Article

Risk Characterization of Firms with ESG Attributes Using a Supervised Machine Learning Method

Prodosh Eugene Simlai

Abstract: We examine the risk–return tradeoff of a portfolio of firms that have tangible environmental, social, and governance (ESG) attributes. We introduce a new type of penalized regression using the Mahalanobis distance-based method and show its usefulness using our sample of ESG firms. Our results show that ESG companies are exposed to financial state variables that capture the changes in investment opportunities. However, we find that there is no economically significant difference between the risk-adjusted returns of various ESG-rating-based portfolios and that the risk associated with a poor ESG rating portfolio is not significantly different than that of a good ESG rating portfolio. Although investors require return compensation for holding ESG stocks, the fact that the risk of a poor ESG rating portfolio is comparable to that of a good ESG rating portfolio suggests risk dimensions that go beyond ESG attributes. We further show that the new covariance-adjusted penalized regression improves the out-of-sample cross-sectional predictions of the ESG portfolio’s expected returns. Overall, our approach is pragmatic and based on the ease of an empirical appeal.

Keywords: greens stocks; large datasets; mahalanobis distance; penalized regressions; risk characterization; prediction

JEL Classification: C51; C53; C55

1. Introduction

The last two decades have witnessed dramatic growth in sustainable investing—an investment strategy focused on environmental, social, and governance (ESG) criteria (Bolton and Kacperczyk 2021; Starks 2023). Green stocks are companies that are supposed to have an ESG footprint, while brown stocks are companies with no such expectations (Pásstor et al. 2022). According to BlackRock (2020), 88% of their clients rank the environment as “the priority most in focus.” In the presidential address of the American Finance Association, Starks (2023) argues that the motivations for investing in ESG products are either value-based or values-based. According to Starks (2023), the term ESG value relates to traditional investment goals such as risk management and return opportunities, while the term ESG values suggests that the investment motive is nonpecuniary. Although the performance of ESG companies is not a black box, little research has been completed on the risks associated with investing in green companies.

In this paper, we use firm-level monthly data from 2009 to 2022 and investigate the risk–return tradeoff of a portfolio of companies that have tangible ESG attributes. We create our sample using companies with ESG ratings from Morningstar Sustainalytics. We introduce a new type of penalized regression using the Mahalanobis distance-based covariance-adjusted method (Mahalanobis 1936) and discuss the usefulness of this covariance-adjusted shrinkage estimate using the sample of ESG firms. We show the relative performance of the resulting shrinkage estimate from a cross-sectional regression model.
that allows non-sparse and heteroskedastic residuals. Similar to the existing work (Bolton and Kacperczyk 2021; Safiullah et al. 2022), our results show that ESG companies are exposed to financial state variables that capture the change in investment opportunity. However, we find that there is no economically significant difference between the risk-adjusted returns of various ESG-rating-based portfolios. Furthermore, investors require return compensation for holding ESG stocks, but the risk associated with a poor ESG rating portfolio is comparable to that of a good ESG rating portfolio. We also find that the new covariance-adjusted penalized regression improves the out-of-sample performance of cross-sectional predictions of the ESG portfolio’s expected returns.

This paper is organized as follows. The following section provides the motivation and our incremental contribution with respect to the existing literature. The section after that describes the basic setup and theoretical results. The next section contains our main empirical results. In the final section, we conclude with brief comments.

2. Motivation and Our Incremental Contribution

In supervised learning, penalized regressions and the resulting shrinkage estimators are useful tools known to tackle real-world issues encountered in big data (see e.g., Athey 2018; Belloni et al. 2012; Chernozhukov et al. 2017; Kapetanios and Zikes 2018; Mullainathan and Spiess 2017; Stock and Watson 2019; and Varian 2014). The commonly used shrinkage estimators, which include Ridge regression (Hoerl and Kennard 1970), LASSO (least absolute shrinkage and selection operator) regression (Tibshirani 1996), and elastic net regression (Zou and Hastie 2005), are usually implemented under sparse and homoskedastic regression errors (Jia et al. 2013). We propose a Mahalanobis distance-based penalized regression in the presence of non-sparse and heteroskedastic residuals and show that the incorporation of covariance-adjusted shrinkage improves the out-of-sample performance. The Mahalanobis distance, first proposed by statistician P. C. Mahalanobis (1893–1972), is a well-known tool in multivariate analysis (Berrendero et al. 2020).

