file consists of the actual implementation of the model estimation and selection procedure of Algorithm 2. It saves each newly chosen 'best model' into the *models* folder.

replication.py: This python file consists of the replication of Cederburg et al.'s results. The outputs of each portion of the replication are printed out in a similar manner to the tables found in Cederburg et al. (2019).

strategy_comparison.py: This python file consists of direct performance comparison between the different weighting strategies: $1/N$, volatility-management, the real-time combination strategy of Cederburg et al. (2019), the Softmax model without model selection, and the Softmax model with model selection.