1 Documentation

We construct a rank-weighted factor for each characteristic. Rank-weighted factors are created by

$$rc_{i,t,p} = \frac{\text{rank}(x_{i,t,p})}{n_t + 1} \tag{1}$$

$$z_{i,t,p} = \frac{rc_{i,t,p} - \bar{rc}_{t,p}}{\sum_{i=1}^{N_t} |rc_{i,t,p} - \bar{rc}_{t,p}|} \tag{2}$$

$$F_{t+1} = Z_t' R_{t+1}, \tag{3}$$

where $\bar{rc}_{t,p} = \frac{1}{N_t}\sum_{i=1}^{n_t} rc_{i,t,p}$ and Z_{t-1} is a $N_t \times P$ matrix of the normalized ranks of characteristics $z_{i,t,p}$, P being the number of characteristics and N_t the number of assets at time t.

As in Kozak et al. (2020), we orthogonalize the rank-weighted factors with respect to the market by estimating their exposures β. To make the method more comparable to the others, we add the market factor to the set of rank-weighted factors.

Kozak et al. (2020) solves the penalized minimization problem

$$\min_w (\hat{\mu} - \hat{\Sigma} w)' \hat{\Sigma}^{-1} (\hat{\mu} - \hat{\Sigma} w) + \frac{\lambda_2}{2} \|w\|_2^2 + \lambda_1 \|w\|_1, \tag{4}$$

where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and covariance matrix of returns of the basis assets, and $\|\cdot\|_1$ is the L^1 norm. The basis assets are either the rank-weighted factors or the principal components of the factors.

We include only L^2 shrinkage when the rank-weighted factors are the basis assets, and include L^1 shrinkage when working with principal components. Usually, L^1 shrinkage with multiple features precludes any closed-form solution. With principal components, however, a closed-form solution does exist since they are uncorrelated by construction

$$\tilde{w}_{PC} = \left(\hat{\Sigma}_{PC} + \lambda_2 I\right)^{-1} \hat{\mu}_{PC} \tag{5}$$

$$\hat{w}_{PC,p} = \text{sign}(\tilde{w}_p) \cdot (|\tilde{w}_p| - \lambda_1)^+, \tag{6}$$

where $\hat{w}_{PC,p}$ is the portfolio weight in the principal components p. The weights $\hat{w}_{PC}$ can be rotated back to weights in the rank-weighted factors.

We determine the hyperparameters λ_1 and λ_2 using cross-validation and grid search. Each month, we use the previous ten years as training data, and we use a standard version of five-fold cross-validation to pick hyperparameters. Specifically, we split the training data into five folds that maintain the temporal order of the data, such that the first fold contains the first two years and the last fold contains the last two years. The folds then take turns being assigned as the validation data, while the remaining four folds are assigned as training data. For example, when the first fold is the validation data, the method is estimated on the remaining four folds for each set of hyperparameters, and the estimated model is then evaluated on the first fold.

The validation metric is

$$1 - \frac{(\hat{\mu}_v - \hat{\Sigma}_v \hat{w})'(\hat{\mu}_v - \hat{\Sigma}_v \hat{w})}{\hat{\mu}_v' \hat{\mu}_v}, \tag{7}$$

where the subscript v indicates that the sample is the validation sample, and $\hat{w}$ are the weights estimated using the training data.

We repeat this procedure five times, average the performance for each hyperparameter set, and choose the best performance. We then re-estimate the model with the chosen hyperparameters on the full rolling training data. We repeat this process every month.

Each month, we estimate the principal components for each fold to be used in the cross-validation when the principal components are the basis assets. The market exposures of each rank-weighted factor are also estimated for each fold. This ensures an unbiased estimate of the performance of the entire method pipeline using cross-validation. After picking a set of hyperparameters λ_1 and λ_2, the market exposures and principal components of the total rolling 10-year window are computed.

Although the portfolio is updated on a monthly basis, we follow Kozak et al. (2020) and estimate the portfolio weight using daily data.

References

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.