In many applications in finance, we come across a large number of primitive predictors \((p)\) compared to the number of observations \((n)\), and a regression error that is not sparse and displays heteroskedasticity with variance as a function of the primitive predictors. One such example is the cross-sectional regression model to evaluate the returns of a group of assets or portfolios by using various risk factor exposures. Examples of such equity risk factor models include the cross-sectional implementation of the capital asset pricing model (CAPM), the Fama–French five factors (Fama and French 2016), the q-factor model (Hou et al. 2015), etc. If the variability of returns of an asset or a portfolio displays cross-sectional clustering, the residuals of a risk factor regression model can be a function of the risk attributes, which can further lead to non-sparse residual covariance structure and heteroskedasticity. This paper contributes to the literature on the implications of ESG ratings in portfolio and risk management. Recent work by Giese et al. (2019) suggests that the traditional systematic risk factor contains ESG risk information, while Cohen (2023) argues that the ESG score is diminishing for US large-cap firms. In related work, Bannier et al. (2023) use a corporate social responsibility (CSR) criterion and argue that the lowest CSR-rated portfolios outperform their higher CSR-rated counterparts. Although the evidence is not always conclusive, the risk characterization of ESG stocks is valuable for portfolio managers. The results of this paper add knowledge to Giese et al. (2019) and Cohen (2023).

In terms of methodology, even though we can improve upon ordinary least squares (OLS) by generalized least squares (GLS) or other weighted least-squares alternatives, similar heteroskedasticity-corrected estimates under the shrinkage method are not readily available. There are reservations against the finite-sample properties of the GLS estimators. As an example, Angrist and Pischke (2008, sect. 3.4.1) argue that “if the conditional variance model is a poor approximation or if the estimates of it are very noisy, [the weighted] estimators may have worse finite-sample properties than unweighted estimators.” We propose an alternative method that uses Mahalanobis distance-based shrinkage
in which we shrink the risk factor regression parameters using their estimated variance–covariance matrix. We show that the resulting estimate is biased, but its variance is less than the corresponding unconstrained GLS variance. Thus, when \( p \) is as large as \( n \), we can obtain efficient estimates of the risk factor regression model parameters that can improve the credibility of the in-sample prediction model used to obtain predictions from out-of-sample observations.

### 3. Basic Setup and Theoretical Results

Suppose that \( Y = (y_1, \ldots, y_n) \) is an \((n \times 1)\) response vector, \( X = (x_1, \ldots, x_n) \) is the \((n \times p)\) matrix of covariates where \( x_j = (x_{ij}, \ldots, x_{nj})' \), \( j = 1, \ldots, p \) are the linearly independent predictors. Let \((Y, X)\) be the transformation of the originally collected data where all the covariates are standardized to have mean 0 and variance 1 and the dependent variable is transformed to have mean 0. We assume that \( Y = X\beta + \varepsilon \) where \( \varepsilon \) is a vector of residuals with \( E(\varepsilon|X) = 0, \) and \( V(\varepsilon|X) = \Omega \neq \sigma^2I \). It is well known that when \( p < n \) and the error term is heteroskedastic, the OLS estimate of \( \beta \) is consistent but the GLS estimate \( \hat{\beta}_{GLS} \) is efficient and can easily be obtained using standard statistical packages with or without the explicit knowledge of the source of heteroskedasticity. When \( p > n \), a possible alternative is to shrink the estimators by methods such as the Ridge, LASSO, and Elastic Net regressions. The shrinkage-based estimators are biased and do not have causal interpretations. Unlike OLS and GLS, both of which produce non-zero estimates to all coefficients, methods such as LASSO and Elastic Net perform automatic variable selection and produce parsimonious final models.

In the Ridge regression (Hoerl and Kennard 1970), we minimize the residual sum of squares (RSS) subject to a bound on the L2 norm (Euclidean distance) of the slope coefficients, whereas in the LASSO regression (Tibshirani 1996), we minimize the RSS subject to an L1 penalty (Manhattan distance) on the regression slope coefficients. In contrast, in the Elastic Net regression (Zou and Hastie 2005), we minimize the RSS subject to a combination of the penalizations using the L1 and L2 norms.

We propose that if the residuals display heteroskedasticity, that is, if \( \varepsilon|X \sim N(0, \Omega) \) with \( \Omega \neq \sigma^2I \), we can obtain an estimate of \( \beta \) using the Mahalanobis distance-based shrinkage (MDS) regression. The premise of the MDS regression is to minimize the sum of RSS subject to a bound on the norm for the beta vector scaled by the variance–covariance matrix as given by the following:

\[
\hat{\beta}_{MDS} = \arg\min_{\beta} \|Y - X\beta\|^2 \Omega^{-1} + \lambda \|X\Omega^{-1}X\|^{-1} \beta
\]  

where the tuning parameter \( \lambda \) can be selected by a cross-validation method. The MDS approach under (1) is a continuous shrinkage method that retains all the covariates but penalizes large coefficients through a modified L2 norm. The first part of the loss function in (1) is the same as that of the GLS. The second part of the loss function involves the regularization of parameters because it penalizes larger coefficient values. It is imperative that as \( \lambda \to 0 \), \( \hat{\beta}_{MDS} \to \hat{\beta}_{GLS} \). In the presence of a non-zero penalty term, if \( \alpha = 0 \) and \( \Omega = \sigma^2I \), \( \hat{\beta}_{MDS} = \hat{\beta}_{Ridge} \), and if \( \alpha = 1 \) and \( \Omega = \sigma^2I \), \( \hat{\beta}_{MDS} = \hat{\beta}_{LASSO} \).

The basic properties of the MDS estimate are given by the following result.

**Lemma 1.** The MDS estimate given by \( \hat{\beta}_{MDS} = QX\Omega^{-1}Y \) with \( Q = [(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}]^{-1} \) is biased and has a variance of \( V(\hat{\beta}_{MDS}) = QX'\Omega^{-1}XQ' \).

**Proof.** See the Appendix A. \( \square \)

Regarding the bias and MSE of regression predictions using MDS estimates, we have the following result.
Lemma 2. The bias and the MSE of $X\hat{\beta}_{MDS}$ are given by $-\lambda XX(Q'(X'\Omega^{-1}X)^{-1})^T\beta$ and $\text{Tr}[(X'\Omega^{-1}X)(XQ)\gamma(XQ)] + \lambda^2(\Omega^{-1}Q^{-1}\beta)^{-1}X'X(\Omega^{-1}Q^{-1}\beta)$, respectively.

**Proof.** See the Appendix A. □

We observe that, when $\lambda = 0$, $Q = (X'\Omega^{-1}X)^{-1}$, and, consequently, $\text{MSE}(X\hat{\beta}_{MDS}) = \text{MSE}(X\hat{\beta}_{GLS})$. Finally, regarding the efficiency of the MDS estimates, we provide the following result.

Lemma 3. The difference between the variance of $\hat{\beta}_{GLS}$ and $\hat{\beta}_{MDS}$ is non-negative definite.

**Proof.** See the Appendix A. □

It is important to recognize that the MDS estimate in (1) allows the minimization of RSS as well as a specific norm for the beta vector scaled by the variance–covariance matrix. Since the GLS estimate has no bias, its MSE is composed of full variance that can cause overfitting, which may further lead to inaccurate predictions. For the MDS estimate, however, we incorporate a covariance-adjusted penalty through the beta vector and adjust the variance of the shrinkage-based estimate. Consequently, we end up with an estimated model that is not overfitted and is thus capable of producing better predictions. This is seen in our empirical results presented in the next section.

We recognize that there are some existing works, such as that of Belloni et al. (2012), who use Lasso and post-Lasso for estimating the first-stage regression of endogenous variables on the instruments. In their method, the Lasso estimator selects instruments and estimates the first-stage regression coefficients via a shrinkage procedure. The post-Lasso estimator of Belloni et al. (2012) discards the Lasso coefficient estimates and uses the data-dependent set of instruments selected by Lasso to refit the first-stage regression via OLS to alleviate Lasso’s shrinkage bias. We stress that our objective in this paper is not to demonstrate efficiency gains from using optimal instruments. Since heteroskedasticity does not introduce bias to the estimator, one should not count on modeling the changing variance to produce superior forecasting results. Instead, we explore a practical penalized regression method in the presence of heteroskedasticity, which may provide a useful complement to existing approaches.

4. Empirical Analysis of ESG Firms

For the empirical illustration, we use the firm-level data from January 2009 to December 2022. We restrict our sample to companies with ESG ratings from Morningstar Sustainalytics. Morningstar rating measures the degree to which individual companies face financial risks from ESG issues, and then it rolls those individual scores up into an overall, portfolio-level score. In addition to Morningstar Sustainalytics, we utilize the CRSP-COMPUSTAT merged database and Kenneth French’s data library. Beginning in January 2009, each month we ranked every stock by its ESG rating from Sustainalytics and created three portfolios using ESG ratings. During the sample period, we also ranked all stocks into three size categories based on the market capitalization of each firm with an ESG rating. As a result, we obtained nine portfolios and obtained the time series of returns for each portfolio. The empirical methodology implemented in this paper is consistent with much of the prior literature.

Table 1 reports the average excess returns of nine portfolios constructed by independent sorts based on ESG rating and market cap. In our sample, firms with poor ESG ratings tend to be small firms, while firms with better ESG ratings tend to have large market capitalization. As is well known in the literature, it is visible in Table 1 that there is an inverse relationship between firm size and average excess returns in our sample. The average excess return of three small and three big portfolios is 0.52% per month and 0.41% per month, respectively. The table also shows an inverse relationship between ESG ratings
and average excess returns. As an example, for the small-cap portfolio, the average excess return monotonically decreases from 0.59% per month for the worst ESG rating portfolio to 0.46% per month for the best ESG rating portfolio. On average, irrespective of any size group we consider, the portfolio of firms with the best ESG ratings produces the lowest average returns than the portfolio of firms with the worst ESG ratings. As shown by the long-short portfolio, the difference between the worst and the best ESG rating portfolio's average return is not statistically significant.

Next, for each portfolio, we obtain the risk-adjusted return using the Fama and French (2015, 2016) multifactor risk model. Let $R_{i,t+1}$ be the excess returns of $i$th portfolio at period $t+1$. A full specification of the 5-factor Fama–French time-series regression model takes the following form:

$$R_{i,t+1} = \alpha_{i} + \beta_{i,Mktf} Mktf_{t+1} + \beta_{i,SMB} SMB_{t+1} + \beta_{i,HML} HML_{t+1} + \beta_{i,CMA} CMA_{t+1} + \beta_{i,RMW} RMW_{t+1} + \epsilon_{i,t+1}$$

where $Mktf$ is the excess market return, $SMB$ is the size factor, $HML$ is the value factor, $CMA$ is the investment factor, and $RMW$ is the profitability factor. While the size and value factors are empirically motivated, the investment and profitability factors are theoretically motivated. The addition of $CMA$ and $RMW$ captures the drivers of expected returns in the q-factor model and the size and value factors are associated with the static dimensions of the firms (Hou et al. 2015, 2020).

Table 1. Performance of nine ESG-rating- and size-sorted portfolios.

<table>
<thead>
<tr>
<th>Size</th>
<th>Small</th>
<th>Medium</th>
<th>Big</th>
<th>Average</th>
<th>Long–Short</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average excess returns</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worst</td>
<td>0.59***</td>
<td>0.54**</td>
<td>0.49**</td>
<td>0.54**</td>
<td>0.10</td>
</tr>
<tr>
<td>Medium</td>
<td>0.52**</td>
<td>0.41**</td>
<td>0.35*</td>
<td>0.43**</td>
<td>0.17</td>
</tr>
<tr>
<td>Best</td>
<td>0.46**</td>
<td>0.37*</td>
<td>0.40*</td>
<td>0.41**</td>
<td>0.06</td>
</tr>
<tr>
<td>Average</td>
<td>0.52**</td>
<td>0.44**</td>
<td>0.41**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long–Short</td>
<td>0.13</td>
<td>0.17</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Risk-adjusted returns</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst</td>
<td>0.009*</td>
<td>-0.005</td>
<td>-0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>-0.005</td>
<td>0.006</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best</td>
<td>0.008*</td>
<td>0.007</td>
<td>-0.003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *, **, and *** denote a rejection of the null hypothesis at the 10%, 5%, and 1% levels, respectively.

If the alphas corresponding to the 5-factor model is significant, it suggests that the risk factors are not successful in explaining the abnormal returns of the portfolio. The risk-adjusted returns of all nine double-sorted portfolios are also reported in the lower panel of Table 1. None of the alphas corresponding to the 5-factor FF regression model are statistically significant, at least at the 5% level. Unlike the average excess returns, the estimated alphas do not display any patterns as well. The insignificant alphas corresponding to the 5-factor model suggest that the risk-adjusted portfolio return does not persist after controlling for the risk factors. Although investors require return compensation for holding ESG stocks, the risk of a poor ESG rating portfolio is comparable to that of a good ESG rating portfolio, which suggests an additional risk dimension not incorporated into the literature.

As a next step, after obtaining the factor exposures from the time-series regression (2), the factor betas are used as explanatory variables in the following cross-sectional regression:
\[
R_i = \gamma_0 + \gamma_1 \beta_{i, MKEF} + \gamma_2 \beta_{i, SMB} + \gamma_3 \beta_{i, HML} + \gamma_4 \beta_{i, CMA} + \gamma_5 \beta_{i, RMW} + \epsilon_i, \forall i = 1, \ldots, 9
\] (3)

Traditionally, for evaluating a portfolio return, the 5-factor model includes a set of five primitive predictors in the time-series regression (2). As a result, the benchmark factor betas shows up in linear form in the cross-sectional regression (3). If the portfolio loadings with respect to the risk factors are important determinants of average returns, the slope coefficients of \(\beta_{i,j}\)'s from (3) should be statistically significant. In order to allow non-sparse covariance structure and heteroskedasticity, we evaluate two variations of the simulated cross-sectional regression. Note that the double-sorted portfolio construction scheme ensures that the test portfolios differ in their level of ESG ratings and market cap. The portfolio-level data that we use in (2) and (3) also helps us to avoid issues such as infrequent trading and outliers.

First, in order to allow non-sparse residuals, we assume that \(\text{Corr}(\epsilon_i, \epsilon_j) = \Omega_{ij} = \rho^{||i-j||}\) for all \(i, j\). Consequently, we construct \(\Omega = (\Omega_{ij})\) and compute the GLS and MDS estimates of the slope coefficients. We identify neighbors using the relative ESG rating and market cap of each portfolio and assume that the errors take a simple autoregressive form \(\epsilon_{i+1} = \rho \epsilon_i + \epsilon_t\) with \(\epsilon_t \sim N(0, \theta^2)\). We use five different pre-assigned values for \(\rho = 0.10, 0.20, 0.50, 0.80,\) and 0.90—and experiment with seven alternative pre-determined values of the tuning parameter \(\lambda = 0.5, 1.0, 3.0, 5.0, 10, 20,\) and 50. The MSE from the testing set using various methods is given in Table 2. We see that irrespective of any \(\rho\) we consider, the MDS regressions corresponding to \(\lambda < 5\) always produce a smaller MSE than the GLS counterpart. Even for high \(\lambda\), the MSE of GLS regressions exceeds those of MDS. For example, for \(\lambda = 5\), the testing sample MSE for MDS regressions varies between 0.0491 and 0.0570, whereas the same for GLS is between 0.0630 and 0.0697. The general trend that the MSE of GLS is higher than the MDS regressions suggests an improvement in the in-sample predictive accuracy.

Table 2. Testing sample MSE from GLS and MDS cross-sectional regressions.

<table>
<thead>
<tr>
<th>(\lambda)</th>
<th>0.5</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\rho)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.0625</td>
<td>0.0476</td>
<td>0.0488</td>
<td>0.0490</td>
<td>0.0494</td>
<td>0.0493</td>
<td>0.0495</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0630</td>
<td>0.0483</td>
<td>0.0493</td>
<td>0.0492</td>
<td>0.0491</td>
<td>0.0487</td>
<td>0.0510</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0638</td>
<td>0.0512</td>
<td>0.0522</td>
<td>0.0437</td>
<td>0.0534</td>
<td>0.0535</td>
<td>0.0537</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0680</td>
<td>0.0550</td>
<td>0.0545</td>
<td>0.0566</td>
<td>0.0565</td>
<td>0.0571</td>
<td>0.0564</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0697</td>
<td>0.0564</td>
<td>0.0570</td>
<td>0.0574</td>
<td>0.0570</td>
<td>0.0572</td>
<td>0.0581</td>
</tr>
</tbody>
</table>

Next, we implement a form of heteroskedastic regression by modeling the error variance with non-linear forms of heteroskedasticity. More specifically, we assume that the error variance is a function of all squares, cubes, and interactions of the factor betas. As a result of the incorporation of all the auxiliary variables, the number of parameters in the variance becomes 26.

Table 3 presents summary statistics for the out-of-sample predictions from heteroskedastic cross-sectional regressions where we split all testing portfolios into training and testing samples. We use the training sample for estimating prediction models and use the testing sample to assess how the model performs. To see the sensitivity of our results, we split the sample 100 times. Each time, we randomly select 56% of observations to create a training set and assign the remaining 44% set of the available data into a testing set. To maintain consistency with earlier results, we report the out-of-sample prediction errors GLS and MDS corresponding to seven alternative pre-determined values of \(\lambda\) as in Table 2.
Table 3. Performance of alternative methods under heteroskedastic cross-sectional regressions for the testing set.

<table>
<thead>
<tr>
<th>Summary Statistics</th>
<th>0.5</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Squared Error</td>
<td>0.2027</td>
<td>0.1905</td>
<td>0.1854</td>
<td>0.1918</td>
<td>0.1939</td>
<td>0.1925</td>
<td>0.1930</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>0.3091</td>
<td>0.2803</td>
<td>0.2250</td>
<td>0.2163</td>
<td>0.2159</td>
<td>0.2147</td>
<td>0.2145</td>
</tr>
<tr>
<td>Median Absolute Error</td>
<td>0.2539</td>
<td>0.2241</td>
<td>0.2245</td>
<td>0.244</td>
<td>0.2334</td>
<td>0.2640</td>
<td>0.2449</td>
</tr>
</tbody>
</table>

A close inspection of the reported results reveals that some MDS regressions have very similar values for average squared error and absolute error. For a large number of \( \lambda \) s, the average squared error is very close to 0.19. Compared to GLS, the MDS regression with \( \lambda = 0.5 \) results in a 5% reduction in average predicted errors. For \( \lambda = 1 \) and 3, the reductions in the average squared prediction errors become 9% and 6%, respectively. Similar observations can be made for average and median absolute error. In sum, we find that, based on the out-of-sample prediction errors, the slight dominance of MDS over GLS persists even in our heteroskedastic regressions. Therefore, the incorporation of covariance-adjusted shrinkage improves the out-of-sample performance of the ESG portfolio’s expected returns using the heteroskedastic cross-sectional regression model.

5. Conclusions

In this paper, we use a new covariance-adjusted shrinkage method and examine the risk–return tradeoff of a portfolio of companies that have tangible environmental, social, and governance (ESG) attributes from Morningstar Sustainalytics. We introduce a new type of penalized regression that minimizes the sum of residual squares subject to Mahalanobis distance and shows that the resulting estimates are biased but efficient compared to unconstrained GLS estimates. The results show that there is an absence of significant risk-adjusted returns of various portfolios constructed using ESG ratings. Furthermore, we find that the risks associated with various ESG rating-based portfolios are comparable. The incorporation of the covariance-adjusted shrinkage method in the cross-sectional regression also improves the predictive accuracy of expected returns. Ours is the first paper to provide a covariance-adjusted penalty that can be important for applied economic analysis. In terms of practical applications, our method is simple to implement and provides a compromise between model interpretation and improved prediction performance. If a researcher is concerned about heteroskedasticity, our method can be complementary to other existing approaches. Note that our methodology is not free from limitations. Our approach in this paper is pragmatic and based on the ease of an empirical appeal. Future work can focus on the theoretical and asymptotic properties of the MDS estimates.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from WRDS and are available with the permission of WRDS.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Proof of Lemma 1.

For the objective function (1), the first-order condition with respect to \( \beta \) is

\[
(X' \Omega^{-1}X + \lambda (X'\Omega^{-1}X)^{-1}) \beta = X' \Omega^{-1}Y
\]

To solve this, we obtain
\[
\beta_{MDS} = Q(X'\Omega^{-1}Y), \text{ where } Q = [(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}]^{-1} \\
= P\hat{\beta}_{GLS}, \text{ where } P = Q(X'\Omega^{-1}X) \text{ and } \hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.
\]

The expectation of \( \hat{\beta}_{MDS} \) is given by
\[
E(\hat{\beta}_{MDS}) = E[Q(X'\Omega^{-1}Y)] = Q[X'\Omega^{-1}E(Y)] = XQ'[\Omega^{-1}X]\beta
\]
\[
= [(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}](X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}]\beta - \lambda Q(X'\Omega^{-1}X)^{-1}\beta
= [\beta - \lambda Q(X'\Omega^{-1}X)^{-1}\beta] \neq \beta
\]

The variance of \( \hat{\beta}_{MDS} \) is given by
\[
V(\hat{\beta}_{MDS}) = PV(\hat{\beta}_{GLS})P' = Q(X'\Omega^{-1}X)(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}X)'Q' = Q(X'\Omega^{-1}X)Q'
\]
□

**Proof of Lemma 2.**

Using Lemma 1, the expectation of \( X\hat{\beta}_{MDS} \) can be expressed by
\[
E(X\hat{\beta}_{MDS}) = X\beta - \lambda XQ(X'\Omega^{-1}X)^{-1}\beta
\]

So, the bias is
\[
\text{Bias}(X\hat{\beta}_{MDS}) = E(X\hat{\beta}_{MDS}) - X\beta = -\lambda XQ(X'\Omega^{-1}X)^{-1}\beta
\]

The variance of \( X\hat{\beta}_{MDS} \) is given by
\[
V(X\hat{\beta}_{MDS}) = XV(\hat{\beta}_{MDS})X' = XQ(X'\Omega^{-1}X)Q'X' = (XQ)(X'\Omega^{-1}X)(XQ)'
\]

And
\[
\|\text{Bias}(X\hat{\beta}_{MDS})\|_2^2 = [E(X\hat{\beta}_{MDS}) - X\beta]'[E(X\hat{\beta}_{MDS}) - X\beta]
\]
\[
= [-\lambda XQ(X'\Omega^{-1}X)^{-1}\beta]'[-\lambda XQ(X'\Omega^{-1}X)^{-1}\beta]
\]
\[
= \lambda^2 \beta'(X'\Omega^{-1}X)^{-1}Q'X'Q(X'\Omega^{-1}X)^{-1}\beta
\]

Hence, the MSE is
\[
\text{MSE}(X\hat{\beta}_{MDS}) = Tr\left(V(X\hat{\beta}_{MDS})\right) + \|\text{Bias}(X\hat{\beta}_{MDS})\|_2^2
\]
\[
= Tr(XQ(X'\Omega^{-1}X)Q'X') + \lambda^2 \beta'(X'\Omega^{-1}X)^{-1}Q'X'Q(X'\Omega^{-1}X)^{-1}\beta
\]
\[
= Tr([X'\Omega^{-1}X](XQ)'(XQ)) + \lambda^2 Q(X'\Omega^{-1}X)^{-1}\beta)'X'X (Q(X'\Omega^{-1}X)^{-1}\beta)
\]
□

**Proof of Lemma 3.**

\[
V(\hat{\beta}_{GLS}) - V(\hat{\beta}_{MDS}) = (X'\Omega^{-1}X)^{-1} - Q(X'\Omega^{-1}X)Q'
\]
\[
= Q^{-1}(X'\Omega^{-1}X)^{-1}Q^{-1} - (X'\Omega^{-1}X)Q'
\]
\[
= Q[(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}((X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1})'] - (X'\Omega^{-1}X)Q'
\]
\[
= Q[(1 + \lambda(X'\Omega^{-1}X)^{-1})(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}) - (X'\Omega^{-1}X)]Q'
\]
\[
= Q[2\lambda(X'\Omega^{-1}X)^{-1} + \lambda^2(X'\Omega^{-1}X)^{-3}]Q'
\]
\[
= [(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}]^{-1}[2\lambda(X'\Omega^{-1}X)^{-1} + \lambda^2(X'\Omega^{-1}X)^{-3}][[(X'\Omega^{-1}X) + \lambda (X'\Omega^{-1}X)^{-1}]^{-1}]'
\]

Since each of the components of the matrix product is non-negative definite, the difference is non-negative definite.
□
Note
1. The rating is expressed as 1 to 5 globes. A higher number of globes indicates that the portfolio has lower (or negligible) ESG risk and vice versa. Accessed on 12 November 2023. See https://www.morningstar.com/sustainable-investing for details.

References

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.